A General Means for Depth Data Error Estimation of Depth Sensors

Qian She[1, a], HongYang Yu[2]
[1,2]Institute of Electronic Science and Technology
University of Electronic Science and Technology Chengdu 611731, China,
[a]Joycesq521@163.com

Abstract. At present the depth error estimation of the RGB-D sensor is aimed at a specific depth camera, the disadvantages of these methods is that it is put forword for a specific sensor and cannot be applied to the other sensors. In order to solve this problem, this paper proposes a general method to estimate the root mean square (RMS) error of depth data provided by general three-dimensional sensors. The method is applicable to three-dimensional sensors based on structured light, time-of-flight, stereo vision and other technologies. Use a common checkerboard to detect corner points and get two point clouds, one is the real point cloud of the image corner, and the other is the estimated point cloud of the corner point given by the device. After registrating of the two point clouds, the RMS error is calculated. The RMS error is generalized as a function of the distance between the RGB-D sensor and the checkerboard .The accuracy and practicability of the proposed method are verified by comparing it with the existing advanced depth error estimation methods.

Keywords: RGB-D sensors, Depth data, point clouds, RMS error.

1. Introduction

The three-dimensional information intuitively reflects the overall information of the target or scene, and the use of the depth data of the RGB-D sensor for information extraction is a typical topic in computer vision research. Several devices are currently available to obtain depth data, such as stereo cameras (Bumblebee, ZED), structured light-based sensors (Kinect v1 and RealSense), and time-of-flight-based (ToF) devices (Kinect v2). Although the above-mentioned devices have been widely used in robotics and computer vision, The depth data obtained have errors due to physical characteristics and measurement algorithm errors. For stereo cameras, the error is mainly due to high lens distortion, unreasonable camera layout, and poor image capture resolution. For devices based on structured light and Time-o-Flight, the error mainly comes from the imaging characteristics of the device itself and the external environment interference. Considering that each sensor has inherent errors, determining or estimating these errors is of great significance for subsequent studies.

2. Related Works

Different solutions to this problem have been proposed by A large number of domestic and foreign researchers. Tang Shengjun proposed a depth data error correction model in the literature[1], which is based on structured light depth sensor (Kinect V1). For the structured light sensor,its depth data accuracy is easily affected by the distortion of the infrared projector, depth camera and system error. The depth error is obtained through the true value and the depth data. The system parameters are optimized by the least squares method, and the systematic error is described by using the exponential function. Khoshelham and Elberink[2] proposed a model to determine the accuracy of the sensor, which takes into account the depth data from the disparity estimation.

Since 2014, Microsoft has released Kinect V2 based on ToF. Shortly after the sensor was put into use, C. Beder and M. Stommel respectivity used some simple linear models to describe and correct the errors[3][4]. Although such method is simple and requires little computation, the correction result is limited. Ying He et al[5] .attempted to introduce the data learning process into the error analysis of RGB-D cameras. Freedman D. et al. proposed to improve the accuracy of depth data by using multiple acquisition frequencies. The error correction methods of the above various ToF depth camera data are simple but lack of accuracy, while others have high precision but complicated calibration process.

The accuracy of the data obtained by 3D sensors has always been an aspect often involved in the selection of capture devices in application development. In the work of Rauscher[6], Zennaro[7], and Jorge[8] et al, several methods can be found to determine the accuracy and precision of sensors using ToF and structured light technology. The first two methods obtain depth data by placing a whiteboard or flat wall in front of the sensor. The depth error of the sensor is determined based on the three-dimensional data and the experiment conducted in [2]. Jorge et al. maintained the method of using point clouds to determine precision and accuracy of Kinect v1 and Asus Xtion. The accuracy of these sensors is calculated by comparing the distance between the acquired three-dimensional coordinates and the values measured by means of an external device.

For stereo cameras, Jin et al. [9] proposed an analysis and experimental method to determine the sensor depth error, and a second-order polynomial mathematical model to represent the relationship between the depth estimate and the actual value. Jing et al.[10] proposed a method based on the existing plane fitting technique to determine the global error of the sensor's working range. The above methods provide a global error for the sensor and do not provide an error model that is comparable to the results herein. It is shown in[11] that there is no significant difference between the results of using one board and two or three boards and that the ZED equipment can work normally within the range of 15 m. Therefore, this paper uses only one checkerboard. The methods presented above are all specific depth cameras. The disadvantage of the methods is that they are proposed for specific sensors and cannot be used to other depth devices.

In order to solve the problem, this paper proposes a universal method of depth data error estimation, which can be applied to various existing depth sensors, and can also be applied to devices that may be developed without geometric modeling of the underlying principle of data acquisition in the future. Use a common checkerboard to detect corner points and get two point clouds, one is the real point cloud of the image corner, and the other is the estimated point cloud of the corner point given by the device. After registrating of the two point clouds, the RMS error is calculated. The RMS error is generalized as a function of the distance between the sensor and the checkerboard .The accuracy and practicability of the proposed method are verified by comparing it with the existing advanced depth error estimation methods[6].

3. RGB-D Camera Depth Error

He et al [5]. believe that there are two main types of depth measurement errors for RGB-D cameras based on the TOF principle: systematic errors and non-systematic errors. Systematic errors are caused by the inherent characteristics of the system and the imaging principle. Non-systematic errors are caused by factors that are unknown in the environment. In practical applications, the system error is relatively fixed and can be eliminated by correction. Non-systematic errors are generally randomly varying. In this paper, the depth camera is first calibrated on the checkerboard and the obtained image is calibrated. Next, the error is measured as a black box. We only measure input and output errors. As mentioned above, the main advantage of the method proposed in this paper is its versatility, which can be applied to depth cameras with different principles.

The experiment uses a 6×9 size checkerboard for correction, and the size of each square is 60mm×60mm. The RGB-D sensor is used to collect 10 images of different angles. The position of the 10 checkerboards is shown in Fig. 1. The obtained depth image has more noise and the contrast is not high. In order to improve the precision of calibration, the histogram is equalized by the grayscale image to improve the contrast. The internal parameter matrix of the RGB-D camera can be obtained by checkerboard calibration. Table 1 shows the intrinsic parameters of the different RGB-D sensors.

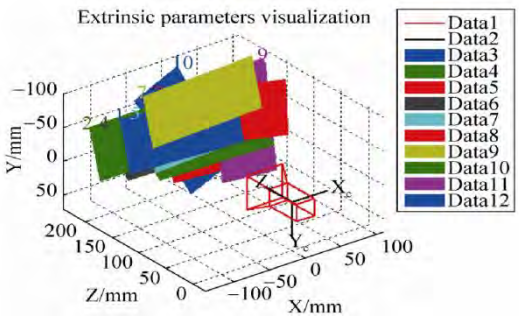
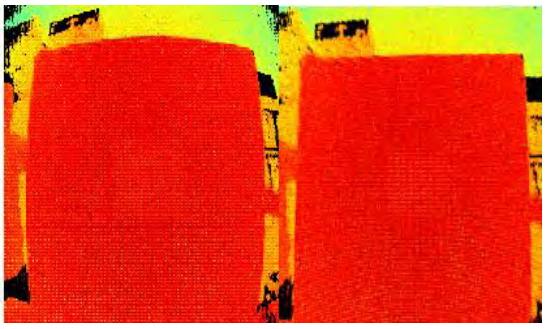


Fig. 1 The space diagram of the checkerboard correction picture

Table 1 Intrinsic parameters of the RGB-D sensors

	ZED (672px*376px)	**Kinect v1 (640px*480px)**	**Kinect v2 (512px*424px)**
f_x	337.042	521.856	367.213
f_y	337.042	524.648	367.213
C_x	353.668	316.627	257.168
C_y	172.565	250.795	201.542

Using the above internal reference matrix to correct the depth image, Fig. 2 takes the depth image obtained by Kinect v2 as an example. It can be seen that the distortion of the original image in the edge portion is well corrected.



（a）The original image（b）Rendering

Fig. 2 The correction of the depth image

To achieve depth, the RGB-D sensor uses techniques such as ToF, structured light, and stereo vision. There are certain errors in the depth values obtained by these techniques. In the actual environment, there are many objects with regular geometric shapes. When the depth sensor captures regular geometric objects and obtainsmeasurement values that do not conform to the geometric shapes, it can use the regular geometric shapes to correct the measured depth data. Next, this argument will be used to demonstrate a straightforward approach to correcting errors, demonstrating the applicability of this approach. The depth D_C obtained by the depth camera is expressed as follows:

$$D_c = D + e * s \tag{1}$$

$$S = \begin{bmatrix} s_{1,1} & s_{1,2} & s_{1,3} & \cdots & s_{1,u} \\ s_{2,1} & s_{2,2} & s_{2,3} & \cdots & s_{2,u} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ s_{v,1} & s_{v,2} & s_{v,3} & & s_{v,u} \end{bmatrix},$$

$$where \ \ s_{v,u} = \begin{cases} +1, & if \ D_{v,u} \geq \mu_D \\ -1, & otherwise \end{cases} \tag{2}$$

Where, for the depth pixel (v, u), $D_{v,u}$ is the depth value of the pixel, and μ_D is the average depth value of the matrix D.

4. Get the Depth Data RMS Error

The purpose of this paper is to quantify the error of depth data as the distance of the target object from the sensor changes, The method has the following characteristics:

Generality, ignoring the underlying principle of different depth sensors.

The concept is simple, the error model is calculated from the three-dimensional space, rather than being evaluated in the two-dimensional image coordinates as proposed in [2].

Practicality, the calibration process uses only a laser ruler and a flat checkerboard. The laser ruler is used to determine that the device is parallel to the checkerboard plane, and the checkerboard is used for camera calibration. The position of the RGB-D camera and the checkerboard is shown in Fig. 3:

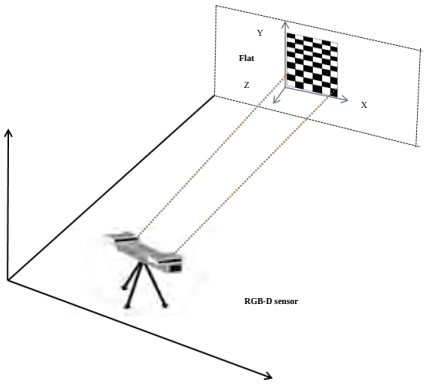


Fig. 3 Data acquisition scheme

4.1 Assumption

The method proposed in this paper assumes that the color image and the depth image have been calibrated and registered. Specifically, for a given color image I and the depth image D associated with it, the correspondence between the color image of the pixel point (u, v) and the depth map is as shown in Fig. 4:

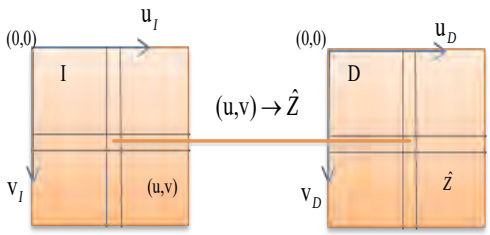


Fig. 4 Correspondence of RGB-D camera between color image and depth image

In the world coordinate system, the 3D point $P=[X, Y, Z]^t$ corresponds to $\hat{P} = [\hat{X}, \hat{Y}, \hat{Z}]^t$ in the camera coordinate system. According to the imaging characteristics of the RGB-D sensor, by Equation 3, $\hat{X}$ and $\hat{Y}$ can be calculated by the depth $\hat{Z}$, where f_x and f_y represent the focal lengths in the horizontal and vertical directions, in the units of pixels, C_X and C_Y are projection center coordinates. These parameters are intrinsic parameters of RGB-D sensor and can be obtained through the calibration procedure.

$$\hat{X} = \frac{u - C_x}{f_x}\hat{Z} \qquad \hat{Y} = \frac{v - C_y}{f_y}\hat{Z} \tag{3}$$

4.2 Point Cloud Generation

In order to obtain the depth error data of a general RGB-D sensor, this paper evaluates the calibration error in three dimensions. For this, two stereo planar point clouds are constructed: $\tau = \{P_1, P_2, \ldots, P_N\}$ is the ideal point cloud, $\epsilon = \{\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_N\}$ is the estimated point cloud.

The ideal point cloud τ is obtained through the checkerboard. For a checkerboard, the total number of points is N=ST, S means that the checkerboard has S lines, and T means that the checkerboard has N columns.

$\tau = \{[0,0,0]^t, [d, 0,0]^t, [2d, 0,0]^t, \ldots, [0, d, 0]^t, [0,2d, 0]^t, \ldots, [(T-1)d, (S-1)d, 0]^t\}$, where d is the size of the checkerboard square.

Estimated point cloud ε is obtained by detecting the corner points of the checkerboard. Specifically, the image coordinates $(u_i, v_i), i = 1,2, \ldots, N$ are detected by the checkerboard corner

detection algorithm[12], $\hat{Z}_i = D(u_i, v_i)$, $\hat{X}_i$ and $\hat{Y}_i$ can be calculated by Equation 3. The specific process is shown in Fig. 5.

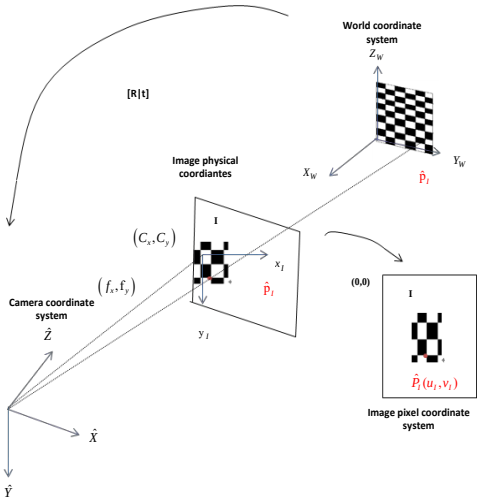


Fig. 5 The projection of estimated point and the ideal

RGB-D sensors, especially structured light sensors and ToF sensors, have a relatively close working range (about 0.5-5 m), while stereo cameras can work up to 20 m, such as ZED. Therefore, it is necessary to detect the corner points of the checkerboard in images captured at short and long distances. Because the method proposed in the paper needs to detect the corner points from a long distance, the Geiger et al. corner detection algorithm is used in the literature[12]. The algorithm can successfully detect corner points indoors and outdoors and under different illumination conditions.

4.3 Point Cloud Registration depth Error Estimation

After obtaining the ideal point cloud and the estimated point cloud, the depth error of the device can be estimated. Before estimating the depth error, the two point clouds are first registered in the same coordinate system, and the rotation matrix R and the translation matrix T can be derived from Equation 4.

$$R, T = \operatorname{argmin} \sum_i \left\| P_i - \left(\hat{R}\hat{P}_i + \hat{T} \right) \right\|^2 \tag{4}$$

After registration in the same coordinate system, the depth error e_i^2 of point i is obtained throuth calculating the square of the Euclidean distance of points P_i and $\hat{P}_i$, as shown in Equation 5:

$$e_i^2 = \left\| P_i - \hat{P}_i \right\|^2 = \left(X_i - \hat{X}_i \right)^2 + \left(Y_i - \hat{Y}_i \right)^2 + \left(Z_i - \hat{Z}_i \right)^2 \tag{5}$$

Note that if the estimated point P_i is in the system coordinate system as $\hat{P}_i$, then e_i^2 is 0, but this is not possible. Specifically, if no error occurs during the matching process, e_i^2 is directly related to the depth error of the point.

The corresponding error $E(Z^j)$ for a specific distance Z^j is equal to the sum of e_i^2 of N points, as shown in Equation 6:

$$E(Z^j) = \sum_{i=1}^{n} e_i^2 \tag{6}$$

4.4 RMS Error Model Estimation

The RMS error is calculated by Z^j， j=1,…,M at different distances. As shown in Equation 7:

$$e_{RMS}^j = \sqrt{\frac{1}{n} E(Z^j)} \tag{7}$$

The values of the exponential or polynomial interpolation function can be estimated by different e_{RMS}^j values. Equation 8 is an exponential model with two parameters a, b. Equation 9 is a polynomial model with three parameters a, b, and c.

$$f_1(\hat{z}^j) = a e^{b\hat{z}^j} \tag{8}$$

$$f_2(\hat{z}^j) \quad a \quad b\hat{z}^j \quad c\hat{z}^{j\,2} \tag{9}$$

Both models calculate the average distance $\hat{z}^j$ for all n points.

$$\hat{z}^j = \frac{1}{n} \sum_1^n \hat{z}_i \tag{10}$$

Finally, the two models' parameters are obtained by least squares method by means of N pairs $(\hat{z}^j, e_{RMS}^j)$ values， j=1,…,N.

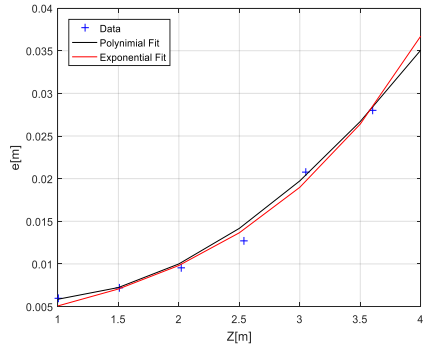
5. Experiments and Results Analysis

In order to prove the versatility of the proposed method in this paper, we use three kind s of equipment: change the distance between each sensor (kinect v1, kinect v2, ZED) and t he checkerboard, as shown in Fig 3, capture data every 0.25m.

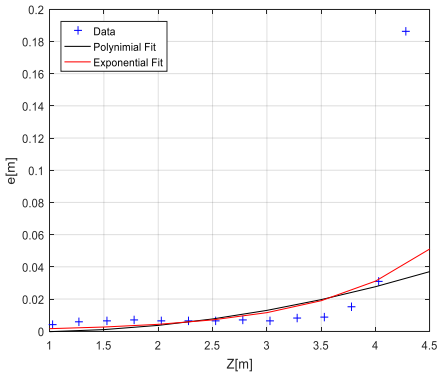
5.1 Get the RMS Error

The paper has obtained the intrinsic parameters of each depth sensors, as shown in Table 1. Using a calibrated checkerboard, in order to obtain more accurate depth data, 150 frames of data were acquired for each specific distance.

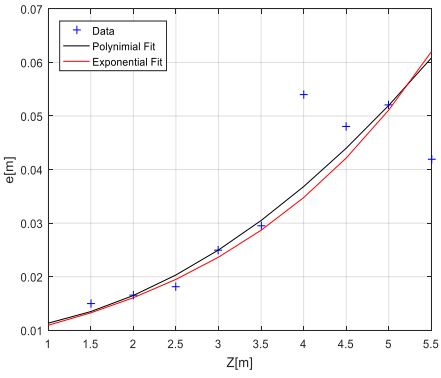
Using the RMS error, a model of kinect v1, v2, and ZED was obtained. As can be seen in Fig. 6, the exponential model and polynomial model can continuously represent this error. In order to more accurately verify whether the fitted model is correct, we determine whether the obtained error model is accurate by fitting the goodness statistic (SSE, R_squre and standard deviation S.).When SSE and S are close to 0, R squre is close to 1 as the best match model. Table 2 gives the model goodness-of-fit statistics for each device. For Kinect v1 and ZED sensors, the best model is a model represented by an exponent. For Kinect v2, the best model is a model represented by a polynomial. Table 3 gives the parameters of the best fit model for the different sensors.



（a）



（b）



（c）

Fig. 6 Depth error for (a) Kinect v1, (b) Kinect v2 and (3) ZED

Table 2 Model goodness-of-fit of statistics analysis

RGB-D sensor	Mathematical model	R_{square}	SSE	S	Best mode
Kinect v1	$a\ b\hat{z}^j\ c\hat{z}^{j^2}$	0.98	1.543×10^{-6}	0.00069	Yes
	$ae^{b\hat{z}^j}$	0.96	0.0001783	0.00603	NO
Kinect v2	$a\ b\hat{z}^j\ c\hat{z}^{j^2}$	0.86	0.0001225	0.00354	NO
	$ae^{b\hat{z}^j}$	0.97	5.728×10^{-5}	0.00268	Yes
ZED	$a\ b\hat{z}^j\ c\hat{z}^{j^2}$	0.96	3.976×10^{-5}	0.00302	NO
	$ae^{b\hat{z}^j}$	0.99	3.538×10^{-6}	0.00104	Yes

Table 3 Parameters of the best fit models

RGB-D sensor	Best mathematical model	Parameters
Kinect v1	$a\ b\hat{z}^j\ c\hat{z}^{j^2}$	a=0.002797;b=-0.004297;c=0.007311
Kinect v2	$ae^{b\hat{z}^j}$	a=0.0006022;b=0.9894
ZED	$ae^{b\hat{z}^j}$	a=0.007512;b=0.3798

5.2 Verification of the Correctness of the Results

Next, in order to verify the accuracy and practicability of the method proposed in this paper, the RMS error curve is compared with the curve in literature[6] and the trend of the curve is compared. A number of papers have been thoroughly analyzed for the depth error of kinect v2. The error curve obtained in this paper is consistent with the error curve obtained in the paper Rauscher et al.[6] , as shown in Fig. 7. As the distance of the sensor increases from the target, the two polynomial depth error curves tend to overlap. As can be seen from the scatter gram of Fig. 8, there is a strong positive correlation between the two curves with a correlation coefficient of 0.99. The method of this scheme can be applied to other 3D sensor well.

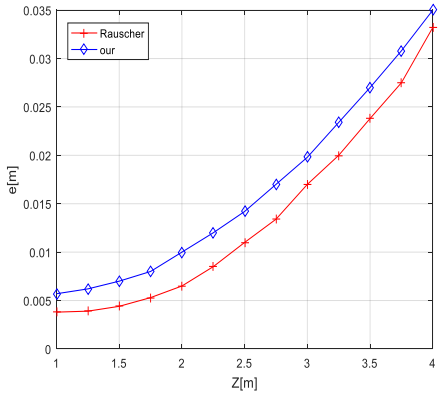


Fig. 7 Depth error comparison of Kinect v1

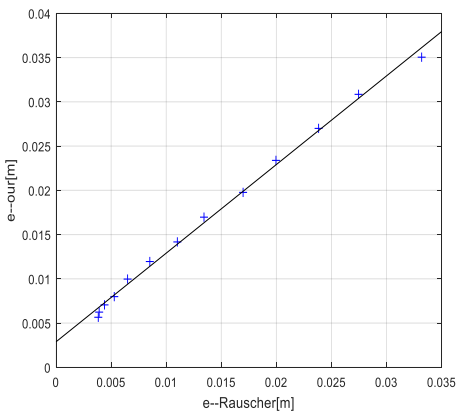


Fig. 8 Dispersion graphs of Kinect v1 depth error curves

6. Conclusion

The method proposed in this paper can obtain the depth RMS error of the general RGB-D sensor without ignoring the underlying implementation principle of the depth sensor. The calibration method is simple and practical. It can calculate and compare the RMS error of the three most commonly used depth sensors in computer vision and robotics applications. It can also be used to quickly evaluate the quality of RGB-D sensors. Future work will verify the use of SLAM algorithms for robots. The applicability of navigation in complex tasks.

References

[1]. ShengJun Tang. RGB-D indoor high-precision 3D mapping method with multi-view image enhancement [D], Wuhan: Wuhan University; 2017.

[2]. Khoshelham, K.; Elberink, S.O.: Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. Sensors 2012, 12, 1437–1454.10.3390/s120201437.

[3]. Beder C, Koch R. Calibration of Focal Length and 3D Pose Based on the Reflectance and Depth Image of a Planar Object[J]. International Journal of Intelligent SystemsTechnologies&Applications, 2008, 5(3/4):285-294.

[4]. Mills J P, Chandler J H. ISPRS Commission v Sysposium Image Engineering And Vision Metrology[J]. Photogrammetric Record, 2007, 22(117):94-96.

[5]. He Y, Liang B, Zou Y,at al. Depth Errors and Correction for Time-of-Fligth(ToF) Cameras[J]. Sensors (Basel, Switzerland), 2017, 17(1):92-110.

[6]. Rauscher G, Dube D, Zell A. A Comparison of 3D Sensors for Wheeled Mobile Robots. In Intelligent Autonomous Systems 13[M]. Springer International Publishing, 2016.

[7]. Zennaro S, Munaro M, Milani S, at al. Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications[C]. In Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME), 2015:1-6.

[8]. Gonzalez-Jorge H, RiveiroB, Vazquez-FernandezE, at al. Metrological evaluation of Microsoft Kinect and Asus Xtion sensors[J]. Measurement, 2013, 46:1800-1806.

[9]. Jin B, Zhao Li. Error modelling of depth estimation based on simplified stereo vision for mobile robots[J]. Comput. Model. New Technol, 2014, 18:450-454.

[10]. Jing C, Potgieter J, Noble F, at al. A comparison and analysis of RGB-D cameras' depth performance for robotics application[C]. In Proceedings of the 2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), 2017:1-6.

[11]. Ortiz L, Cabrera V. Depth Data Error Modeling of the ZED 3D Vision Sensor from Stereolabs [J]. ELCVIA Electron, 2018, 17:1-15.

[12]. Geige A, Moosmann F, Car O, at al. Automatic camera and range sensor calibration using a single shot [C]. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, 2012:3936-3943.