RGB-D Sensors Calibration for Service Robots SLAM

Yue Sun[1,2], Jingtai Liu[1,2(✉)], and Lei Sun[1,2]

[1] Institute of Robotics and Automatic Information System, Nankai University, Tianjin 300071, China
[2] Tianjin Key Laboratory of Intelligent Robotics, Tianjin 300071, China
liujt@nankai.edu.cn

Abstract. One of the major research directions in robotic vision focuses on SLAM using RGB-D sensors. The information can be used for decision making of robots and other areas that require precise position as a feature. This paper presents a novel algorithm to calibrate the RGB-D sensors for service robots SLAM. The distortions of the RGB and depth images are calibrated before the sensor is used as a measuring device for robot navigation. The calibration procedure includes the correction of the RGB and depth image as well as alignment of the RGB lens with the depth lens. The key advances in this paper are: a new method for RDB-D sensors calibration, and use of a depth distortion correcting model to help improve measurement precision. We experimentally verify our algorithm using varies of methods. The results show that, typically, our approach provides accurate calibration and the RGB-D sensors could provide reliable measurement information for robots navigating in unknown environments.

Keywords: RGB-D sensors · Camera calibration · Robot SLAM · Depth distortion correcting

1 Introduction

When a robot is navigating in an unknown environment, it relies on sensors to recognize the outside world and estimate the state of the robot itself to achieve the task of autonomous navigation. RGB-D sensors as low cost Kinect are widely used in robotics applications. Obstacle Avoidance (OA), Augmented Reality (AR), Simultaneous Localization and Mapping (SLAM), Mobile Object Tracking (MOT) all are needed accurate information about the position of objects in the environment.

Depth information is an important cue for robot SLAM and any other scenes in a real world. In all of the depth needed applications, it's very important to gain robust and accurate depth image of the relevant RGB image. Different kinds of RGB-D sensors can provide real time depth estimation at each pixel, which give us 3D information. The depth information has generally been acquired using stereo cameras or other expensive hardware such as laser range scanner or Time of Flight cameras. But most of them are too expensive to be extensively used or capture only two-dimensional images without depth information of environmental objects. This study uses Microsoft

Y.-J. Zhang (Ed.): ICIG 2015, Part III, LNCS 9219, pp. 197–211, 2015.
DOI: 10.1007/978-3-319-21969-1_18

Kinect as the RGB-D sensor to capture color and depth images as environmental information for service robots SLAM because of its functions and low price.

The distortions of color images and depth images need to be calibrated before the RGB-D sensor is applied as a measuring device. The calibration includes internal calibration of color and depth camera as well as relative pose calibration between the camera pair. Many procedures have been developed to calibrate the color image by determining the camera intrinsic parameters [1, 2]. However, such earlier methods are fail to produce an answer in depth sensor calibration. On the other hand, the distortion of the depth image is due to the relative pose between the RGB sensor and the depth sensor. In the literature, Herrera [3, 4], Khoshelham [5] proposed calibration algorithm for structured light depth sensor. Zhang [6] adopt the method of depth camera calibration from depth images. Smisek [7] analyzed the measurement error of depth camera calibration from infrared projected light. Raposo [8] proposed several modifications to the work of Herrera [4] that improved runtime using less images. Nakayama [9] propose an alignment method which is based on line segments to improve the camera pose accuracy. However, all these methods are either not robust, or produce a result which is not accurate enough for real applications, or require relatively complex processing procedure. So there is no popular method to transfer and align the depth image with the RGB image.

In this paper, the proposed a novel method uses algorithm based on distortion for RGB-D sensors calibration. We choose the origin of the RGB sensor as the reference frame and transform the depth sensor to align with the RGB sensor frame. Depth information is used to distorting the disparity error. The method was validated in the experiments by performing SLAM tasks using a RGB-D sensor as the only sensing device. The contribution of this paper is to solve the problem of aligning the depth sensor with the RGB sensor as well as calibrate the RGB-D sensor.

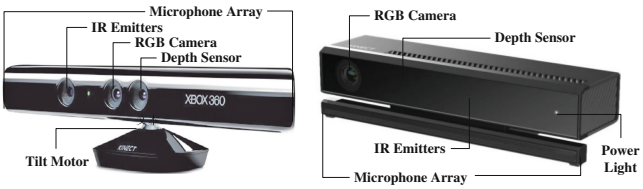
The rest of the paper is structured as follows: Sect. 2 studies the principle of RGB-D sensors like Kinect and introduces the background of camera calibration. Section 3 presents a novel calibration method of typical RGB-D sensors and depth distortion algorithm proposed by this paper. In addition, we use image processing technology to locate landmark and measure its distance in the coordinate system of the robot. In Sect. 4 the experimental setups are detailed and the results are discussed. This paper ends with conclusion and future work in Sect. 5.

2 Overview of RGB-D Sensors and Background

2.1 RGB-D Sensors

This section will provide brief introduction to the RGB-D sensors and how they work. There are two kinds of RGB-D sensors, one is structured light depth sensor such as Kinect 1, Xtion Pro live, Creative Senze3D and so on. The Kinect1's depth camera is not a real camera; it's a virtual camera, created by combining images from the real IR camera with light patterns projected by the IR emitter. But for Kinect 2, the technology for the depth estimation changes from structured light to time of flight (ToF), another type of method to sense depth. In a time-of-flight depth camera, the depth camera is a real camera, with every pixel containing a real depth measurement.

The Kinect has two cameras and a laser-based IR projector. Figure 1 shows their placement on the device. The sensors are able to produce two images: a color image from RGB camera and a depth image from the couple of IR projector and camera. The depth in a depth image is the perpendicular distance from the object to the sensor plane rather than the actual distance from the object to the sensor as shown in the Fig. 2.



(a) Kinect 1 for XBOX (b) Kinect 2 for windows.

Fig. 1. RGB-D sensor Kinect.

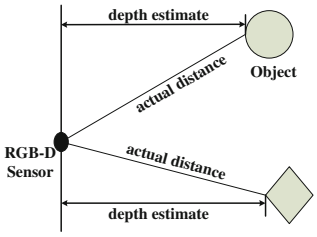


Fig. 2. Depth distance vs actual distance.

2.2 RGB Camera Model

In this work, we use a similar intrinsic model of the color camera as Heikkila [10]. Let $X_c = [x_c \quad y_c \quad z_c]^T$ be a point in RGB camera coordinates while $P_c = [u_c \quad v_c]^T$ be a point in color image coordinates. The point is normalized by $X_n = [x_n \quad y_n]^T = [x_c/z_c \quad y_c/z_c]^T$. Distortion is performed:

$$X_g = \begin{bmatrix} 2k_3 x_n y_n + k_4\left(r^2 + 2x_n^2\right) \\ k_3\left(r^2 + 2y_n^2\right) + 2k_4 x_n y_n \end{bmatrix} \tag{1}$$

$$X_k = \left(1 + k_1 r^2 + k_2 r^4 + k_5 r^6\right) X_n + X_g \tag{2}$$

where $k_c = [k_1 \quad k_2 \quad k_3 \quad k_4 \quad k_5]$ is a vector containing the distortion coefficients. The image coordinates are obtained:

$$\begin{bmatrix} u_c \\ v_c \end{bmatrix} = \begin{bmatrix} f_{cx} & 0 \\ 0 & f_{cy} \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix} + \begin{bmatrix} u_{0c} \\ v_{0c} \end{bmatrix} \tag{3}$$

where $f_c = [f_{cx} \quad f_{cy}]$ are the focal lengths and $p_{0c} = [u_{0c} \quad v_{0c}]$ is the principal point. The model for RGB camera is described by $L_c = \{f_c, p_{0c}, k_c\}$.

2.3 Depth Camera Model

The transformation between depth camera coordinates $X_d = [x_d \quad y_d \quad z_d]^T$ and depth image coordinate $P_d = [u_d \quad v_d]^T$ uses a similar model to the RGB camera.

The relation between the disparity value d_k and the depth z_d is modeled by the equation:

$$z_d = \frac{1}{c_1 d_k + c_0} \tag{4}$$

where c_1 and c_0 are part of depth sensor intrinsic parameters. The depth camera presents a depth distortion which has been modeled by Herrera [4]:

$$d_k = d + D_\delta(u, v) \cdot \exp(\alpha_0 - \alpha_1 d) \tag{5}$$

where d is the distorted disparity returned by the Kinect. The depth camera model is described by $L_d = \{f_d, p_{0d}, k_d, c_0, c_1, D_\delta, \alpha\}$.

2.4 Extrinsic and Relative Pose

Figure 3 shows the different reference frames present in a scene. The relative pose between the sensors can be denoted by $T = \{R, t\}$, where R is a rotation and t is a translation.

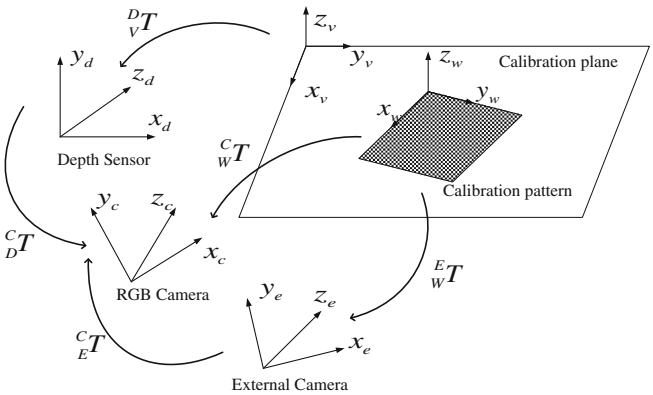


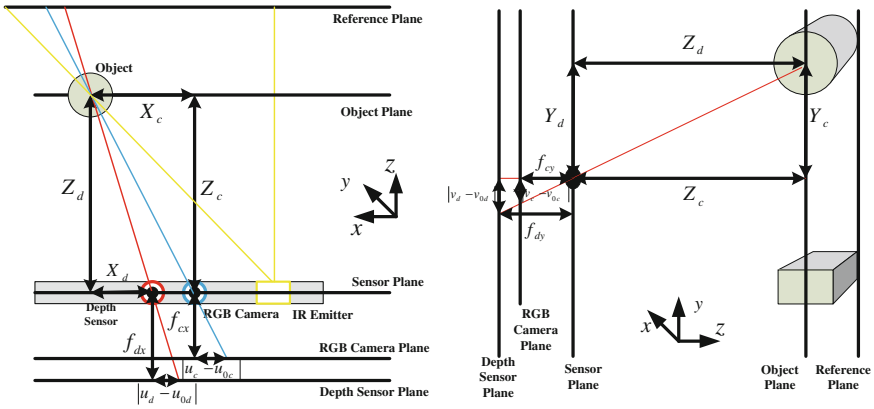
Fig. 3. The relative pose between different sensors. {C}, {D} and {E} are the RGB, depth and external cameras. {W} and {V} are the calibration pattern and calibration plane.

3 Calibration of RGB-D Sensors

In this study, we develop an algorithm to calibrate the RGB-D sensor based on depth distortion corrected. All the calibrations done below are based on RGB and infrared images of chessboard patterns, as well as depth images returned from RGB-D sensor. The concept and procedures are described in the following subsections.

3.1 Calibration Principle and Method

Figure 4 shows the structure of depth visual and the calibration principle of RGB-D sensor in both front view and side view.



(a) Front View (a) Side View

Fig. 4. The structure of depth visual.

In this work, we compare our method with the work of Herrera [4] that uses image-disparity map pairs of planes to accurately calibrate a RGB-D sensor. The method relies on the steps of calibration and the format of images. So we propose several modifications to this pipeline that improve applicability and stability, as show in Fig. 5. We collect depth images by OpenNI and convert them into disparity images instead of catching disparity images directly so that we can finish calibration in real time. But the edges and corners of the depth images are not obvious leading to low precision. So this paper proposes a new calibration method that calibrate depth camera with infrared image and increase a depth distortion correcting model.

The calibration model described below will be used to retrieve the intrinsic and extrinsic parameters of the RGB-D sensors. As show in Fig. 6, a block diagram of our calibration method is presented. We consider that the intrinsic matrices of RGB camera and depth sensor can be obtained using calibration toolbox for Matlab [11]. While the extrinsic parameters which are the translational and rotational displacements between the RGB camera and the depth sensor providing the depth image will be estimated.

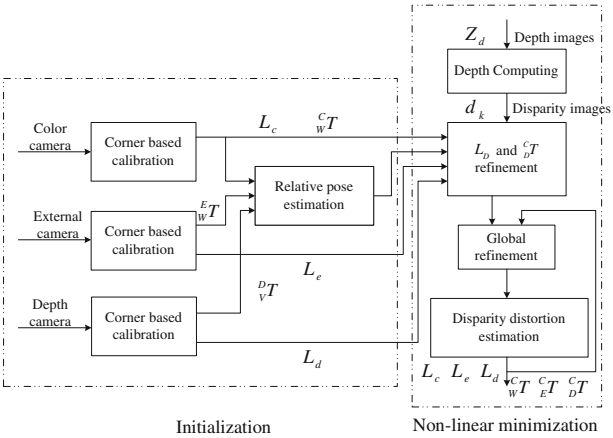


Fig. 5. The framework of the proposed calibration method.

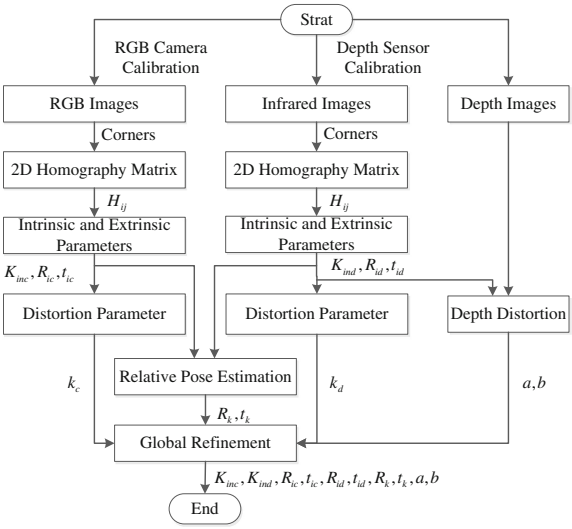


Fig. 6. Calibration algorithm.

After calibrating the intrinsic and extrinsic parameters, the model will be nonlinear optimized by a global cost function:

$$\min\{L_c + L_i + \alpha L_d\} \tag{6}$$

where α is ratio to optimize the relationship between depth distortion correction and image reprojection error.

The reprojection error of color camera and infrared camera, as well as the deviation of depth distortion for the cost function can be obtained by the equations:

$$L_c = \sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\hat{m}_{ijc}\left(K_{inc}, R_{ic}, t_{ic}, M_{ij}\right) - m_{ijc}(K_{inc}, k_c)\right\|^2 \tag{7}$$

$$L_i = \sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\hat{m}_{ijc}\left(K_{ind}, R_{id}, t_{id}, M_{ij}\right) - m_{ijd}(K_{ind}, k_d)\right\|^2 \tag{8}$$

$$L_d = \sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\hat{d}_{ijd}\left(R_{ic}, t_{ic}, M_{ij}\right) - d_{ijd}\right\|^2 \tag{9}$$

which can be solved by Newton iteration solver or the Levenberg-Marquardt algorithm.

3.2 Depth Distortion Correcting Algorithm

Using the calibration results obtained with data sets acquired by RGB-D sensor, we estimate the depth distortion correcting model with different images of a plane at different depths. We get the measured distance read from depth image and the real distance after calibration and fitted the linear relationship between them as shown in Fig. 7. It can be seen that the measured value and the true value are on the whole close but there are some deviations between them. And experimental data (Fig. 8) shows that the farther the distance of the target object, the greater the depth error of measurement.

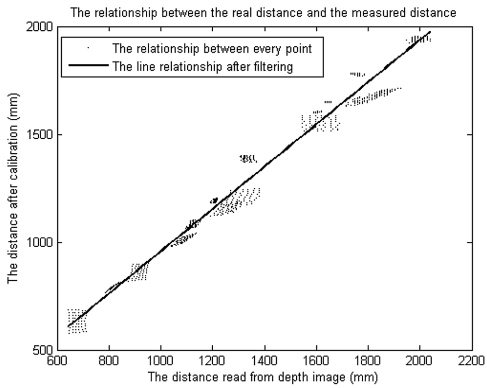


Fig. 7. The relationship between true value and measured value.

The linear relationship between the measured depth and the real depth can be established by:

$$d_c = ad_0 + b \tag{10}$$

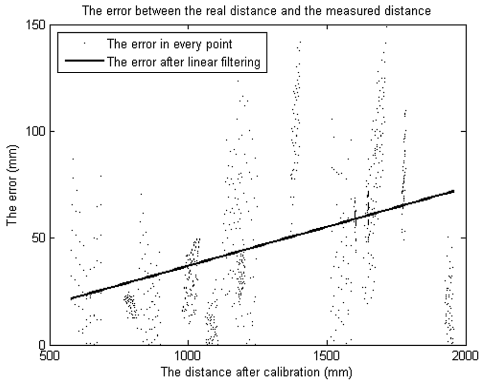


Fig. 8. The depth error between true value and measured value.

where d_0 is distance read from depth image and d_c is true distance. The coefficients a and b are varied along with different RGB-D sensors.

3.3 Robot SLAM Calibration

When a robot performs SLAM tasks, the states of robot and landmarks in the environment are estimated on the basis of measurement information. The observed image feature can be initialized using 3D coordinates in the world frame as shown in Fig. 9.

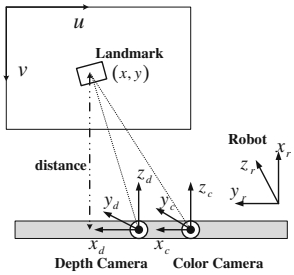


Fig. 9. RGB-D sensor localization system.

For calculating the coordinates of landmark under the robot coordinate system, we need to calibrate the relative pose between robot system and RGB-D sensor system by transformation matrix as follow:

$$ {}^{r}_{c}T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{11} $$

Suppose that the coordinates of landmark under RGB-D sensor system can be described as $X_c = [x_c \quad y_c \quad z_c]$ while the coordinates of landmark under robot system can be described as $X_r = [x_r \quad y_r \quad z_r]$.

1. Distortion correction for RGB and depth images with calibration results.
2. Calculate the depth of each pixel in RGB image by:

$$s\begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = K_{inc}\left(Z_d R_k \begin{bmatrix} \frac{u_d - u_{0d}}{f_{dx}} \\ \frac{v_d - v_{0d}}{f_{dy}} \\ 1 \end{bmatrix} + t_k\right) \tag{12}$$

3. Do a binary morphology operation on registration RGB image to get the connected domain of landmark.
4. Get the coordinates of landmark under robot system from the center of landmark on RGB image by:

$$X_r = R^{-1}(X_c - t) \tag{13}$$

5. Remove the glitter according to the position relationship between center and corners of landmark.

4 Experiments and Results

In this section the methods to test and verify the robust and accuracy of RGB-D sensors calibration will be discussed. The calibration results for different RGB-D sensors will be given in the first part and two experiments including ground truth and robot SLAM are carried out to validate the proposed algorithm.

4.1 Calibration

Two sets of experiments were conducted in order to prove the accuracy of our calibration method. The first one uses the data set acquired by Kinect1 (Fig. 10) while the second set uses images acquired by Kinect2 (Fig. 11), in order to further validate the results. For each set, the captured images were divided into calibration and validation images. All results presented here were obtained from the validation images.

The calibration results and average RMS reprojection errors for two sets are shown in Table 1. It can be seen that the model was corrected since the reprojection errors significantly decreased from 0.4 to 0.1. Our method, on the other hand, yields good results with both Kinect1 and Kinect2.

4.2 Ground Truth

To test the performance of robot SLAM using the RGB-D sensor, the sensor is carried to follow a track with each landmark at a distance of one meter, as shown in Fig. 12.

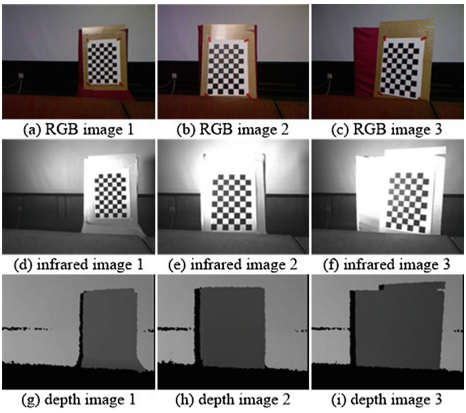


Fig. 10. Images used for calibration from Kinect1.

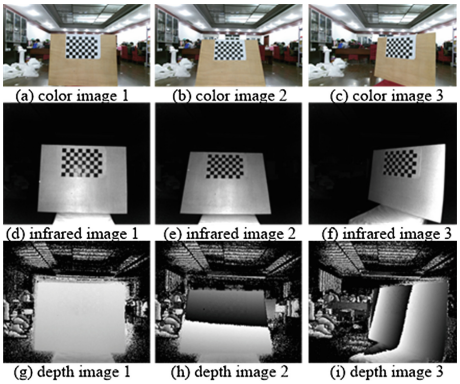


Fig. 11. Images used for calibration from Kinect2.

The experimental results of ground truth are listed in Table 2. The coordinates in the table are localization errors that the difference between estimated locations and ground truth locations of each landmark. The four columns of the table are acquired with the parameters of manufacturer calibration, uncorrected calibration, Herrera's calibration [4] and our calibration. As shown in the table, the localization results are successfully estimated to be much more close to the true coordinate of the real world using the algorithm proposed in our paper. Results clearly show that under the same conditions, our method achieves a better accuracy. This can be confirmed in Figs. 13 and 14 where the average localization errors obtained in each landmark for different images are shown. It can be seen that after applying the distortion correction, the depth errors significantly decrease from 56 mm to 21 mm on average.

Table 1. Kinect calibration result

Intrinsic parameters of RGB camera						
Coefficients	f_{cx}	f_{cx}	u_{c0}	v_{c0}	e_{cx}	e_{cy}
Kinect1	530.77	530.64	295.66	253.79	0.1096	0.1104
Kinect2	1132.22	1125.53	959.50	539.50	0.1895	0.1895
Coefficients	k_1	k_2	k_3	k_4	k_5	
Kinect1	0.2116	-0.4111	-0.0031	-0.0054	0.0000	
Kinect2	-0.0242	0.8815	0.0042	0.0055	0.0000	
Intrinsic parameters of depth sensor						
Coefficients	f_{dx}	f_{dy}	u_{d0}	v_{d0}	e_{dx}	e_{dy}
Kinect1	600.13	596.84	301.98	241.65	0.1217	0.1182
Kinect2	401.27	393.84	221.68	186.14	0.2132	0.1981
Coefficients	k_1	k_2	k_3	k_4	k_5	
Kinect1	-0.0904	0.2373	-0.0008	-0.0081	0.0000	
Kinect2	0.2889	-0.3514	-0.0131	-0.0664	0.0000	
Extrinsic parameters (position of RGB camera wrt depth sensor)						
Coefficients	R_x	R_y	R_z	T_x	T_y	T_z
Kinect1	-0.0011	0.0042	-0.0001	23.755	0.1001	-3.9342
Kinect2	0.0289	-0.0942	-0.0157	40.883	-0.6866	-16.368

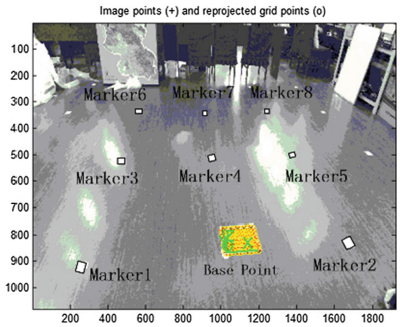


Fig. 12. Trajectory of ground truth.

4.3 Robot SLAM

This paper also testify the accuracy of calibration by reconstructing a scene. The fully calibrated system can be used to obtain a colored point cloud [12] in metric coordinates. For illustration purposes, Fig. 15 shows an example scene and a reconstruction from a different view point. Applying distortion correction leads to a more accurate reconstruction in our experiment. To reconstruct and interact with a 3D environment, we also testified our method by using KinectFusion [13] as shown in Fig. 16.

Table 2. Localization result

Error	Uncalibration (mm)	Uncorrected (mm)	Herrera's method (mm)	Our method (mm)
Mark1	(-39,-71,-41)	(-66,-7,-26)	(-59,-7,-33)	(-45,-12,-3)
Mark2	(-21,-130,-37)	(-82,-23,-60)	(-76,-48,-39)	(-41,-32,-16)
Mark3	(-15,-125,-35)	(-89,-5,-59)	(-81,-19,-33)	(-43,-15,-9)
Mark4	(-37,-110,-51)	(-100,-24,-52)	(-89,12,-36)	(-49,-35,2)
Mark5	(-25,-172,-42)	(-127,-26,-85)	(-139,-54,-50)	(-61,-40,-14)
Mark6	(-37,-156,-62)	(-183,10,-115)	(-81,-30,-26)	(-111,-5,-37)
Mark7	(27,-70,-29)	(1,-4,-14)	(8,-5,-20)	(20,-9,7)
Mark8	(-36,-83,-74)	(-54,39,-68)	(-70,3,-62)	(-17,29,-25)
Mark9	(48,-67,-42)	(-16,51,-66)	(-1,33,-35)	(24,40,-17)
Mark10	(92,-78,-31)	(49,24,-37)	(48,27,-1)	(94,12,15)
Mark11	(-84,-81,-114)	(-160,73,-147)	(-209,17,-118)	(-99,57,-73)
Mark12	(5,-47,-87)	(-81,109,-123)	(-88,89,-90)	(-16,91,-45)

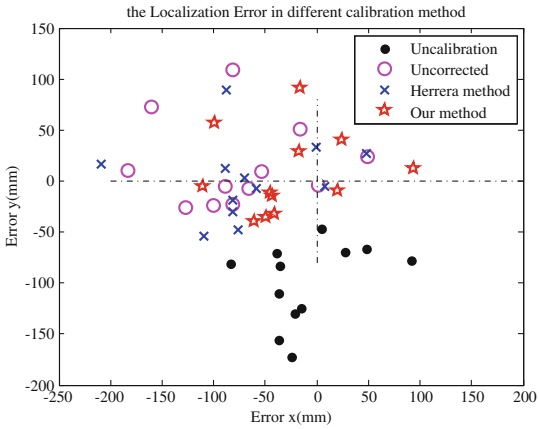


Fig. 13. Localization error in different calibration method.

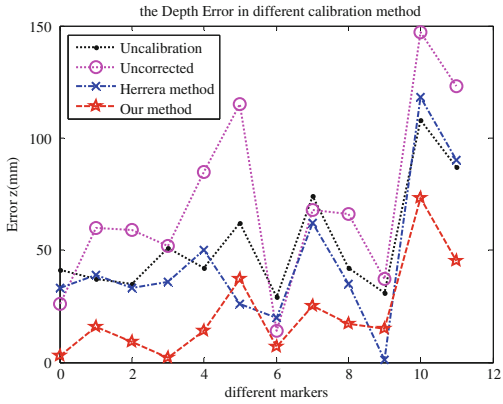
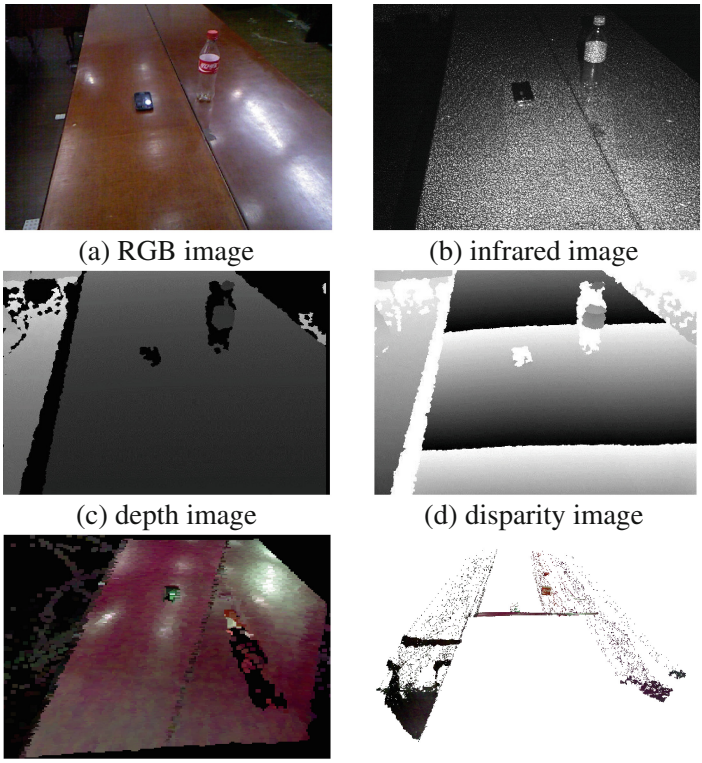


Fig. 14. Depth error in different calibration method.



(a) RGB image (b) infrared image

(c) depth image (d) disparity image

(e) 2D point cloud (f) 3D point cloud

Fig. 15. Sample scene.

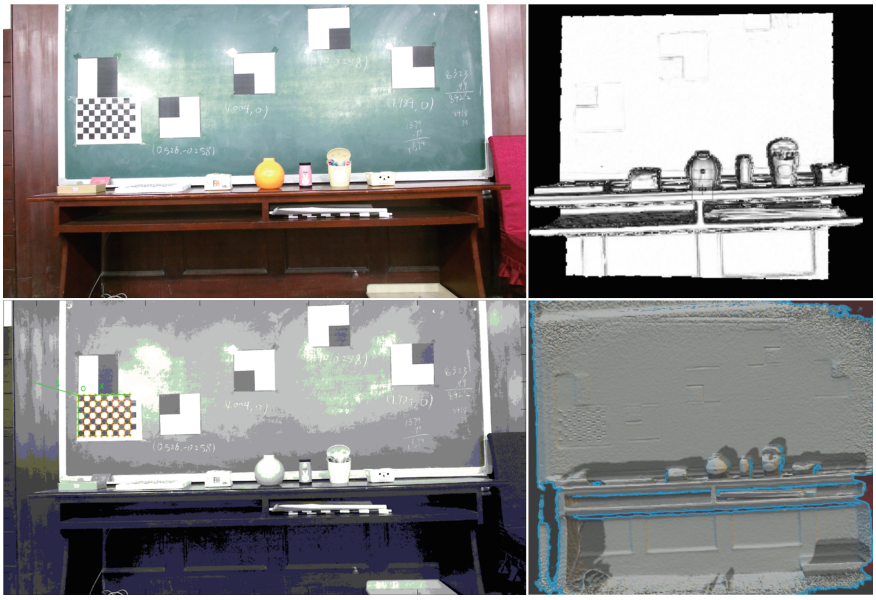


Fig. 16. Robot SLAM.

5 Conclusion

Depth measurement equipment based on structured light or TOF made great progress in recent years, which promoted the progress of depth perception technology such as 3D reconstruction, object detection, human tracking and so on.

In this paper we analyzed the imaging principle and realized the calibration method of RGB-D sensors. What's more, we compared the results with different calibration methods and completed the registration between depth image and color image. The experiments show that our method is able to accomplish better accuracy to [4], using for not only structured light camera but also time-of-flight camera.

Future work will be devoted to: (1) improve the calibration algorithm we presented for different RGB-D sensors in different kinds of surrounding environment, thereby it will be more fast and robust; (2) design a human-robot interactive system based on RGB-D information to replace sensors previously used for indoor service robotics platforms, such as stereo cameras and laser range finders, while increasing the precision and reducing the required computational burden for robotics applications.

Acknowledgements. This work was supported by National Natural Science Foundation of China (Grant No.61375087), National High Technology Research and Development Program (863 Program) of China under Grant 2012AA041403.

References

1. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)
2. Heikkilä, J., Silven, O.: A four-step camera calibration procedure with implicit image correction. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1106–1112 (1997)
3. Herrera C., D., Kannala, J., Heikkilä, J.: Accurate and practical calibration of a depth and color camera pair. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011, Part II. LNCS, vol. 6855, pp. 437–445. Springer, Heidelberg (2011)
4. Herrera C., D., Kannala, J., Heikkilä, J.: Joint depth and color camera calibration with distortion correction. IEEE Trans. Pattern Anal. Mach. Intell. **34**(10), 2058–2064 (2012)
5. Khoshelham, K., Elberink, S.O.: Accuracy and resolution of Kinect depth data for indoor mapping applications. Sensors **12**(2), 1437–1454 (2012)
6. Zhang, C., Zhang, Z.: Calibration between depth and color sensors for commodity depth cameras. In: Proceedings of the Springer International Publishing on Computer Vision and Machine Learning with RGB-D Sensors, pp. 47–64 (2014)
7. Smisek, J., Jancosek, M., Pajdla, T.: 3D with Kinect. In: Proceedings of the Springer London on Consumer Depth Cameras for Computer Vision, pp. 3–25 (2013)
8. Raposo, C., Barreto, J.P., Nunes, U.: Fast and accurate calibration of a Kinect sensor. In: Proceedings of the IEEE International Conference on 3D Vision-3DV, pp. 342–349 (2013)
9. Nakayama, Y., Honda, T., Saito, H.: Accurate camera pose estimation for kinectfusion based on line segment matching by LEHF. In: Proceedings of the IEEE International Conference on Pattern Recognition (ICPR), pp. 2149–2154 (2014)

10. Heikkila, J.: Geometric camera calibration using circular control points. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1066–1077 (2000)
11. Bouguet, J.Y.: Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/
12. Rusu, R.B., Cousins, S.: 3D is here: Point cloud library (PCL). In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 1–4 (2011)
13. Newcombe, R.A., Izadi, S., Hilliges, O., et al.: KinectFusion: real-time dense surface mapping and tracking. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 127–136 (2011)