A Robust Calibration Method for Consumer Grade RGB-D Sensors for Precise Indoor Reconstruction

WALID DARWISH[1,2], **WENBIN LI**[1], **SHENGJUN TANG**[3], **BO WU**[1], **AND WU CHEN**[1]

[1]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
[2]Department of Electronics and Informatics, Vrije Universiteit Brussel, 1050 Brussel, Belgium
[3]Shenzhen Key Laboratory of Spatial Smart Sensing and Services and The Key Laboratory for Geo-Environment Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and GeoInformation, Shenzhen University, Shenzhen 518060, China

Corresponding author: Walid Darwish (w.darwish@connect.polyu.hk)

ABSTRACT RGB-D cameras, which can be attached to any mobile device and work under different operation platforms (e.g., iOS, Android, and Windows), have great potential for indoor 3D modeling and navigation due to their low cost and small size. The main problems of RGB-D cameras for such applications are their range limitations and deteriorated depth accuracy. For example, for a 7-m range, the distance error of structure sensor (one type of RGB-D camera) reaches nearly 0.5 m. We propose a new calibration procedure for RGB-D sensors to improve the depth accuracy. First, the baseline between RGB and IR cameras is calibrated using the direct linear transform method. The distortions of the RGB and IR cameras and the IR projector are then calibrated using the newly proposed two-lens distortion model. Finally, the remaining depth systematic errors are calibrated using an empirical model. Compared to existing calibration methods, the new calibration method considers distortions from both the IR camera and projector and significantly improves the accuracy of far-range depth measurements. The experimental results show that the proposed calibration method can precisely calibrate the full range of the RGB-D sensor, up to 7 m, with an overall depth accuracy of 1.9%, compared to the 5.5% accuracy of the manufacturer's depth estimation. To demonstrate the significance of calibration in indoor mapping, the 3D point cloud of a room (4.5 m x 3.5 m) is generated using the RGB-D SLAM system. The accuracy of the 3D model with the proposed calibration method is approximately 1.5 cm, compared to 7.0 cm using the manufacturer's calibration parameters.

INDEX TERMS RGB-D, structured light, calibration, RGB-D distortion, depth modeling.

I. INTRODUCTION

Since RGB-D sensors were released on the market as advanced game controllers, many research efforts have been made to switch these sensors from gaming to surveying and industrial applications. RGB-D sensors have been used for indoor 3D modeling, navigation, object recognition, and computer vision applications [1]. They are manufactured based on two major concepts: structured light (SL) and time of flight (ToF). The SL concept uses the apparent difference between the IR pattern projected by an IR projector and the IR pattern reflected by an object and received by an IR camera. Kinect v1 [2] and Structure Sensor [3] are examples of SL RGB-D sensors. The ToF concept uses the time difference between the emitted and captured IR patterns to compute the distance between an object and an IR camera based on velocity [4] (i.e., Kinect v2 and Tango [5]). Although RGB-D sensors can produce a real-time 3D colored point cloud for

an observed scene, the working range and the depth precision limitations still curb those sensors to be utilized in surveying applications [6]. Extensive research related to RGB-D sensor depth quality enhancement has been conducted [7]. Most such research has been related to sensor calibration and color-depth registration improvements [8]–[10].

For SL RGB-D sensors, different calibration procedures are proposed. The three-lens based calibration method is introduced to calibrate lenses' geometric parameters and depth systematic error based on the conventional photogrammetric bundle adjustment method [11], [12]. The method adopts Brown's distortion models [13] to compensate for radial and tangential distortion [14], [15]. Furthermore, the depth systematic error is assumed to be a function of radial distortion parameters. The reliability and accuracy of the IR projector calibrated parameters and the initial values to begin bundle adjustment are the basic limitations of this

method [11]. Moreover, the method does not investigate the depth range variation effect on the calibration parameters for both RGB and depth sensors.

Another method to calibrate RGB-D sensors concerns two sensors: the RGB camera and depth sensor, which is combined from the IR camera and the IR projector [9]. This calibration method uses either an empirical distortion model for combined depth sensors [7], [16] or applies the IR camera distortion parameters to the depth image to compensate for the depth distortion [17]. This method is highly affected by the baseline between an IR camera and an IR projector and the depth distortion model.

The calibration of RGB-D sensors includes three main stages: the RGB-IR camera baseline estimation, the distortion model for depth sensors, and the depth error with different operation ranges [18]. The majority of the existing methods use a traditional 2D checkerboard to calibrate the external baseline between the RGB camera and the IR camera by adopting the method proposed by Zhang [19], which has a limitation on the baseline calibration accuracy [20]. The conventional calibration methods usually calibrate a specific depth range to fit a desired application and, in some cases, use external hardware [7] to compute the true depth. This restricts the conventional calibration method to calibrating a certain range for a specific application.

The depth image is derived from the IR pattern received by the IR camera, the pattern emitted by the IR projector, and the manufacturer constants, which are the focal length and baseline between the IR camera and the IR projector. The depth systematic error has attracted great attention to be thoroughly modeled to enhance depth accuracy [21]. The traditional depth error model is presented as a function of the radial distortion parameters of the depth sensor [11]. However, systematic error stems from various sources, such as the baseline between the IR camera and the IR projector, the standard projected depth, and the IR pattern incident angle, rounding off, and the correlation algorithm of disparity.

Darwish *et al.* [22] presented a new model to calibrate the distortion model for depth sensors. The distortion model is defined by six parameters. Two represent the systematic error resulting from the baseline between an IR camera and an IR projector. The remaining four parameters represent the combined distortion effect resulting from an IR camera and an IR projector. The depth error model is assumed to be a function of distortion parameters similar to [11]. The calibration method deals with the fitted plane and ignores systematic depth bias [23]. Furthermore, the calibration method can calibrate only 35% of the sensor working range (i.e., the method calibrates the depth up to 3 m out of 9 m).

Three main factors must be calibrated for the SL RGB-D camera. The first factor is the baseline between RGB and IR cameras, which is used to map the texture information to the depth information. This factor is crucial for pixel correspondences in the RGB-D SLAM system. The second factor is depth distortion and disparity bias, which result from

IR sensor lenses and the IR camera and IR projector baselines, respectively. The third factor is the depth error bias resulting from the remaining error sources [24], such as depth uncertainty, correlation algorithms, and rounding off the disparity, incident angle, and object distance. In this study, we develop a new procedure for calibrating all three factors. We adopt the direct linear transform (DLT) method [25] to calibrate the external baseline between RGB and IR cameras. We use a 3D calibration framework and space resection instead of a 2D calibration framework and homography to improve the accuracy of RGB-IR baseline parameters. We use a two-lens based distortion model [22] to model the depth sensor distortion and calibrate the sensor manufacturer parameters to compensate for the IR camera and IR projector baseline related error. Finally, based on the covariance disparity error propagation concept, we introduce a new depth error model to compensate for the remaining depth error resulting from imaging condition and disparity related errors.

The remainder of this manuscript is organized as follows. Section II shows the depth perception concept for RGB-D sensors based on the SL concept and its related error. Section III presents the proposed two-phase calibration method to thoroughly handle RGB-D camera calibration. Section IV reports the calibration results with three different examples to show the improvements on depth precision and 3D modeling quality, both quantitatively and qualitatively. Finally, Section V provides concluding remarks and describes expected future work.

II. SL RGB-D DEPTH COMPUTATION AND ERROR ANALYSIS

An RGB-D camera combines three sensing lenses, an RGB camera, an IR camera, and an IR projector. Normally, the three lenses are designed to be in a fixed frame (i.e., Kinect [2]) or RGB separated from IR sensors (i.e., Structure Sensor [3]). Figure 1 shows the configurations of the latest RGB-D sensor based on SL.
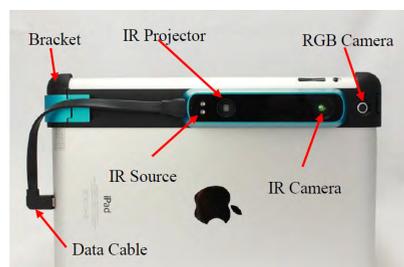

FIGURE 1. Main elements of RGB-D sensors (Structure Sensor [3]).

Structure Sensor uses IR sensors to compute the distance between the IR camera and an observed object. Figure 2 illustrates the manufacturer constants involved in the depth computation. The IR camera and IR projector are separated with a baseline (w), whereas the focal lengths (f) of both the IR camera and projector are the same. A reference pattern for a planar surface captured from a designed distance (Z_0) is stored in the depth sensor's memory.
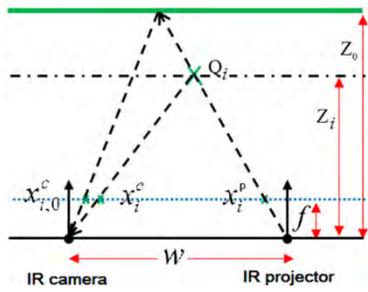
FIGURE 2. Manufacturer constants for RGB-D sensors [18].

Assuming that the firmware stores the standard IR pattern for a feature point (Q_i) as ($x_{i,0}^c$), the current IR projected pattern from the same feature is (x_i^c), the firmware stores (Z_0), (w), and (f) as manufacturer constants. The depth of the feature point (Z_i) can then be computed [12]. The apparent difference between the IR projected and standard patterns is defined as disparity ($d_i = x_i^c - x_{i,0}^c$). As in Figure 2, we find that $x_{i,0}^c = x_i^p + fw/Z_0$, where ($x_i^p$) is the feature location in the IR projector. Substituting this relationship into the (d_i) formula yields the following:

$$d_i = x_i^c - x_i^p - fw/Z_0 \tag{1}$$

Equation (1) can be rewritten using the relationship $x_i^c - x_i^p = -fw/Z_i$ between the IR camera and IR projector feature's locations. The result is $d_i = -\frac{fw}{Z_i} - \frac{fw}{Z_0}$. By rearranging Equation (1), we can determine the general depth computation model:

$$Z_i = \frac{fw}{\frac{fw}{Z_0} + d_i} \tag{2}$$

Equation (2) is the fundamental formula used by the firmware to compute per-pixel depth value. The firmware of the RGB-D sensor does not output the measured disparity (d_i). However, it delivers a normalized rounded-off disparity value from 0 to 2,047 Kinect disparity units (Kdu) as d_i^n, where $d_i = \alpha \left(d_i^n \right) + \beta$ and α and β are two linear factors assigned by the firmware. By substituting the normal disparity formula in Equation (2) and combining all of the constants with the assigned factors (a and b), Equation (2) becomes Equation (3):

$$Z_i = \frac{1}{a + bd_i^n} \tag{3}$$

where a and b are constants and can be expressed as follows:

$$a = \frac{1}{Z_0} + \frac{\beta}{fw} \tag{4a}$$

$$b = \frac{\alpha}{fw} \tag{4b}$$

The final coordinates (X_i, Y_i, and Z_i) of the image point (Q_i) can be computed as follows:

$$X_i = \frac{x_i^c Z_i}{f}; \quad Y_i = \frac{y_i^c Z_i}{f}; \quad Z_i = \frac{1}{a + bd_i^n} \tag{5}$$

The depth produced by RGB-D cameras contains three main error sources. The first type of error results from the deflection of the emitted IR projector ray and the received IR camera ray. It is mainly caused by the distortion of both the IR camera and the IR projector lenses. The second type of error results from the errors of the manufacturer parameters (a and b), which may bias the depth computations. The third type of error is the uncertainty of depth measurement due to the disparity correlation and rounding off the normalized disparity. This error becomes more significant in far-range depth and especially for low-resolution imaging devices, such as Structure Sensor. Figure 3 shows the disparity values corresponding to the measured depth. It should be noted that the disparity value is an integer and varies from 250 Kdu to 1,100 Kdu, covering the entire sensor depth range (0.3 m to 9.0 m). This leads to an important conclusion, which is that the depth image is always formulated from only around a depth value of 850, which may meet with the true value or not, depending on the correlation model of the disparity and rounding-off stage. Figure 3 shows that the far range is critically sensitive to the manufacturer constants (a and b) compared to the close range.
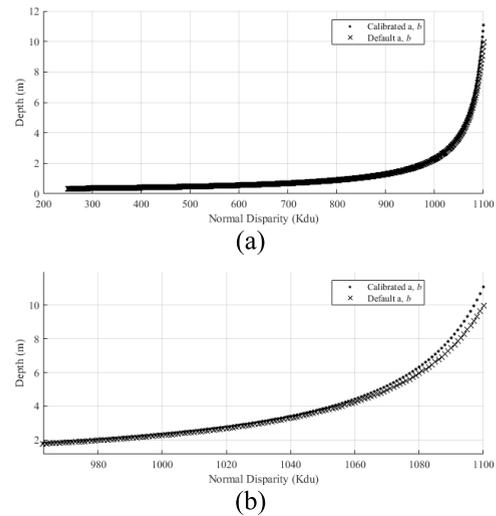

FIGURE 3. Depth and normal disparity relationship, where (a) is the full range of the sensor and (b) is zoomed in for far range.

The error propagation of the depth is related to the normalized disparity. From Equation (3), the depth error can be expressed as follows:

$$\sigma_z^2 = \left(\frac{\partial z}{\partial d} \right)^2 \sigma_d^2 \tag{6}$$

By differentiation in Equation (3) and substitution in Equation (6), the relationship between the error propagated in the calculated depth and measured disparity is a function of squared depth and multiplied by constant b, which is related to the baseline between the projector and camera and the focal length of the depth sensor.

$$\sigma_z = bz^2 \sigma_d \tag{7}$$

From Equation (7), the longer the baseline between the IR camera and projector that the sensor designs, the more accurate the depth that can be obtained with the same focal length.

III. CALIBRATION METHOD AND DEPTH ERROR MODELLING

We propose a three-stage calibration method for SL RGB-D sensors to overcome the drawbacks of traditional calibration methods. First, to improve baseline calibration accuracy between RGB and IR camera calibration, we adopt the DLT method [25] and a 3D chessboard to calibrate the RGB-IR baseline. Second, the distortion model introduced by Darwish *et al.* [22] is used to calibrate the depth distortion. In addition to the depth distortion, the manufacturer constants are also calibrated using full range depth observation. This can overcome the far-range depth bias due to the IR camera and IR projector baseline. Finally, an empirical error model is designed to model the remaining depth error and biases.

A. FIRST CALIBRATION STAGE

The DLT method [25] is adopted for baseline calibration with full distortion parameters [15]. The reason for using DLT over the ordinary photogrammetric bundle adjustment is the data availability for CCD size, due to the absence of initial parameters to begin the bundle solution [26]. Nevertheless, the accuracy of the DLT method is acceptable compared to the ordinary bundle adjustment method. Figure 4 shows the relationship between the image space coordinate system and the object space coordinate system.
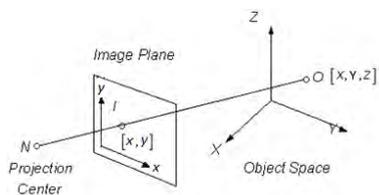

FIGURE 4. Camera coordinate system versus object coordinate system definitions.

The pinhole camera model is a widely applicable model used in close-range photogrammetry [16]. It expresses the relationship between the image point coordinates (x, y) and the corresponding ground point coordinates (X, Y, Z):

$$s \begin{bmatrix} x & y & 1 \end{bmatrix} = \begin{bmatrix} X & Y & Z & 1 \end{bmatrix} \begin{bmatrix} R \\ T \end{bmatrix} [K] \quad (8)$$

where

s	scale factor
x, y	image point coordinates in pixels
X, Y, Z	ground point coordinates

R	3x3 rotation matrix
T	3x1 translation vector, where $T = \begin{bmatrix} dx & dy & dz \end{bmatrix}$
K	3x3 intrinsic matrix, where $K = \begin{bmatrix} f_x & e & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$
f_x, f_y	focal length in pixels
c_x, c_y	coordinate of the principal point in pixels
e	skew between the x and y direction

For consumer grade cameras (e.g., smart phone cameras), the camera lens material and the exact location of the camera lens plane may deviate from the ideal case. This can cause a bending of the line between the focal point and the ground point passing through the image point. The bending error can be modeled by radial and tangential distortion concepts. Both distortion types can be eliminated using Brown's models [19]. Equation (9) describes the radial distortion model, while Equation (10) describes the tangential distortion model.

$$x_d = x \left(1 + r^2 k_1 + r^4 k_2 + r^6 k_3 \right)$$
$$y_d = y \left(1 + r^2 k_1 + r^4 k_2 + r^6 k_3 \right) \quad (9)$$
$$x_d = x + 2y p_1 + p_2 (r^2 + 2x^2)$$
$$y_d = y + 2x p_2 + p_1 (r^2 + 2y^2) \quad (10)$$

where

x_d, y_d	distorted image point coordinates
x, y	coordinates of the free distortion points
k_1, k_2, k_3	radial distortion parameters
p_1, p_2	tangential distortion parameters

To estimate the internal and external parameters of both cameras, the pinhole camera model is solved by the DLT method containing 16 parameters, including the interior and exterior orientations for both cameras and the distortion parameters [15]. Five distortion parameters $\begin{bmatrix} k_1 & k_2 & p_1 & p_2 & k_3 \end{bmatrix}$ represent the full vector of the distortion model. Equation (8) can be rewritten in the format of the bundle adjustment model as follows:

$$x = \frac{X L_1 + Y L_2 + Z L_3 + L_4}{X L_9 + Y L_{10} + Z L_{11} + 1}$$
$$y = \frac{X L_5 + Y L_6 + Z L_7 + L_8}{X L_9 + Y L_{10} + Z L_{11} + 1} \quad (11)$$

where L_i (i = 1, 2, 3...11) is the normalized factor containing the external and internal calibration parameters for the camera. To eliminate the distortion effect from the camera lens, five other parameters (L_{12} to L_{16}) that represent radial and tangential distortion are added to the DLT model. The mathematical relations between the physical parameters illustrated in bundle adjustment and the 16 DLT parameters can be found in [26].

After estimating the internal and external parameters of both the RGB and IR cameras and the two-camera system with a separate baseline, we apply a general cost function, Equation (12), to optimize the external and internal

parameters for the system:

$$min \sum_{n}^{N} \sum_{m}^{M} \left(\left\| p_{mn} - \left(P_{mn} \begin{bmatrix} R_n \\ T_n \end{bmatrix} K \right) \right\|_{color}^2 \right.$$
$$\left. + \left\| p_{mn} - \left(P_{mn} \begin{bmatrix} R_n \\ T_n \end{bmatrix} K \right) \right\|_{IR}^2 \right) \quad (12)$$

where

N	number of captured images
M	number of control points
P_{mn}	ground point coordinates of point m appearing in image n
p_{mn}	image point coordinates of point m appearing in image n
IR	IR camera system
$color$	RGB camera system
K	intrinsic matrix
R_n, T_n	rotation and translation matrix of image n

B. SECOND CALIBRATION STAGE

The second calibration phase is to calibrate the distortion of IR sensor lenses using the two-lens distortion model [18]. Different from the method proposed in [22], we first calibrate the manufacturer parameters based on the full-range depth observation. The distortion models are then used only to model the distortion without considering depth-related systematic errors. The proposed method can mitigate both the radial and tangential distortion effect resulting from both the IR sensor lenses and is baseline independent. The distortion model is derived from Brown's [13] distortion models of both the IR camera and IR projector.

The distortion model has four parameters: W_1, W_2, W_3, and W_4. W_1 and W_2 model the error resulting from the relative orientation between the IR camera and IR projector lenses with respect to the baseline between the IR camera and IR projector. W_3 and W_4 represent the relative ray distortion effect resulting from the combined radial distortion of the IR camera and the IR projector. The distortion model can be presented as follows:

$$dis_i = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{bmatrix}^T \begin{bmatrix} 3Ad_{ti} \\ 2y_i d_{ti} \\ x_i Ad_{ti} \\ x_i Ad_{ti} \left(Ad_{ti} + 2(x_i - d_{ti})^2 + 2y_i^2 \right) \end{bmatrix} \quad (13)$$

where dis_i is the combined IR camera and IR projector distortion error for pixel i. x_i and y_i are the pixel locations in the IR camera space. d_{ti} is the true disparity. A is a term defined as $A = (2x_i - d_{ti})$.

As the true disparity is unknown, Equation (13) can be solved iteratively. We start with d_{ti} as the measured disparity to compute the distortion. We then update d_{ti} using the distortion and measured disparity. This can reach convergence after two or three iterations, as the estimated disparity error is very small compared to the computed disparity (for near range is around 2 Kdu, and for the far range is less than 1 Kdu).

After calibrating the baseline between the RGB and IR cameras, revealing the distortion models of the depth sensor, and calibrating the manufacturer constants (a and b), the RGB and IR sensors are fully geometrically calibrated. However, the geometric sensor calibration does not deal with the imaging conditions and properties of the imaged scene. Thus, an extended calibration depth model is proposed to handle the non-geometric effect of RGB-D cameras. We adopt a polynomial function to calibrate the remaining depth error after correcting the bias from the IR sensor baseline and the IR camera and IR projector distortion effects. The depth error model is proposed as follows:

$$d_{sys} = Ad^3 + Bd^2 + Cd + D \quad (14)$$

where

d_{sys}	systematic depth error remaining after applying the distortion
A, B, C, D	polynomial coefficients
d	undistorted depth

Equation (14) is introduced as per pixel depth error model, therefore the coefficients A, B, C, and D might depend on pixel coordinates.
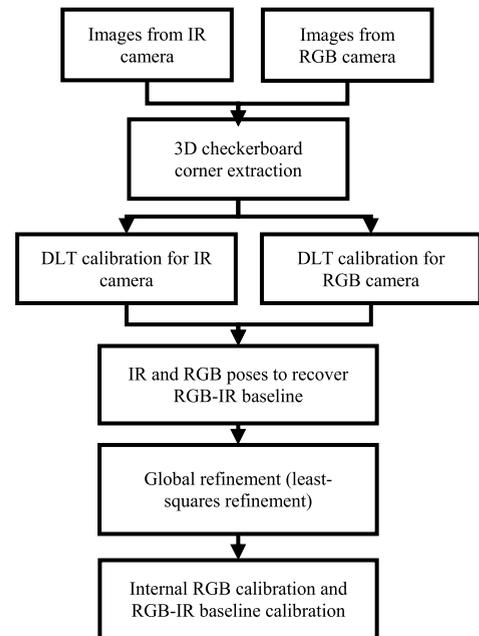

FIGURE 5. Calibration methodology for the RGB-IR camera baseline.

The proposed method (shown in Figure 5) is applied to calibrate the RGB-IR baseline with global optimization stated in Equation (12). We develop an automatic checkerboard corner detection to be applied to a 3D checkerboard, which is shown in Figure 7.

After calibrating the external baseline between the RGB and IR cameras, the second step is proposed (shown in Figure 6) to calibrate the manufacturer constants (a and b) and to model both the distortion and depth systematic error.
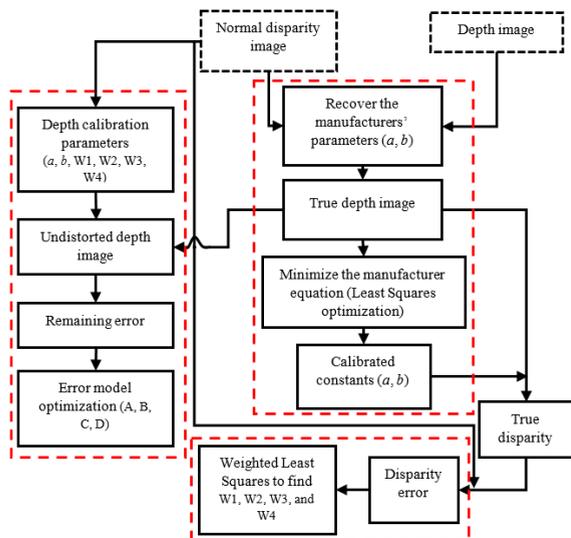
FIGURE 6. Depth sensor calibration methodology divided into three parts, where the middle part is related to manufacturer constant calibration, the lower part is the distortion model estimation, and the left part is the systematic error model.

FIGURE 7. Three-dimensional checkerboard captured by RGB and IR cameras from Structure Sensor.

TABLE 1. Calibration results for three samples of structure sensor RGB-D cameras (S1, S2, AND S3).

Sensor		S1		S2		S3	
		Value	STD	Value	STD	Value	STD
RGB camera internal parameters	F_x	552.690	0.490	559.670	0.850	550.060	0.620
	F_y	551.160	0.460	559.810	0.800	549.040	0.600
	C_x	315.850	0.620	312.770	1.040	321.380	0.640
	C_y	241.640	0.660	239.940	1.130	234.890	0.870
	k_1	0.160	0.007	0.071	0.014	0.149	0.010
	k_2	-0.338	0.051	0.432	0.118	-0.302	0.083
	p_1	0.002	0.001	0.002	0.001	-0.004	0.001
	p_2	0.003	0.001	-0.002	0.001	0.009	0.001
	k_3	-0.005	0.110	-1.877	0.309	-0.286	0.201
IR camera internal parameters	F_x	552.440	0.450	549.670	0.780	552.150	0.590
	F_y	550.850	0.440	549.890	0.760	550.470	0.570
	C_x	316.080	0.550	305.210	0.850	314.440	0.610
	C_y	238.930	0.690	251.690	1.150	233.640	0.850
	k_1	0.058	0.008	0.106	0.014	0.110	0.013
	k_2	-0.577	0.067	-0.930	0.118	-1.090	0.120
	p_1	-0.001	0.000	0.004	0.001	-0.004	0.001
	p_2	0.004	0.000	-0.001	0.001	0.005	0.000
	k_3	1.065	0.165	1.926	0.299	2.456	0.347
RGB-IR baseline	dx	37.502	0.121	33.678	0.223	36.399	0.180
	dy	2.877	0.113	0.877	0.199	3.357	0.192
	dz	21.539	0.464	19.861	0.840	18.812	0.770
	Rx	-0.010	0.001	0.021	0.002	0.005	0.002
	Ry	0.013	0.001	-0.003	0.002	0.001	0.002
	Rz	-0.006	0.000	0.004	0.000	-0.008	0.000

The proposed method uses the distortion model stated in Equation (13) and adopts Equation (3) to calibrate the manufacturer constants. It then uses the error model stated in Equation (14) to model the depth remaining error. Figure 6 illustrates the methodology for the depth sensor calibration. The depth calibration methodology is fully automatic. The inputs for this step are the image pairs of the depth and disparity. The outputs are the final calibration parameters for the depth sensor.

By this stage, the RGB-D sensor is calibrated with respect to two perspectives. The first perspective is the baseline between the RGB camera and the IR camera, which produces a precise set of color and depth information. The second perspective is the depth sensor, which is divided into three parts. The first part is related to the manufacturer constants (*a* and *b*), the second part is related to the distortion effect resulting from the lens material and relative orientation between the lens plane and focal points, and the third part is the calibration of depth error. To examine the calibration method, we apply this calibration methodology to three samples of RGB-D cameras. The following section outlines our experimental designs, results and discussions.

IV. EXPERIMENTS AND RESULTS
A. STRUCTURE SENSOR CALIBRATION
In this experiment, we calibrate three Structure Sensors (S1, S2, and S3) using the proposed method. For the RGB-IR baseline calibration, we use the 3D checkerboard shown in Figure 7. The checkerboard is combined from two nearly perpendicular A3 checkerboards. The ground truth points are measured using a high-grade laser total station. For the depth calibration, we collect depth and disparity data for planar surfaces and adopt the methodology shown in Figure 6.

Table 1 shows the internal parameters of the RGB camera and the RGB-IR baseline calibration results for the three samples. The calibration data are given for each sensor as focal lengths (F_x, F_y) and principal point (C_x, C_y) in pixels with five distortion parameters (k_1 k_2 p_1 p_2 k_3). Furthermore, the external baseline is expressed as three translation components (dx, dy, dz) in mm and three rotation Euler angles (Rx, Ry, Rz) in radians. The focal lengths, principal point, and distortion parameters vary significantly among the examined sensors, which reflects the importance of individually calibrating each sensor.

Table 2 illustrates the depth sensor calibration parameters, which are divided into two sets: the calibrated manufacturer parameters (*a* and *b*) and the distortion parameters (W_1, W_2, W_3, and W_4). The depth error model parameters for the depth error model are per-pixel values. Thus, each coefficient is

TABLE 2. Depth calibration parameters.

Sensor		S1	S2	S3
Parameter	*a*	-3.3916E-06	-3.3882E-06	-3.3880E-06
	b	3.8296E-03	3.8267E-03	3.8266E-03
	W1	-2.8368E-07	6.5743E-07	4.4216E-07
	W2	-3.8742E-07	2.4282E-07	2.6113E-07
	W3	2.6348E-08	1.1180E-09	2.1900E-08
	W4	-3.7213E-14	-1.8772E-14	-5.7055E-14
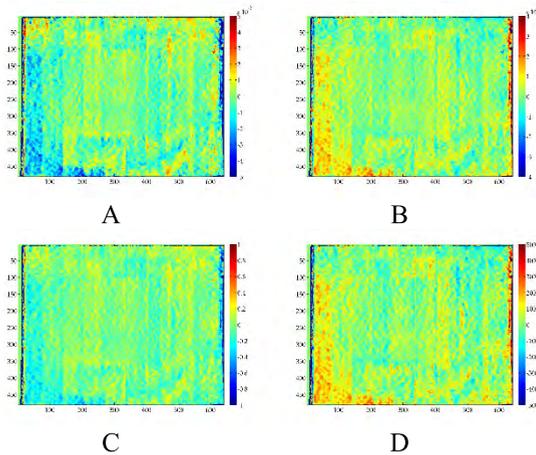
FIGURE 8. Calibrated coefficients (A, B, C, and D) of the depth error model for the S1 sensor.

FIGURE 9. Depth precision performance of the S1 sensor.
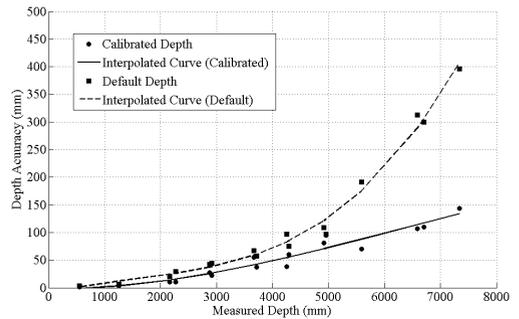
FIGURE 10. Calibration depth performance versus default depth.

represented by an image. Figure 8 shows the four parameters for the S1 sensor as an example. Tables 1 and 2 show that even with the same sensor type, the calibration parameters vary significantly. This reflects the importance of the calibration step before adopting this kind of sensor in surveying applications (e.g., 3D modeling). Figure 8 shows that the artifacts are detected and consumed from the observed data. Ten zones are vertically distributed around the image center.

To check our calibration performance, we acquire depth images for a planar surface from different distances varying from 0.5 m to 2.0 m. The performance is quantified based on the same procedure proposed by [7], [11], and [18], which is the RMSE of the fitted plane. Figure 9 compares the depth calibration performance of our proposed method to both the manufacturer calibration (default calibration) and the method indicated in [18]. The data shown in the figure include the default depth, the undistorted depth based on [18] calibration method, and the modeled depth using the calibration procedure illustrated in this research.

The depth precision before applying the error model exceeded 20 mm for a 2-m range. After applying the combined error model and the calibration for the depth sensor, the error decreased to 6 mm, signifying 70% and 50% improvement compared to default and [18] calibration methods, respectively.

Considering the full range calibration, the experiment is extended to the full depth range of the Structure Sensor
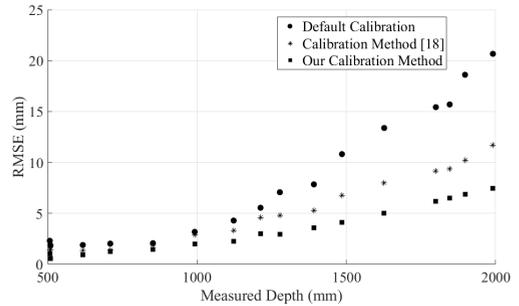
(approximately 8 m). The calibration method indicated in [18] does not give a reliable results for far range measurements. Therefore, the far depth performance is not reported based on that method. Figure 10 shows the depth error performance for the calibrated and default depth. The calibration method can cover the entire depth range and restrains the depth accuracy to 1.9% compared to 5.5% for the default depth.

Figure 11 shows a constructed point cloud using our proposed method calibration parameters and the manufacturer calibration parameters. The sensor is placed 9.0 m away from a wall.
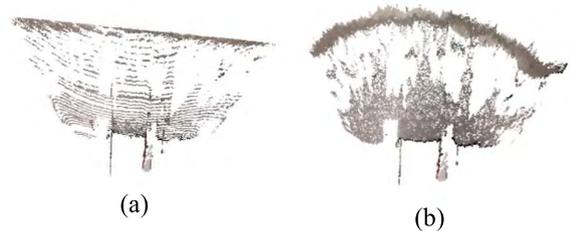

FIGURE 11. Calibration effect on point cloud results, where (a) is the point cloud after calibration and (b) is the point cloud before calibration.

B. 3D INDOOR RECONSTRUCTION

To check the proposed calibration method, we collect a set of images illustrating a room with dimensions of 4.5 m × 3.5 m. The data are processed using the visual RGB-D SLAM algorithm [27]–[30] to form the 3D point
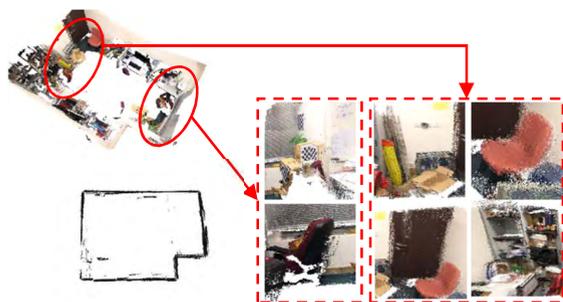
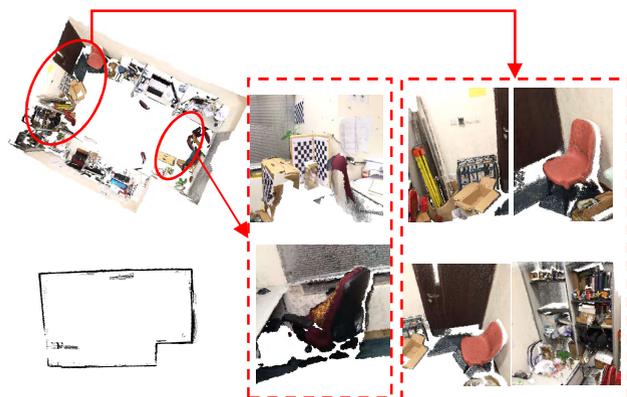**FIGURE 12.** Visual RGB-D SLAM reconstruction with default data.

FIGURE 13. Visual RGB-D SLAM reconstruction with calibrated data using method indicated in [18].

cloud model. Qualitative and quantitative analyses for the reconstructed model before and after calibration are applied. The qualitative analysis is based on the shape of the room, the wall noise, and the extracted details. To carry out the quantitative analysis, we use a total station to measure 11 preset distances for the room and then compare the RGB-D results before and after calibration. Figures 12, 13, and 14 show the difference between the default model, the calibrated model using [18], and the calibrated model using our method, respectively.

The calibrated model using the proposed method can easily detect the chair edges, the boundary of the door, and the edges of the bookshelf storage unit. Furthermore, the projected wall to floor plane is more precise compared to the model based on calibration procedure indicated in [18]. However, the default calibrated data can hardly detect the objects' boundaries with large noise and fewer details.

Table 3 shows the quantitative analysis for the calibrated data based on our method, the calibrated data according to [18], and the default data regarding the 11 measured distances. The differences and relative errors before and after calibration are shown in Table 3.

The quantitative results show a significant improvement in both absolute error and standard deviation. The calibrated model has a precision of ±1.5 cm compared to ±4.4 cm for the method indicated in [18] and ±7 cm for the default model. The relative error of the default calibration method is 3.94%, which decreases to 1.82% and 0.77% after

calibration using [18] method and our proposed method, respectively.

The experiments show a significant improvement in indoor modeling. The main contributions are achieved by accurately applying the external parameters for the precise mapping of the image point from color to depth, as more than a one-pixel error in alignment can cause a large drift in the RGB-D SLAM system. Furthermore, depth precision can reduce the number of inliers used to estimate the camera pose. After applying calibration for depth information and color camera distortion, we can achieve centimeter accuracy for room dimensions compared to real measurements.

C. DISCUSSIONS

1) CALIBRATION RESULTS

Three different RGB-D sensors were calibrated using our calibration method. However the sensors belong to the same manufacturer company, the calibration parameters include geometric parameters (table 1) and depth calibration parameters (table 2), vary among the sensors. Adopting the DLT method can achieve half pixel accuracy compared to two-pixels accuracy for the method mentioned in [18], This is because of the precise estimation of the camera pose during the DLT calibration process.

The proposed calibration method individually calibrates the systematic depth error and the lens distortion effect, this can significantly help to remove the artifacts from RGB-D cameras. The current calibration methods couple the depth distortion and depth systematic error models, thus, they cannot mitigate the artifacts [11] and cannot calibrate the far range [7], [18].

The proposed method considerably estimates the shape of the captured scene, as it can be seen from figure 11. Also, figure 11 is consonant with the calibrated depth performance in figure 10, as the method estimates the wall with a 15cm accuracy, compared with a 50cm wall from the default calibrated point cloud.
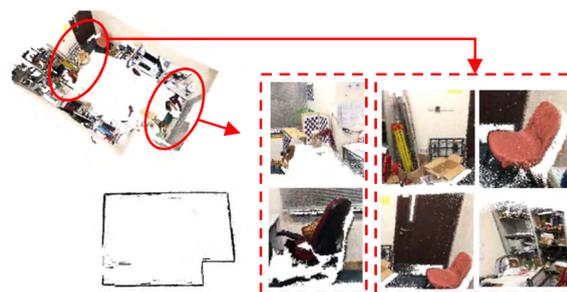

FIGURE 14. Visual RGB-D SLAM reconstruction with calibrated data.

2) 3D INDOOR RECONSTRUCTION

Figures 12, 13, and 14 show the performance of the reconstructed 3D model using the default calibration method, the calibration method discussed in [18], and our calibration method, respectively. It should be noted that the calibrated

TABLE 3. Comparison between calibrated and default data for room reconstruction (Meters).

Check distances	Distances measured by a total station	Distance from calibrated data (our method)	Error	Relative E (%)	Distance from calibrated data (method in [18])	Error	Relative E (%)	Distance from default data	Error	Relative E (%)
d1	0.800	0.807	0.007	0.867	0.830	0.030	3.750	0.767	-0.033	-4.329
d2	0.802	0.792	-0.011	-1.346	0.784	-0.018	-2.244	0.835	0.033	3.968
d3	0.947	0.942	-0.005	-0.520	0.983	0.036	3.801	0.841	-0.106	-12.56
d4	1.110	1.113	0.003	0.298	1.119	0.009	0.811	1.076	-0.034	-3.145
d5	2.337	2.299	-0.039	-1.675	2.332	-0.005	-0.214	2.173	-0.165	-7.571
d6	2.560	2.564	0.004	0.159	2.607	0.047	1.836	2.441	-0.119	-4.888
d7	3.067	3.071	0.004	0.131	3.144	0.077	2.511	2.909	-0.158	-5.434
d8	3.360	3.356	-0.004	-0.131	3.465	0.105	3.125	3.291	-0.069	-2.102
d9	3.402	3.423	0.02	0.597	3.497	0.095	2.792	3.262	-0.141	-4.317
d10	3.855	3.858	0.003	0.079	3.952	0.097	2.516	3.720	-0.135	-3.638
d11	4.670	4.672	0.002	0.043	4.752	0.082	1.756	4.478	-0.192	-4.292
Mean	--	--	-0.001	-0.136	--	0.051	1.858	--	-0.102	-4.392
RMSE	--	--	0.015	0.771	--	0.044	1.815	--	0.068	3.942

model, based on our calibration method, has less noise compared with the model based on the calibration method shown in [18]. Moreover, the projected wall, using [18] calibrated data, is curved compared to both default and our calibrated models. it is reasonable as the depth calibration model of [18] is computed based on near depth.

Table 3 shows a significant improvement in relative and absolute error comparing our calibrated model with the default model and the calibrated model based on [18]. This improvement in relative error results from two major parts. The first is the correct registration between the RGB image and the depth image, which is a critical issue for the RGB-D SLAM algorithm. The second is the depth precision, as the SLAM system can assign inlier points to be outliers not only due to false matching but also if the depth precision is higher than the preset threshold.

V. CONCLUSIONS AND FUTURE WORK

RGB-D sensors have great potential to replace expensive high-grade laser scanners in the mapping of indoor spaces. The calibration method including the modeling of RGB-D camera depth plays a great role in enhancing sensor-mapping ability. Diverse errors affect the depth precision of RGB-D cameras, such as incident angle, IR projector and IR camera baseline, disparity uncertainty, and the distortion of IR cameras and projectors. Most existing calibration methods combine all of these errors into one empirical model, yielding low-accuracy calibration. We propose a calibration method to calibrate each error source effect separately, thereby significantly improving the depth accuracy. For the color depth registration problem, the proposed method adopts the DLT method to calibrate the baseline between the RGB camera and the IR camera.

The calibration method is used to calibrate three sets of Structure Sensor RGB-D cameras. The calibration parameters of each sensor are significantly different, which reflects the importance of individually calibrating each sensor. We compare our calibration results to those of

the manufacturer and lately introduced calibration in [18]. We show that the proposed calibration method improves depth accuracy by 70% for the whole working range of Structure Sensor. Implementing the proposed calibration method in the RGB-D SLAM system significantly improves the 3D modeling error from 7 cm (default calibration) to 1.5 cm. Moreover, the calibrated model shows a significant improvement on the model quality represented by the object's boundary and color.

REFERENCES

[1] M. Camplani, T. Mantecon, and L. Salgado, "Depth-color fusion strategy for 3-d scene modeling with Kinect," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1560–1571, Dec. 2013.

[2] Microsoft. *Kinect*. Accessed: Jan. 15, 2016. [Online]. Available: https://dev.windows.com/en-us/kinect

[3] *Occipital*. Accessed: Nov. 22, 2015. [Online]. Available: http://structure.io/

[4] J. Jung, J. Y. Lee, Y. Jeong, and I. S. Kweon, "Time-of-flight sensor calibration for a color and depth camera pair," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1501–1513, Jul. 2015.

[5] Google. *Tango*. Accessed: Nov. 10, 2015. [Online]. Available: http://get.google.com/tango/

[6] W. Darwish, W. Li, S. Tang, and W. Chen, "Coarse to fine global RGB-D frames registration for precise indoor 3D model reconstruction," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, Jun. 2017, pp. 1–5.

[7] D. H. C, J. Kannala, and J. Heikkilä, "Joint depth and color camera calibration with distortion correction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2058–2064, Oct. 2012.

[8] C. Zhang and Z. Zhang, "Calibration between depth and color sensors for commodity depth cameras," in *Computer Vision and Machine Learning with RGB-D Sensors*. Cham, Switzerland: Springer, 2014, pp. 47–64.

[9] A. Perez-Yus, E. Fernandez-Moral, G. Lopez-Nicolas, J. J. Guerrero, and P. Rives, "Extrinsic calibration of multiple RGB-D cameras from line observations," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 273–280, Jan. 2018.

[10] X. Song, J. Zheng, F. Zhong, and X. Qin, "Modeling deviations of RGB-D cameras for accurate depth map and color image registration," *Multimedia Tools Appl.*, vol. 77, no. 12, pp. 14951–14977, 2018.

[11] J. C. K. Chow and D. D. Lichti, "Photogrammetric bundle adjustment with self-calibration of the PrimeSense 3D camera technology: Microsoft Kinect," *IEEE Access*, vol. 1, pp. 465–474, 2013.

[12] H. Yamazoe, H. Habe, I. Mitsugami, and Y. Yagi, "Easy depth sensor calibration," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 465–468.

[13] J. G. Fryer and D. C. Brown, "Lens distortion for close-range photogrammetry," *Photogramm. Eng. Remote Sens.*, vol. 52, pp. 51–58, Jan. 1986.

[14] J. Heikkilä and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 1106–1112.

[15] J. G. Fryer, "Camera calibration in non-topographic photogrammetry," in *Non-Topographic Photogrammetry*, H. M. Karara, Ed. Falls Church, VA, USA: American Society for Photogrammetry and Remote Sensing, 1989, pp. 59–70.

[16] C. Raposo, J. P. Barreto, and U. Nunes, "Fast and accurate calibration of a Kinect sensor," in *Proc. Int. Conf. 3D Vis. (3DV)*, Jun./Jul. 2013, pp. 342–349.

[17] R. Macknojia, A. Chávez-Aragón, P. Payeur, and R. Laganière, "Calibration of a network of Kinect sensors for robotic inspection over a large workspace," in *Proc. IEEE Workshop Robot Vis. (WORV)*, Jan. 2013, pp. 184–190.

[18] W. Darwish, S. Tang, W. Li, and W. Chen, "A new calibration method for commercial RGB-D sensors," *Sensors*, vol. 17, no. 6, p. 1204, 2017.

[19] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[20] K. Khoshelham, D. R. Dos Santos, and G. Vosselman, "Generation and weighting of 3D point correspondences for improved registration of RGB-D data," *ISPRS Ann. Photogram. Remote Sens. Spatial Inf. Sci.*, vol. II-5/W2, pp. 127–132, Oct. 2013.

[21] T. Mallick, P. P. Das, and A. K. Majumdar, "Characterizations of noise in Kinect depth images: A review," *IEEE Sensors*, vol. 14, no. 6, pp. 1731–1740, Jun. 2014.

[22] W. Darwish, W. Chen, S. Tang, and W. Li, "Full parameter calibration for low cost depth sensors," presented at the Melaha Int. Conf. Exhib., Cairo, Egypt, 2016.

[23] D. Pagliari and L. Pinto, "Calibration of Kinect for Xbox one and comparison between the two generations of Microsoft sensors," *Sensors*, vol. 15, no. 11, pp. 27569–27589, 2015.

[24] K. Khoshelham, "Accuracy analysis of Kinect depth data," in *ISPRS Workshop Laser Scanning*, vol. 38, no. 5, pp. 133–138, Sep. 2011.

[25] Y. I. Abdel-Aziz, H. M. Karara, and M. Hauck, "Direct linear transformation from comparator coordinates into object space in close-range photogrammetry," in *Proc. Symp. Close-Range Photogram.*, Falls Church, VA, USA, 1971, pp. 1–18.

[26] J. C. McGlone, "Analytic data-reduction schemes in non-topographic photogrammetry," in *Non-Topographic Photogrammetry*, 2nd ed., H. M. Karara, Ed. Falls Church, VA, USA: American Society for Photogrammetry and Remote Sensing, vol. 1989, pp. 37–55.

[27] I. Dryanovski, R. G. Valenti, and J. Xiao, "Fast visual odometry and mapping from RGB-D data," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2013, pp. 2305–2310.

[28] G. Hu, S. Huang, L. Zhao, A. Alempijevic, and G. Dissanayake, "A robust RGB-D SLAM algorithm," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 1714–1719.

[29] T. Whelan, H. Johannsson, M. Kaess, J. J. Leonard, and J. McDonald, "Robust real-time visual odometry for dense RGB-D mapping," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2013, pp. 5724–5731.

[30] T. Whelan, M. Kaess, H. Johannsson, M. Fallon, J. J. Leonard, and J. McDonald, "Real-time large-scale dense RGB-D SLAM with volumetric fusion," *Int. J. Robot. Res.*, vol. 34, pp. 598–626, Apr. 2015.

Authors' photograph and biography not available at the time of publication.

• • •