

Article

A Range-Independent Disparity-Based Calibration Model for Structured Light Pattern-Based RGBD Sensor

Wenbin Li ^{1,2,*}, Yaxin Li ² , Walid Darwish ^{3,4} , Shengjun Tang ⁵, Yuling Hu ² and Wu Chen ¹

¹ Shenzhen Research Institute, The Hong Kong Polytechnic University, Shenzhen 518057, China; wu.chen@polyu.edu.hk

² Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China; yaxin.pu.li@connect.polyu.hk (Y.L.); lindsay444@outlook.com (Y.H.)

³ Geomatics Engineering Lab, Civil Engineering Department, Faculty of Engineering, Cairo University, Cairo 12613, Egypt; dawalid@etrovub.be

⁴ Department of Electronic and Informatics, Faculty of Engineering, Vrije Universiteit Brussel, 1050 Brussels, Belgium

⁵ Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services & Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) & Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518050, China; shengjuntang@szu.edu.cn

* Correspondence: wb.li@polyu.edu.hk

Received: 2 December 2019; Accepted: 21 January 2020; Published: 23 January 2020



Abstract: Consumer-grade RGBD sensors that provide both colour and depth information have many potential applications, such as robotics control, localization, and mapping, due to their low cost and simple operation. However, the depth measurement provided by consumer-grade RGBD sensors is still inadequate for many high-precision applications, such as rich 3D reconstruction, accurate object recognition and precise localization, due to the fact that the systematic errors of RGB sensors increase exponentially with the ranging distance. Most existing calibration models for depth measurement must be carried out with different distances. In this paper, we reveal the mechanism of how an infrared (IR) camera and IR projector contribute to the overall non-centrosymmetric distortion of a structured light pattern-based RGBD sensor. Then, a new two-step calibration method for RGBD sensors based on the disparity measurement is proposed, which is range-independent and has full frame coverage. Three independent calibration models are used for the calibration for the three main components of the RGBD sensor errors: the infrared camera distortion, the infrared projection distortion, and the infrared cone-caused bias. Experiments show the proposed calibration method can provide precise calibration results in full-range and full-frame coverage of depth measurement. The offset in the edge area of long-range depth (8 m) is reduced from 86 cm to 30 cm, and the relative error is reduced from 11% to 3% of the range distance. Overall, at far range the proposed calibration method can improve the depth accuracy by 70% in the central region of depth frame and 65% in the edge region.

Keywords: RGBD sensor; calibration model; disparity

1. Introduction

The consumer-grade RGBD sensor has the potential to be used in many fields, such as 3D reconstruction [1,2], camera simultaneous localization and mapping (SLAM) [3], robotics exploration [4], and obstacle avoidance [5], due to its low cost and ability to provide depth with pixel-corresponded

RGB information. Based on the ranging principle, consumer-grade RGBD sensors can be categorized into three groups: time-of-flight (ToF)-based RGBD sensors (i.e., Kinect V2 [6]), structured light pattern (SLP)-based RGBD sensors (i.e., Kinect V1 [6], Structure Sensor [7], Asus Xtion [8], Intel RealSense [9]), and stereo vision-based RGBD sensors (i.e., ZED camera [10]). Although all RGBD sensors provide the same data format, SLP RGBD has become one of the most popular RGBD solutions due to its low requirement for computational power [11]. Despite the wide use and popularity, the SLP-based RGBD sensor has one significant drawback: the errors in depth frame exponentially increase with ranging distance [6]. Existing applications of consumer-grade SLP RGBD sensor are mainly for gaming purposes, not for those requiring high-precision, such as SLAM and visual odometry [12]. With high-precision applications, the accuracy of depth measurement is crucial as errors in depth will be accumulated during the frame matching process, which will significantly affect the quality of final 3D point cloud and position products and may cause a frequent loss tracking problem. Thus, proper calibration of depth measurements is required for SLP RGBD sensors [13].

The SLP RGBD sensor calculates ranging information based on the disparity d_i (pixel difference, $d_i = |x_i - x_0|$) between the reference pixel x_0 location of the pre-structured pattern X_p in the reference plane Π_0 and the measure pixel location x_i of captured pattern X_i on the object plane Π_{obj} . in the received image frame, as shown in Figure 1a. The reference pixel x_0 of the pre-structured pattern is fixed in the image plane Π_c . The pre-structured pattern as shown in Figure 1b, is a fixed pattern of light and dark speckles. The depth is calculated by triangulation against the pre-structured pattern. If the captured feature x_i can be matched to the reference pixel x_0 , the disparity d_i can be obtained. For a commercial SLP RGBD sensor, information including the actual pattern x_p in the projector plane Π_p and the distance Z_0 from the IR projector to the reference plane Π_0 remains undisclosed.

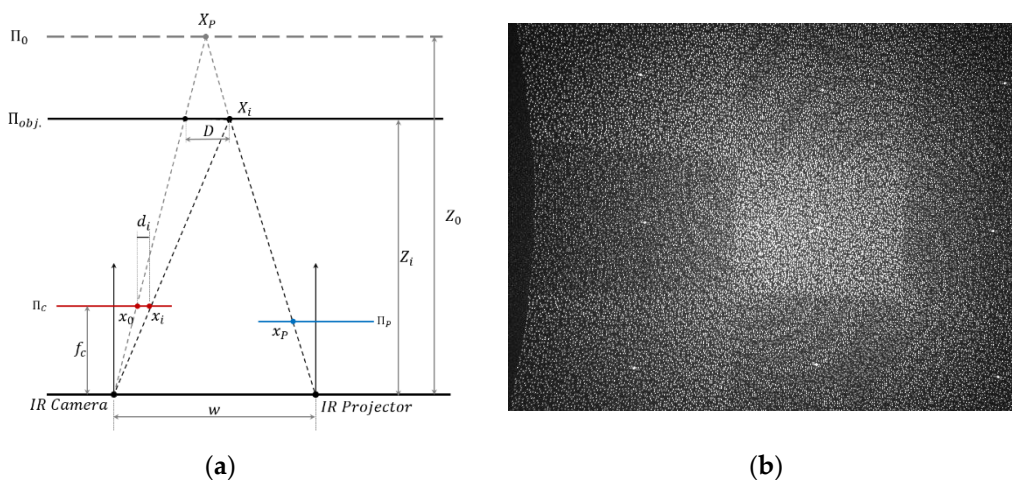


Figure 1. (a) Projection model of structured light depth sensor; (b) example of pre-structured pattern [14].

For the object X_i , the depth Z_i can be expressed as:

$$Z_i = \frac{1}{\frac{1}{Z_0} + \frac{1}{f_c w} d_i} \quad (1)$$

where,

Z_0 is the reference distance from the sensor to the reference plane Π_0 ;

f_c is the focal length of IR camera;

w is the baseline between the IR camera and IR projector;

d_i is the disparity for x_i .

For commercial RGBD sensors, such as Structure Sensor and Kinect V1, the actual disparity d_i in pixel unit is not available to the users. Instead, a normalized disparity d'_i value from 0 to 2047 is outputted. Normalized disparity is expressed as Equation (2), m , n are the normalized parameters that remain unknown:

$$d'_i = m \cdot d_i + n \quad (2)$$

Combine (1) and (2):

$$Z_i = \frac{1}{\frac{1}{Z_o} - \frac{n}{f_c w m} + \frac{1}{f_c w m} d'_i} = \frac{1}{\alpha + \beta d'_i} \quad (3)$$

where: $\alpha = \frac{1}{Z_o} - \frac{n}{f_c w m}$, $\beta = \frac{1}{f_c w m}$, α & β are fixed parameters of a RGBD sensor, and are provided by the manufacturer to calculate the depth Z_i from the normalized disparity d'_i .

Early works on SLP RGBD calibration [6,15] described the systematic error in the depth frame as a function of radial distortion parameters and adopted the classic RGB camera distortion model [16] to correct the radially symmetric error on the SLP-based RGBD sensor. However, they failed to address the fact that the distortion increases with distance. Zhang and Zhang [17] proposed a depth dependent calibration model that separately fits a linear equation of the actual depth and the rotation angle for each pixel. The limitations for this method are the instability of the algorithm, the requirement of accurate initial parameters to avoid divergence, and the high computational cost. Similarly, Canessa et al. [18] proposed a labour intense empirical calibration model by using the sample images captured from 0.6 to 2 m to fit a second degree polynomial model for each pixel. Recently, Darwish et al. [19] proposed a calibration model, treating the depth distortion of a RGBD sensor as a combined centrosymmetric error of the IR camera distortion and IR projector distortion with same optical centre. A look up table of calibration parameters was calculated at different distance, i.e., from 0.5 to 3 m with 0.5 m interval. With insufficient investigation of the mechanism of how the IR camera and IR projector contribute to the overall non-centrosymmetric distortion of the RGBD sensor, and also being a distance-based calibration model, this method suffers large offsets in the edge area of depth frame and requires a look up table for parameters in different distance. Similar problems can be found in [20].

A visual camera or a RGB camera often have a centrosymmetric or near centrosymmetric distortion pattern. When the distortion model is adopted in RGBD sensor calibration, the assumption that the distortion of SLP RGBD is also centrosymmetric and can be modeled by a classic distortion model is accepted by many researchers [15,20–22]. However, compared to visual distortion that can use one model to represent the distortion, current calibration models for SLP RGBD sensors are much more complex due to the additional work such as fitting the model separately based on distance, the requirement of good initial parameters, and the larger number of parameters to provide full range coverage. The complexity of current calibration models as discussed later in this paper is mainly due to the negligence of the fundamental difference between the visual camera and the SLP RGBD sensor. One significant difference between visual camera and SLP RGBD sensor is that the distortion in the depth frame of the RGBD sensor is caused by both camera lens and projector lens while image distortion is only caused by the camera lens. In fact, due to the combined effects of camera and projector, the depth distortion is non-centrosymmetric rather than the centrosymmetric distortion in the visual camera. Therefore, the existing methods that only use a centrosymmetric distortion model to calibrate the depth distortion, for example the method of Darwish et al. [19], will inevitably suffer large offset in the frame edge area. The edge area is extremely sensitive to the high order part in the visual camera distortion model [16]. Furthermore, the target measurement of SLP RGBD sensors to apply the distortion model should be disparity rather than depth, as disparity is the raw measurements of distance measurement. A camera distortion model was originally designed for calibrating the distorted pixel caused by camera lens. Although a similar distortion pattern can be discovered in depth frames, it is actually a reciprocal product of the pixel difference (disparity) between the captured pattern and reference pattern. By analyzing the distortions of the camera and the projector separately, this paper proposes a new distortion model to calibrate disparity which can be applied to full range

depth frames of SLP-based RGBD sensors. Compared to the current methods, the proposed method is distance independent.

This paper firstly makes a comprehensive discussion of the sources and forms a mechanism of depth distortion of the SLP RGBD sensors. By compositing the calibration model that is strictly based on actual physical model of camera distortion and projector distortion, we propose a new method which can effectively and accurately calibrate the full range depth data. The proposed method calibrates the whole range depth with just one model and one set of parameters. Compared to the existing model that requires to separately fit calibration parameters for different distance range or a look up table of parameters, the proposed method largely improved the efficiency of the calibration procedure. By applying the proposed calibration model, we can accurately improve the ranging accuracy by 70% in the central area and 65% in the edge area of the depth frame. Two distortions and one infrared cone related bias will be used to model the overall systematic error of SLP sensors. The two distortions are a pincushion distortion caused by the projector lens and a barrel distortion caused by infrared camera lens, shown in Figure 2. A bias caused by the infrared cone that cannot illuminate the homogeneously for a unified pattern recognition performance [6,23] is also modeled in the proposed calibration method. As the systematic error is extremely sensitive to how these two distortions and the bias overlapped, accurately determining the forming mechanism of the overall distortion is crucial for calibration.

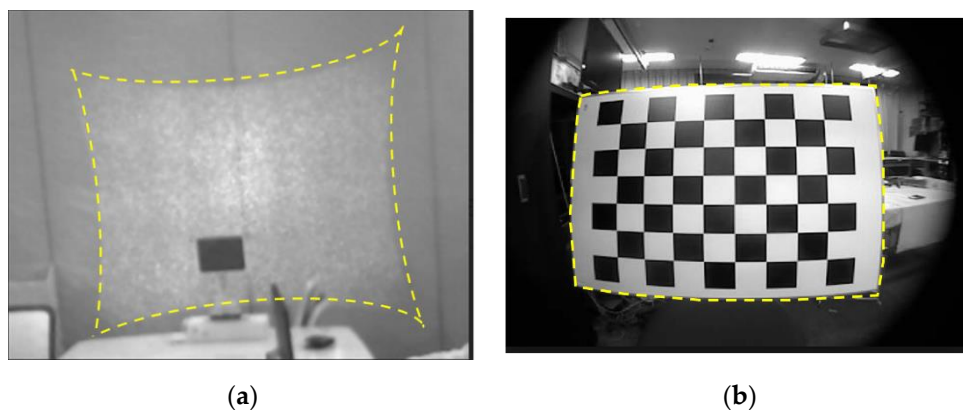


Figure 2. (a) Pincushion distortion of infrared (IR) projector; (b) barrel distortion of IR camera.

Then a two-step calibration procedure is proposed to overcome the problem that the commuter-grade RGBD sensor often remains a black box to the users. SLP RGBD sensors such as Kinect V1 [24] and Structure Sensor [7] only provide a normalized disparity unit, but the raw disparity measurement and the normalization parameters for disparity remain unknown. Other information, such as original reference pattern and the actual measurement of the captured pattern, is also unavailable to users. Such black box environment is extremely problematic for the sensor calibration. Although the camera distortion and camera internal parameters can be calibrated by traditional checkboard method [25], the projector distortion cannot be easily calibrated with existing methods [26–28] which all require knowing the original reference pattern and take measurement of the casted pattern in real world. The two-step calibration procedure is proposed to achieve a precise calibration procedure in such black box environment for RGBD sensor. The two-step calibration procedure includes: (1) the IR camera related distortion is calibrated with classic checkboard method; (2) the projector distortion in normalized disparity is modeled by a new combined objective function.

The calibration method is evaluated by two experiments in an indoor environment. One is an experiment designed to test the performance of the calibration model. A set of test data with ground truth error is acquired in three different scenes: (1) flat surface, (2) non-flat surface, and (3) flat surface with distance variation. The ground truth error is obtained by plane fitting and compared with the modeled error generated from the proposed method. Comparison results demonstrate

that the proposed method can precisely model the systematic error caused by non-centrosymmetric distortion of the RGBD sensor. The other one is an evaluation of proposed method in actual point cloud collected in the indoor environment. By comparing the calibrated and uncalibrated point cloud in the whole working range of the tested SLP sensor, the improvement of the calibration model can be demonstrated clearly.

2. RGBD Distortion Model

To calibrate the SLP based RGBD sensor, it is important to understand how the error is introduced in the RGBD sensor and how that error affects the depth measurements. The radical systematic error appeared in the depth frame is mainly a combined result of the distortions caused by camera lens and projector lens. The infrared cone also causes a certain radical error to the depth frame. In this section, the mechanism of the non-centrosymmetric distortion of RGBD sensor will be revealed.

The distortion appearing in the disparity measurement of SLP RGBD sensor can be traced back to the two optical lenses in its hardware, the projector lens and camera lens. The projector lens causes a pincushion distortion as shown in Figure 2a; and the camera lens causes a barrel distortion as shown in Figure 2b. As the pattern is casted in IR and can't be captured by normal digital camera, Figure 2a is captured by a calibrated IR camera called MIQUS M3 [29]. Figure 2b is captured by the IR camera. Both pincushion and barrel distortion are optical distortions, therefore, they occur as a result of optical design and lens error. They both appear and overlap in the disparity measurement of a SLP RGBD sensor.

The data accessed by users are the normalized disparity unit, which is a normalized pixel difference between the pre-stored reference pattern and the captured pattern. The most basic observation of SLP RGBD sensor is the disparity d_i between the pre-stored distortion-free reference pixel x_o and the captured distortion affected measure pixel x_i . The distortion first appeared in the measured pixel x_i and then passed down to the disparity and normalized disparity.

The pincushion distortion occurred when the pre-stored pattern was casting through the projector lens into the object scene. As shown in Figure 2a, the reference pattern (rectangle shape) would be distorted into a pincushion shape (marked as yellow in Figure 2a). When the pincushion distortion pattern in the object space was then captured by the IR camera's complementary metal-oxide-semiconductor (CMOS) sensor [30], the camera lens would cause additional barrel distortion (Figure 2b) to the already distorted casted pattern. As a result, the captured patterns, which is used to calculate disparity, contains the distortion caused by both IR projector lens and IR camera lens and two types of distortion overlapped in the image plane. Since the disparity, or normalized disparity, are the linear products of the captured pattern, the distortion model that is applicable to capture pattern in pixel unit can also be applied in the disparity or normalized disparity.

Considering both camera and projector distortions, the measured disparity in pixel can be described as:

$$d_M = d_i + \delta_c^{ep} + \delta_p^{ep} \quad (4)$$

where,

d_M is the measurement value of disparity in pixel;

d_i is the true value of disparity in pixel with no error;

δ_c^{ep} is the distortion error caused by camera lens in pixel;

δ_p^{ep} is the distortion error caused by projector lens in pixel.

According to Zhang [17], the optical distortion of the optical lens can be modeled as:

$$\begin{aligned} x_u &= x_d + (x_d - x_c)(K_1 r^2 + K_2 r^4 + \dots) + (P_1(r^2 + 2(x_d - x_c)^2) + 2P_2(x_d - x_c)(y_d - y_c)) \\ y_u &= y_d + (y_d - y_c)(K_1 r^2 + K_2 r^4 + \dots) + (2P_1(x_d - x_c)(y_d - y_c) + P_2(r^2 + 2(y_d - y_c)^2)) \end{aligned} \quad (5)$$

where,

(x_d, y_d) = distorted image point as projected on image plane;
 (x_u, y_u) = undistorted image point as projected by an ideal pinhole camera;
 (x_c, y_c) = distortion centre;
 $K_n = n^{\text{th}}$ radial distortion coefficient;
 $P_n = n^{\text{th}}$ tangential distortion coefficient; and
 $r = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2}$.

Camera distortion is normally barrel distortion with a negative term for K_1 , and projector distortion is often pincushion distortion with a positive K_1 .

Both the camera-caused barrel distortion and projector-caused pincushion distortion are uncorrelated to the distance. The camera-caused barrel distortion in the disparity is a fixed centrosymmetric distortion with a centre at the optical centre of the IR image plane for the whole distance. For the pincushion distortion of the projector lens, it can be proved that the casted projector distortion's effect in the image plane is uncorrelated with distance, and same as the barrel distortion caused by the camera lens, the pincushion distortion on the image plane should be a fixed distortion for all distances. It is because the ratio between the original pincushion distortion in projector plane and the captured pincushion distortion in image plane is a constant value for all the distances. Figure 3 illustrates that the pixel distortion d in projector plane Π_P caused by the projector lens in the object scene has a fixed relationship with the pincushion distortion d_C that appeared in the image plane Π_C .

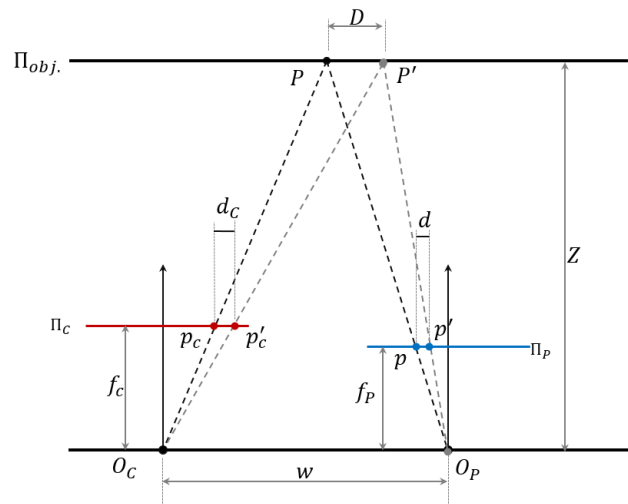


Figure 3. Relationship between the project distortion in object space and in image plane.

Where,

(f_P, f_C) is the focal length of projector lens and camera lens;
 (O_P, O_C) is the optical centre of the projector lens and camera lens;
 Π_{obj} is the object plane;
 Π_P is the projector pattern plane;
 Π_C is the camera image plane;
 (p', p) is the undistorted and distorted pattern location in projector pattern plane Π_P ;
 (P', P) is the undistorted and distorted pattern location casted into the object plane Π_{obj} ;
 (p'_C, p_C) is the undistorted and distorted pixel captured by camera in image plane Π_C ;
 d is the pixel difference in projector plane Π_P ;
 d_C is the pixel difference in image plane Π_C ;
 D is the location difference in object plane Π_{obj} ;

w is the baseline between camera centre and projector centre;
 Z is the distance from baseline to object plane.

As shown in Figure 3, pixel difference d is the pincushion distortion in the projector plane Π_P . When this distortion is projected into the object plane Π_{obj} , it is presented as the location difference D between the actual casting location P and the undistorted location P' . As here we are discussing the projector distortion's effect on the image plane, the camera distortion is not considered for better visualization. In actual sensor measurement, the projector distortion and camera distortion are overlapped in image plane and have a combined effect on the distorted pixel p_C . With no camera lens barrel distortion considered, the location difference D in the object plane Π_{obj} will be captured as the pixel difference d_C in image plane Π_C . From Figure 3, the following equation can be obtained:

$$\begin{cases} \frac{d}{D} = \frac{f_P}{Z} \\ \frac{d_C}{D} = \frac{f_C}{Z} \end{cases} \quad (6)$$

Combine the two parts in Equation (6), the following relationship can be obtained:

$$d_C = \frac{f_C}{f_P} d \quad (7)$$

Since $\frac{f_C}{f_P}$ is a constant value, therefore, no matter what the distance Z is, the pixel shift d_C in the image plane Π_C caused by the projector lens follows the same model as the original pincushion distortion in the projector plane Π_P and just with an additional scale $\frac{f_C}{f_P}$. The scale $\frac{f_C}{f_P}$ can be put into the parameters $K_1 K_2 P_1 P_2$ of Equation (5) as new parameters $K'_1 K'_2 P'_1 P'_2$ in the distortion model. Therefore, the distortion impact from projector in image plane Π_C follows the same pincushion model in projector plane Π_P and is not correlated to the distance from the projector to the object.

Both camera barrel distortion and projector pincushion distortion are centrosymmetric distortion in the normalized disparity unit. The camera barrel distortion is constant at the optical centre of the IR image plane for all the distance range. For the pincushion distortion appeared in the image plane, although it is still a centrosymmetric distortion, it has varying distortion centre (x_0^p, y_0^p) based on the distance Z . In the SLP RGBD sensor's chipset, the disparity is only calculated from the pixel difference in the x axis, and for this reason, the y coordinates y_0^p of the pincushion distortion centre are always equal to 0. The x coordinates of (x_0^p, y_0^p) varies based on the distance. Based on the geometry shown in Figure 4, the image coordinates of projector centre (x_0^p, y_0^p) in image plane can be calculated as Equation (8).

$$\begin{cases} x_0^p = \frac{f_C w}{Z} \\ y_0^p = 0 \end{cases} \quad (8)$$

Although both the barrel distortion and pincushion are centrosymmetric distortion in the disparity unit, the combined overall distortion is a non-centrosymmetric distortion due to the pincushion distortion's varying centre. If the centrosymmetric barrel distortion and pincushion distortion have the same distortion centre in the image plane, as shown in Figure 5a,b, then the overlapped distortion, shown in Figure 5c, is also a centrosymmetric distortion with the distortion in the optical centre of the image frame. According to Equation (8), the pincushion distortion centre (x_0^p, y_0^p) in the image frame is always different from the barrel distortion centre. Although the pincushion distortion in a shifted centre is still a centrosymmetric pattern (Figure 5d), its overlapping result with a barrel distortion of different distortion centre will become a non-centrosymmetric pattern, as shown in Figure 5e. Therefore, the overall distortion in the SLP RGBD sensor has a non-centrosymmetric pattern. The typical distortion pattern is similar to that illustrated in Figure 5e, with distortion centre (the least distorted part) shifted from the optical centre to the side of the projector module and the far side away from the projector centre suffering more severe distortion than the other side.

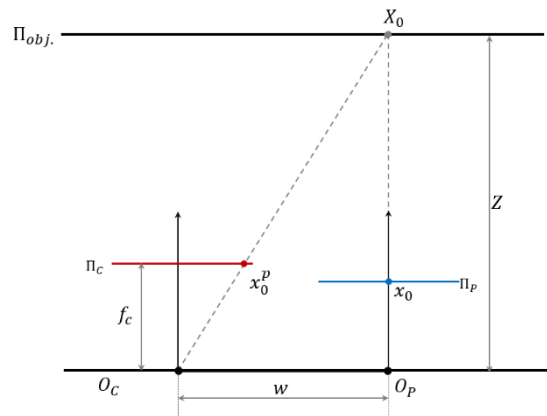


Figure 4. Projector centre in image plane.

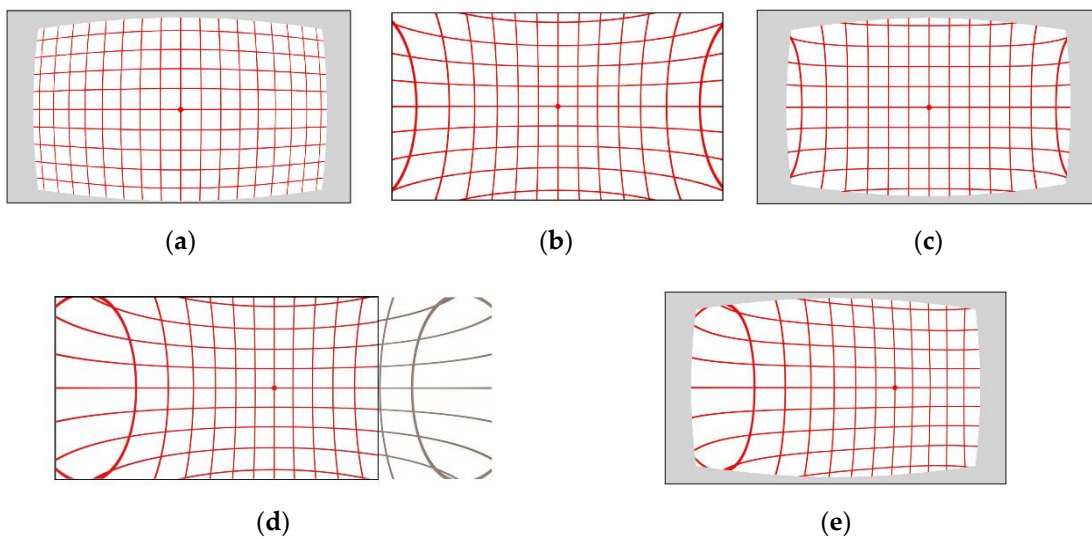


Figure 5. Different overlapping distortion in image plane. (a) Centrosymmetric barrel distortion with centre in the optical centre of image frame; (b) centrosymmetric pincushion distortion with centre in the optical centre of image frame; (c) centrosymmetric overlapped distortion with centre in the optical centre of image frame; (d) centrosymmetric pincushion distortion with centre shifted to the side of image frame; (e) non-centrosymmetric overlapped distortion with barrel distortion centre in the optical centre of image frame and pincushion distortion centre in the shifted side of image frame.

Normalize the disparity in Equation (4) using Equation (2):

$$d'_M = m(d_i + \delta_c^{ep} + \delta_p^{ep}) + n = d'_i + m\delta_c^{ep} + m\delta_p^{ep} \tag{9}$$

where,

d'_M is the measurement value of normalized disparity;

m, n are the normalize parameter of disparity;

d_i is the true value of disparity in pixel with no error;

δ_c^{ep} is the distortion error caused by camera lens in pixel;

δ_p^{ep} is the distortion error caused by projector lens in pixel;

d'_i is the true value of normalized disparity.

Based on the Brown's model [31], δ_c^{ep} and δ_p^{ep} can be expanded to radial distortion part and tangential distortion part as:

$$\begin{cases} \delta_c^{ep} = \delta_c^{rad.} + \delta_c^{tang.} \\ \delta_p^{ep} = \delta_p^{rad.} + \delta_p^{tang.} \end{cases} \quad (10)$$

The distortion model can be expressed as:

$$\begin{cases} \delta^{rad.} = x_i(K_1r^2 + K_2r^4 + K_3r^6) \\ \delta^{tang.} = P_1(r^2 + 2x_i^2) + P_2x_iy_i \end{cases} \quad (11)$$

where,

$\delta_c^{rad.}$ is the radial distortion of the IR camera;

$\delta_c^{tang.}$ is the tangential distortion of the IR camera;

$\delta_p^{rad.}$ is the radial distortion of the IR projector;

$\delta_p^{tang.}$ is the tangential distortion of vIR projector;

$$r = \sqrt{x_i^2 + y_i^2}; x_i = x_i^m - x_0, y_i = y_i^m - y_0;$$

(x_0, y_0) is the camera or projector centre;

(x_i^m, y_i^m) is measured image coordinate;

K_1, K_2, K_3 is the parameters of radial distortion model;

P_1, P_2 is the parameters of tangential distortion model.

In this way, Equation (10) can be further expanded as:

$$\begin{cases} \delta_c^{ep} = x_i^c(K_1^c r_c^2 + K_2^c r_c^4 + K_3^c r_c^6) + P_1^c(r_c^2 + 2x_i^{c2}) + P_2^c x_i^c y_i^c \\ \delta_p^{ep} = x_i^p(K_1^p r_p^2 + K_2^p r_p^4 + K_3^p r_p^6) + P_1^p(r_p^2 + 2x_i^{p2}) + P_2^p x_i^p y_i^p \end{cases} \quad (12)$$

where,

$K_1^c, K_2^c, K_3^c, P_1^c, P_2^c$ are the distortion parameters of the IR camera;

$$r_c = \sqrt{x_i^{c2} + y_i^{c2}}; x_i^c = x_i^m - x_0^c; y_i^c = y_i^m - y_0^c;$$

(x_0^c, y_0^c) is the centre point of camera;

(x_i^m, y_i^m) is the measured image position;

$K_1^p, K_2^p, K_3^p, P_1^p, P_2^p$ are the distortion parameters of the IR projector;

$$r_p = \sqrt{x_i^{p2} + y_i^{p2}}; x_i^p = x_i^m - x_0^p; y_i^p = y_i^m - y_0^p;$$

(x_0^p, y_0^p) is the centre point of projector distortion model in image coordinates;

(x_i^m, y_i^m) is the measured image position without the influence of camera distortion.

$$\begin{cases} x_p^m = x_c^m + \delta_c^{ep} \\ y_p^m = y_c^m + \delta_c^{ep} \end{cases} \quad (13)$$

Combine (9), (10) & (12), the proposed calibration model in normalized disparity can be expressed as:

$$\begin{aligned} \delta_{d_i} = d'_M - d'_i = & mx_i^c(K_1^c r_c^2 + K_2^c r_c^4 + K_3^c r_c^6) + mP_1^c(r_c^2 + 2x_i^{c2}) + mP_2^c x_i^c y_i^c \\ & + mx_i^p(K_1^p r_p^2 + K_2^p r_p^4 + K_3^p r_p^6) + mP_1^p(r_p^2 + 2x_i^{p2}) + mP_2^p x_i^p y_i^p \end{aligned} \quad (14)$$

Therefore, for the SLP-based RGBD sensor, the distortion error in disparity consists of two parts, one is a barrel distortion caused by camera lens with fixed location at optical centre. The other is pincushion distortion caused by the projector lens with a varying location over the camera image plane.

In practice, since the actual distance Z of projector to the object scene is often unknown, the proposed calibration model uses the measured distance Z_M in Equation (8) to calculate the distortion centre (x_0^p, y_0^p) as an alternative.

3. Calibration of Structured Light Pattern (SLP) RGBD Sensor

In the SLP sensor, both camera distortion and projector distortion occur. In the image plane, the camera distortion is fixed as its optical centre, and the projector lens distortion has a varying optical centre in the image plane according to the distance from projector to the object. Therefore, if the distortion model for the camera and projectors is known, the combined distortion can be calculated proximately by the measured distance. The distortion calibration for infrared camera can be conducted in the traditional ways. But for the raw observation, the structured pattern of infrared projector of RGBD sensor is inaccessible, and there is no way to determine the projector distortion model by using the existing method [32] that requires a known reference pattern and ability to take measurement of the actual distorted casted pattern in the object space. The RGBD sensor only provides a normalized disparity unit that contains the combined distortion. To calibrate such combined non-centrosymmetric distortion, a two-step calibration procedure is proposed.

A two-step calibration procedure is designed to specifically calibrate the SLP-based RGBD sensors. The RGBD sensor used in this paper is the Structure Sensor from Occipital [7], shown in Figure 6. It is a SLP-based RGBD sensor released in 2013. Structure Sensor shares the same ranging principle with Kinect V1 and Asus Xtion that are all based on the PrimeSense technology for SLP ranging [11]. The calibration procedure developed in this paper is also applicable for other PrimeSense-based SLP RGBD sensor. To obtain the parameters of the proposed calibration model (Equation (14)), the standard checkboard calibration method for the RGB camera is first used to get the parameters related to the infrared camera lens. Direct sunlight or infrared light is used to illuminate the checkboard. Then a test field is set up to calibrate the distortion related to the infrared projector. Sample data is captured at a different distance and plane fitted to estimate the distortion parameters. The camera distortion and projector distortion are combined based on the varying focal centre relating to the measured distance. At a different distance, there will be different combined effects on the depth frame due to the shifting location of the projector focal centre on the image plane. The remaining error after modelling the camera distortion and projector distortion is regarded as bias caused by the infrared cone.

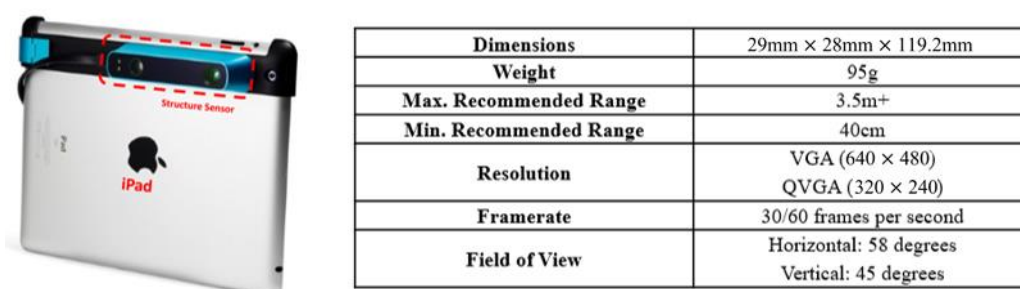


Figure 6. Structure sensor specifications.

3.1. Data Acquisition and Data Processing

In order to calibrate the SLP RGBD sensor, a calibration field is set in an open space. Figure 7 shows the test field used in this paper. The wall in the test field is served as the control plane for calibration. The control plane is a plane with image identifiable markers and serves as the ground truth for calibration. The 3D coordinates of the image markers are measured by total station, an electronic surveying instrument that can measure high-precision 3D coordinates in mm level accuracy. Those image identifiable markers can provide a calibration dataset of image coordinates with its corresponding 3D coordinates in objective space. The calibration dataset is captured at a different

distance to have a full coverage of the SLP RGBD sensor's whole working range. In our test, the calibration dataset is captured from 0.5 m to 6.5 m with an interval about 0.5 m to cover the working range of the Structure Sensor. Since the wall with calibration mark covers less image area when the distance from the sensor to the wall increases, more than one station should be set for longer distance to make sure the calibration markers can cover the whole image.



Figure 7. Test field for calibration.

The data acquisition and processing procedure is as follows:

1. Establish a control plane by placing the control markers with known coordinates measured by total stations.
2. Fit the plane equation of the control plane.
3. Place the Structure Sensor in front of the control wall and take multiple observations (100 observations) at each designed distance (from 0.5 m to 6.5 m with an interval of about 0.5 m), as shown in Figure 8a,b.
4. Calculate the pixel-wise average depth frame for each distance to reduce the effect of random noise in each pixel.
5. Use the perspective-n-point algorithm [33] to calculate the camera extrinsic parameters for each distance based on the correspondences between the control wall and the depth frame.
6. Unify the coordinates system between Structure Sensor and the control wall based on the calculated camera extrinsic parameters in 5.
7. Calculate the ground truth depth for each distance based on the fitted control plane equation.
8. Calculate the ground truth disparity based on the processed ground truth depth.
9. Use the ground truth disparity to calculate the disparity error of the averaged disparity measurement in each distance.

The disparity error obtained after the data processing is used for model parameters estimation. The plane-fitted method is used to fit the equation of the control plane to provide adequate data for the parameter estimation. Figure 8b shows the different stations settled for data acquisition (camera markers) and the fitted planes based on different measurements (different colour on the control wall). The fitted plane can provide a significant amount of data for the estimation. If only the control markers measured in different distance are used for parameters estimation, the valid point that can be used for estimation is still too sparse to provide the full frame coverage, as shown in Figure 8c. By using the control markers, only 864 pixels within the frame have the associated disparity error and are valid for estimation, which covers about 0.28% of the whole 640×480 frame. By using the plane fitting to calculate the ground truth disparity for each control plane pixel, we can improve the number of valid pixel to 293.196, which cover about 94.44% of the whole frame. Furthermore, each pixel as shown in Figure 8d has multiple disparity errors that are measured from different distance. The comparison between Figure 8c,d shows a significant improvement of the valid data for calibration parameters'

estimation by the plane-fitting method. By minimizing the disparity errors for each pixel, we can further reduce random noise for a better parameter estimation.

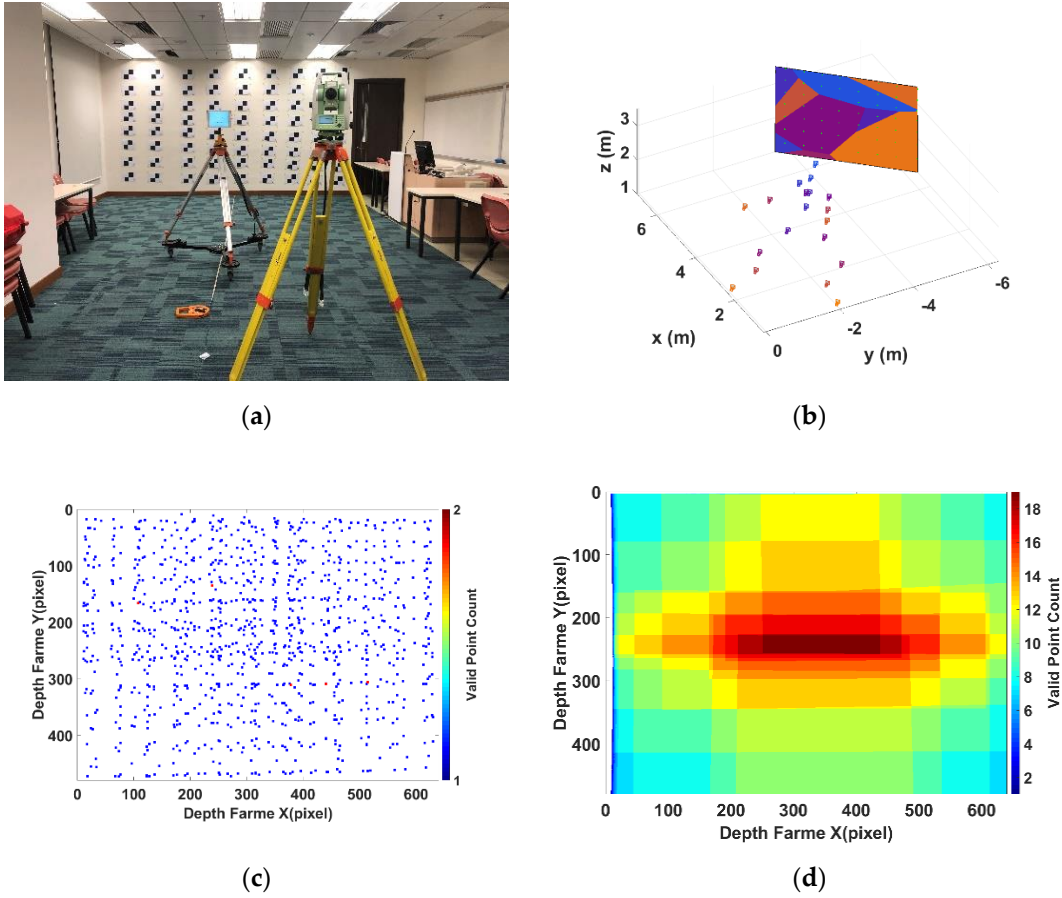


Figure 8. (a) Setup for data acquisition; (b) relationship between the stations and the fitted plane; (c) sparse valid data for estimation; (d) dense valid data calculated from plane fitting.

3.2. Model Parameter Estimation

With the fitted plane $ax + by + cz + d = 0$ and the camera pose $(R, t | R \in SO(3), t \in \mathbb{R}^3)$ are in the same reference system, the ground truth disparity error can be calculated through the following procedure.

In the reference system of the fitted plane, the projection location $X_i \in \mathbb{R}^3$ of the depth pixel (u, v) on the fitted plane $ax + by + cz + d = 0$ can be given by:

$$X_i = \left(\frac{(P_0 - L_0) \cdot \vec{n}}{\vec{l} \cdot \vec{n}} \right) \vec{l} + L_0 \quad (15)$$

where P_0 can be any point on the fitted plane, L_0 is the coordinates of the camera centre t , \vec{n} is the normal vector of fitted plane, and \vec{l} is the line vector of the line pass through the camera centre t and depth pixel (u, v) .

Then the ground truth depth Z_i for the pixel (u, v) in the depth frame can be calculated as:

$$Z_i = \frac{|\text{col}_3(R) \cdot X_i - \text{col}_3(R) \cdot t|}{\sqrt{\text{col}_3(R) \cdot \text{col}_3(R)}} \quad (16)$$

where $col_3(R)$ stands for the third column of rotation matrix R . Based on Equation (3), the ground truth disparity d'_i can be given as:

$$d'_i = \frac{1 - \alpha Z_i}{\beta Z_i} \quad (17)$$

Therefore, based on Equation (9), the disparity error δ_d of the depth pixel (u, v) can be obtained as:

$$\delta_d = m(\delta_c^{ep} + \delta_p^{ep}) = d'_M - d'_i \quad (18)$$

The disparity errors obtained at different distances in the data acquisition are visualized in the following figure. Figure 9 shows the normalized disparity errors observed from 0.5 m to 6.5 m (with assorted colours). The different colours in Figure 9 represent the different measurements.

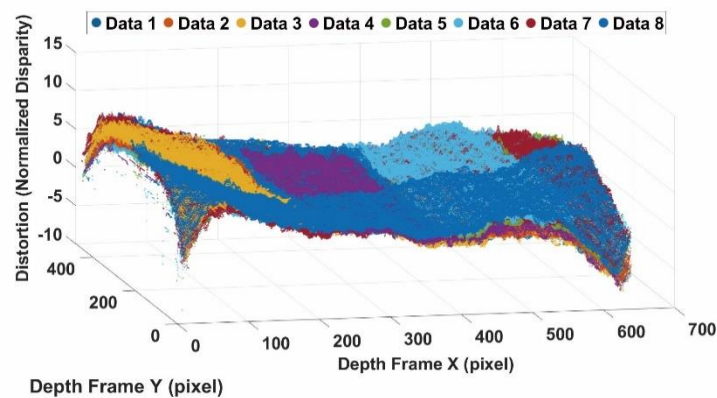


Figure 9. Observed error in normalized disparity.

Firstly, the IR camera distortion part can be calibrated separately by the standard checkboard calibration procedure [16]. It can be seen that the camera distortion is in both x axis and y axis of the image plane. While in disparity, only the x axis is required as the RGBD sensor only uses the pixel difference in the x axis to calculate the disparity. The distortion occurring in the y axis will not affect the disparity. Therefore, only the x axis is used as variables in the proposed disparity calibration model. Figure 10 shows the distortion model of the barrel distortion caused by camera lens that are calibrated by the standard camera calibration procedure. Figure 10a shows the modeled disparity correction for the whole frame coverage. The unit of x , the y axis of Figure 10a is the image pixel. The colour in Figure 10a indicates the disparity correction in normalized disparity. Figure 10b is a 3D visualization of Figure 10a. The x, y axis of Figure 10b is in the image pixel and z axis is in the normalized disparity.

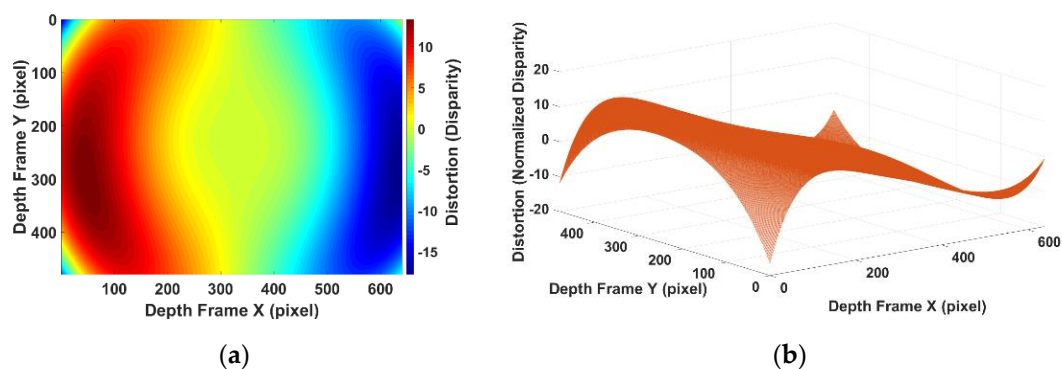


Figure 10. (a) Calibrated camera distortion in normalized disparity unit in image plane; (b) calibrated camera distortion in normalized disparity unit in 3D visualization.

The parameters of Structure Sensor's camera radial distortion and tangential distortion are estimated and listed in Table 1. The parameters are associated with the dimensionless normalized image coordinates, which are calculated from pixel coordinates by translating to the optical centre and dividing by the focal length in the pixel. Also, according to Equation (3), we can have $\beta = \frac{1}{fwm}$. Since β and the baseline w are known manufacturing parameters, and the focal length f can be obtained from standard camera calibration, the normalized coefficient m can be calculated.

Table 1. Calibrated distortion parameters of infrared (IR) camera.

Camera Parameter	Radial Distortion			Tangential Distortion		Normalized Coefficient
	K_1^c	K_2^c	K_3^c	P_1^c	P_2^c	m
Structure Sensor	-0.0578	0.1248	1.0212×10^{-4}	-0.0010	-6.0139×10^{-4}	7.9418

After using the standard calibration method to model the camera-related barrel distortion, the pincushion distortion caused by projector can be simply separated by subtracting the observed data by the camera part (as shown in Figure 11a). As mentioned in Section 2, the remaining pincushion distortion is a centrosymmetric distortion with a shifted distortion centre based on the distance. How the projector caused the distortion overlay with the camera part is determined by the distance from the sensor to the object. As shown in Equation (8), the distortion centre of projector distortion can be calculated. The disparity errors caused by pincushion model shown in Figure 11a have different distortion centre based on the distance from the sensor to the object scene. Actually, the distortion parameters for these data captured at different distances are same. They have the same centrosymmetric shape but with different distortion centre in the image plane. By knowing the exact distance from sensor to the object scene, the shift caused by distance can be precisely calculated. All the disparity errors captured at different distances are then aligned back to the optical centre of the image frame for a better estimation of the distortion parameters, as shown in Figure 11b.

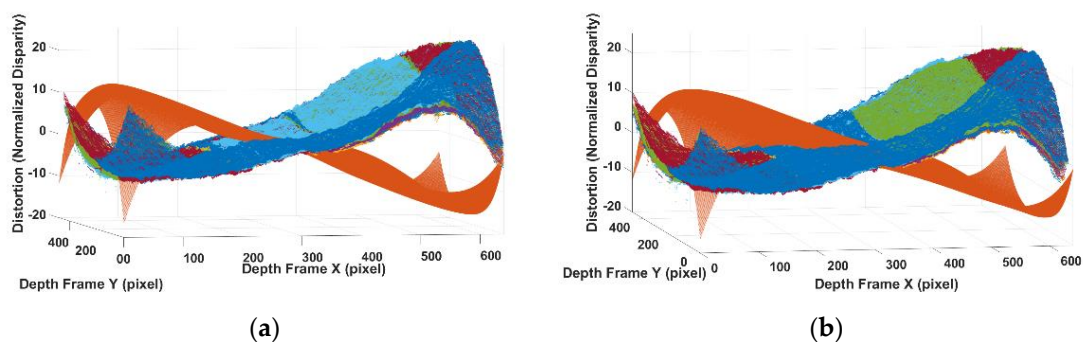


Figure 11. (a) Camera distortion and unaligned projector distortion; (b) camera distortion and aligned projector distortion.

By aligning the shifted projector distortion and taking the average disparity error in each pixel to reduce the effects of random noise, the processed data is regarded as the projector-caused disparity distortion. The aligned and averaged data shown in Figure 12 is the data processed to estimate the IR projector distortion parameters. The stripe pattern that appeared in Figure 12b is a normal phenomenon in SLP RGBD sensor [23,34]. The actual cause of this has not been revealed by the manufacture. Ref. [35] suggests the possible cause for this stripe pattern is the rolling shutter used in the IR camera. As the scale of the stripe pattern is much smaller than the distortion [6], the stripe pattern is uncalibrated in the proposed method.

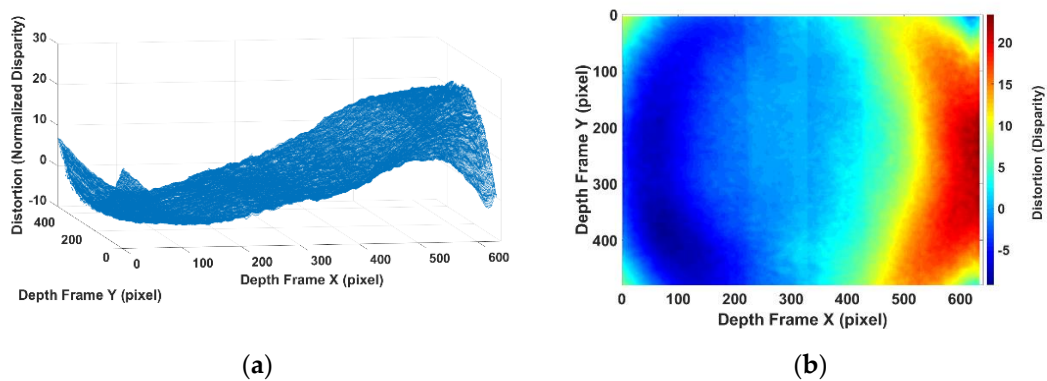


Figure 12. (a) Averaged processed projector distortion in 3D visualization; (b) averaged processed projector distortion in image plane.

Then, the Levenberg–Marquardt fitting algorithm [36] is used to minimize the cost function in Equation (10) and to estimate the parameters of the projector distortion model. The calibrated parameters are listed in Table 2:

Table 2. Calibrated distortion parameters of IR projector.

Projector Parameter	Radial Distortion			Tangential Distortion	
	K_1^p	K_2^p	K_3^p	P_1^p	P_2^p
Structure Sensor	0.0474	-0.0714	-0.1014	0.0019	1.4390×10^{-4}

The parameter estimation result is shown in Figure 13a. The distortion model is compared with the actual observed distortion error. The comparison between the actual projector distortion and the modeled result (Figure 13b) shows that the calibration model can precisely match the actual disparity error. After calibrating the distortions caused by camera lens and projector lens, there are still some remaining errors of the observed disparity error. As shown in Figure 13b, there is still an increasing offset to the edge area between the actual disparity error and the proposed distortion model.

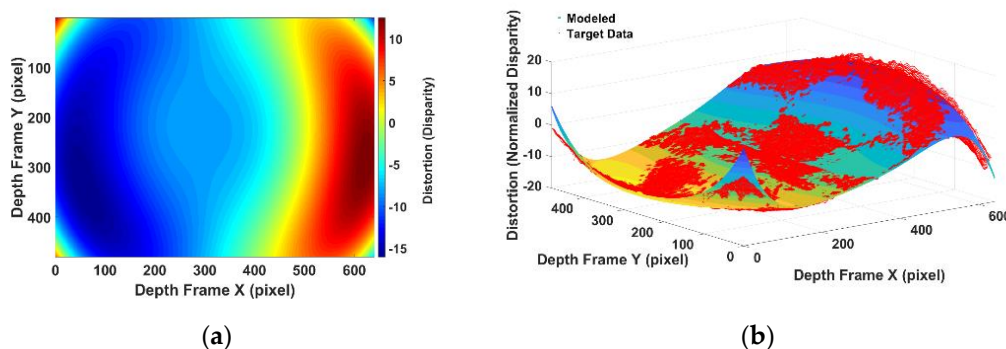


Figure 13. (a) Modelled projector distortion; (b) comparison of modelled projector error and actual data (red markers are the actual data; colored surface is the modelled projector error).

The remaining part of the error according to [6,23] is caused by the reason that the images cannot be illuminated homogeneously, known as the infrared cone error. When the infrared light is emitted by the projector and cast on to the scene, the central area of the cone will have the highest illumination level. The illumination will decay towards the edge area. In the RGBD sensor, the illumination level peaks at the central area and decays in the edge area. As better illumination often benefits the pattern recognition performance [37], the offset caused by infrared cone in the central area is much smaller than the offset in the corner. As a result, it causes a systematic error across the depth frame with small

offset in the central area and increasing offset towards the edge, as presented in Figure 14a. To model this systematic error caused by the infrared cone, a low order polynomial equation, shown in Equation (19), is applied to fit the offset caused by infrared cone. In Equation (19), x, y is the normalized image coordinates by optical centre in pixel, and $p_{00}, p_{10}, \dots, p_{03}$ are the polynomial parameters. The modeled result is shown in Figure 14b. The fitted parameters for Equation (19) are listed in Table 3.

$$f(x, y) = p_{00} + p_{10}x + p_{01}y + p_{20}x^2 + p_{11}xy + p_{02}y^2 + p_{30}x^3 + p_{21}x^2y + p_{12}xy^2 + p_{03}y^3 \quad (19)$$

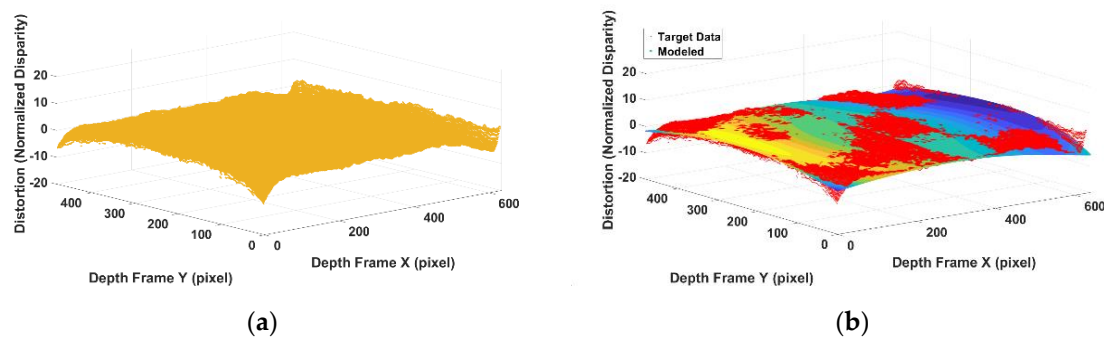


Figure 14. (a) Visualization of the remaining error after eliminate camera and projector distortions; (b) comparison of modelled IR cone error and actual error (red markers are the actual data, coloured surface is the modelled IR cone error).

Table 3. Fitted polynomial parameters for infrared cone error.

Polynomial Parameters	Value	Polynomial Parameters	Value
p_{00}	0.5432	p_{02}	-0.6831
p_{10}	-0.0579	p_{30}	0.0604
p_{01}	0.1775	p_{21}	0.0535
p_{20}	0.1471	p_{12}	-0.0368
p_{11}	0.0113	p_{03}	-0.0515

To summarize, the calibration model consists of three parts (Figure 15): (1) a barrel distortion model calibrated with classic checkboard camera calibration, which is fixed in the optical centre; (2) a pincushion distortion model based on the proposed procedure, of which the effects on the disparity is a fixed centrosymmetric shape while this has a varying distortion centre on the image plane; (3) a low order polynomial equation for the error caused by infrared cone, which has the same varying centre as the pincushion distortion because the cone is caused by the projector when it illuminates the scene. Combining these three parts, we can achieve a range-independent disparity-based calibration model for SLP RGBD sensors. Since the calibration model is developed based on the actual physical model of the SLP RGBD sensor's distortion, the proposed method is a range independent model. No range-based parameter selection is required in the proposed method. It can provide the full range coverage with one single model of 20 parameters (5 parameters for barrel distortion, 5 parameters for pincushion distortion, and 10 parameters for the infrared cone related error). Considering the overlapping mechanism between the barrel distortion and pincushion, the proposed method can use a non-centrosymmetric distortion model to provide full frame calibration of the depth frame. The offset in the edge area is largely calibrated. By accurately calibrating the edge area, the proposed method can significantly increase the quality of the depth frame and provide additional information that is missed out due to the large error.

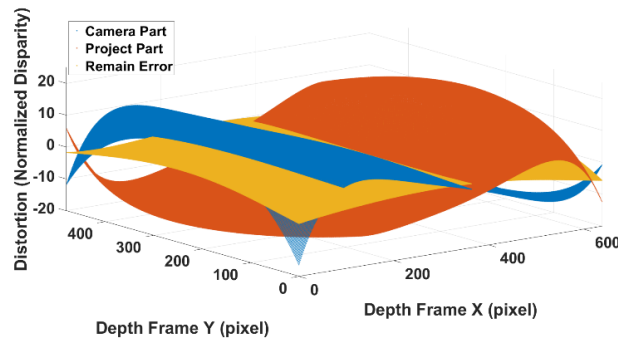


Figure 15. Three main parts of the distortion in the RGBD sensor.

4. Calibration Model Performance

Two experiments were designed to test the performance of the proposed calibration method with the SLP RGBD sensor, Structure Sensor. The performance of the calibration model was evaluated by comparing the difference between the modeled error and the actual error. Then a more straightforward showcase of the calibrated point cloud and the uncalibrated point cloud was provided to demonstrate the performance of the proposed calibration model.

4.1. Model Evaluation

To test the calibration method’s performance, a test dataset that followed the same setup as mentioned in Section 3.1 was obtained in an unfamiliar environment. The proposed calibration model was put into a test to compare the difference between the actual disparity error and the modelled error by the proposed method. The test dataset was captured from 2 m to 7 m with an interval of about 0.5 m. The example shown in Figure 16 is three typical results for close range (1.93 m), middle range (4.01 m) and far range (6.38 m). The first column is the actual observed disparity error. The second column is the modeled disparity error based on the proposed calibration model. The last column is the difference between the first two columns. For all the figures presented in Figure 16, the x and y axis are in image pixels and the colour indicates the value of the disparity error at that pixel location in the normalized disparity. As shown in Figure 16, the proposed model can accurately model the actual error. The difference between the modelled error and the actual error is minor. The root mean squared error (RMSE) of the overall modelled error is 0.76 disparity units. While two right corners still have some large offset that cannot be accurately modelled. It may be caused by the less enough data in the corner area.

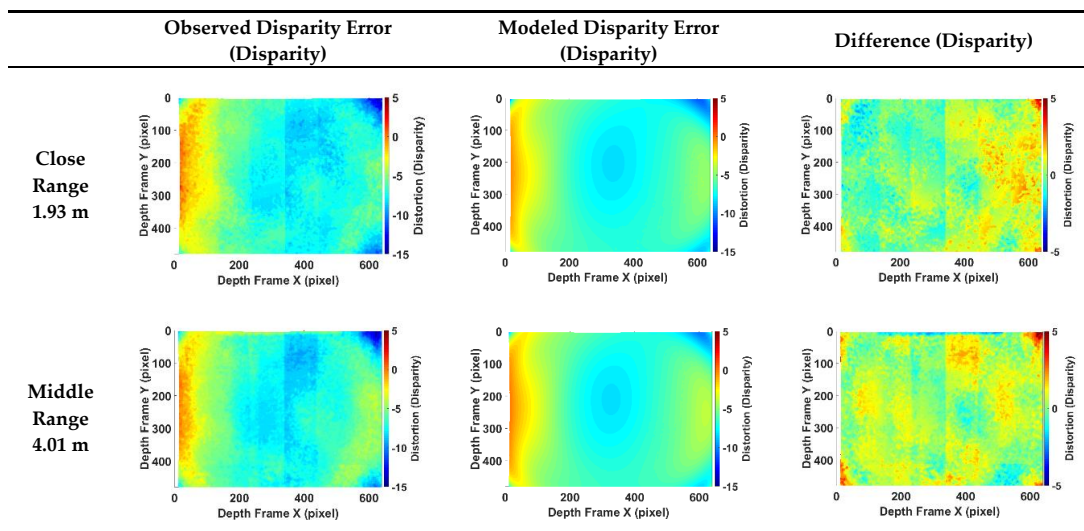


Figure 16. Cont.

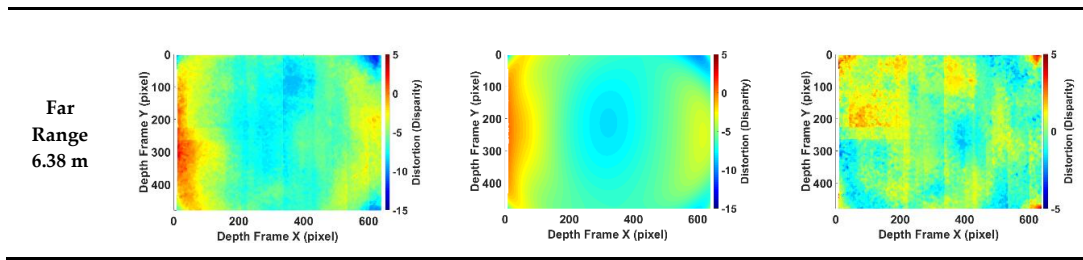


Figure 16. Comparison of ground truth of disparity error, the modelled disparity error based on the proposed model, and the difference between the ground truth and the modelled data.

4.2. Point Cloud Correction

A more direct method to evaluate the quality of the calibration model is the visualization of data quality improvement of the measured point cloud. The proposed method was tested in three different scenes: (1) flat surface; (2) non-flat (curved) surface; and (3) flat surface with distance variation, shown in Figure 17. Darwish et al.’s calibration method [19] was selected as the reference of existing calibration performance for the SLP RGBD sensor. As Figure 17 shows, the proposed calibration model improves the point cloud accuracy in all three scenes. The validations in non-flat surface and the surface with depth variation indicate the calibration model was applicable in all ranges. The 3D data between the sampling distances can also be calibrated by the proposed method.

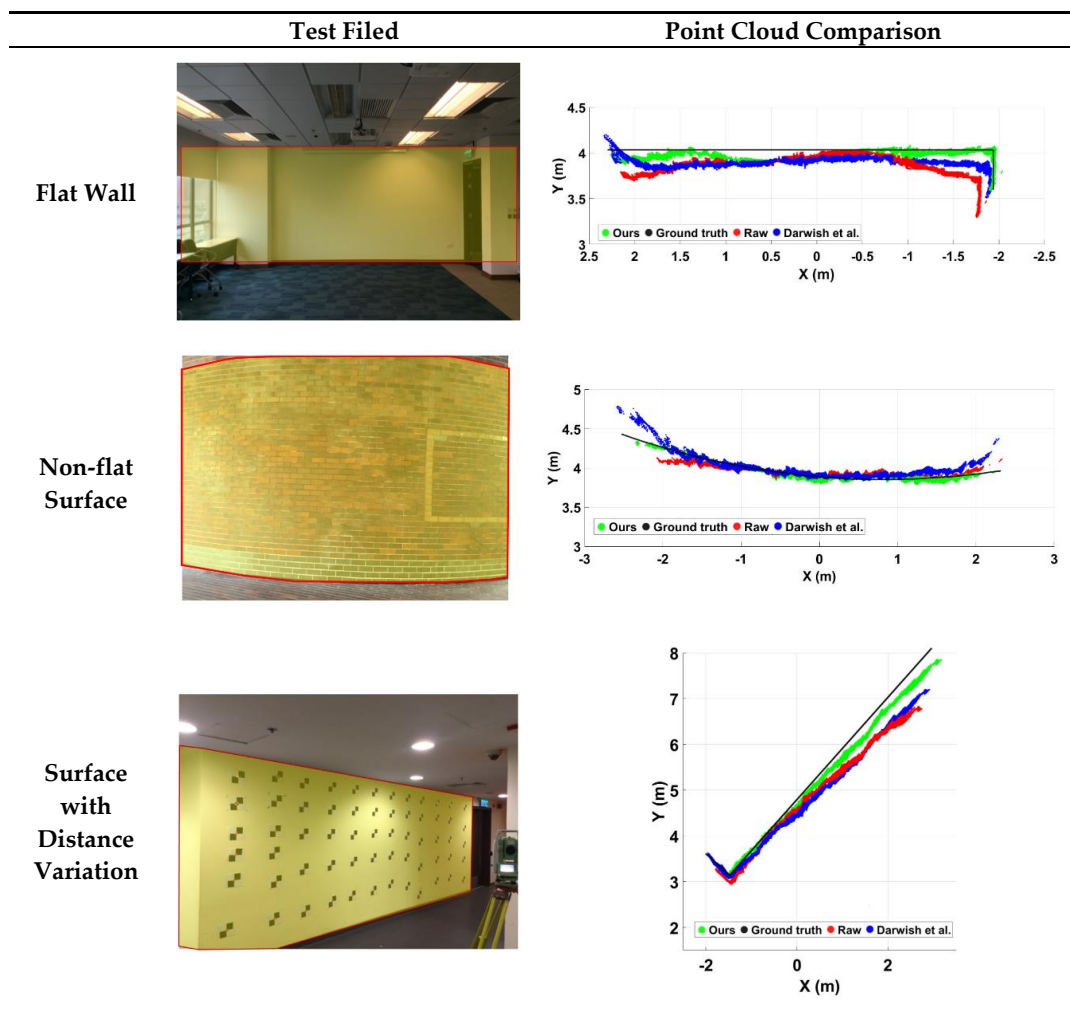


Figure 17. Calibration comparison on flat surface, non-flat surface, and surface with distance variation.

Figure 18 shows the full range improvements of the calibration model. For the full range of the SLP RGBD sensor's working range, the proposed method can greatly improve the quality of the measured point cloud. The improvement of the edge area shown in Figure 18 indicates the full frame coverage of the proposed calibration method. The calibration model is derived from the actual physical properties of how the camera distortion and projector distortion combined in the image frame. Therefore, the proposed calibration can accurately model the actual error to greatly improve the point cloud quality of the low-cost sensor, especially in the far range and edge area of the depth frame.

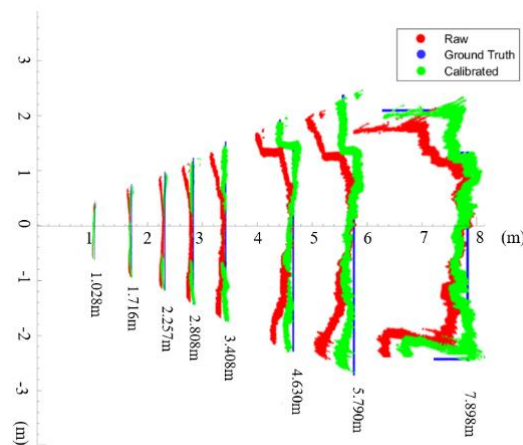


Figure 18. Full range result of the calibration model.

The accuracy improvement of the proposed calibration method can be demonstrated from two different regions within the depth frame. As shown in Figure 18, the offset of the depth measurement increases differently between the central region and edge region of the depth frame. The central region that is near the optical centre of the camera suffers much less errors compared to the edge region. The increasing pace of the offset in the edge region is much faster than the central part. Therefore, it is reasonable to demonstrate the accuracy improvements for different parts of the frame separately. Figure 19 shows how the central region and edge region are defined in the depth frame. A central subset and an edge subset of depth measurement are selected based on Figure 19 for each tested distance. The mean value of each subset is used for accuracy evaluation.

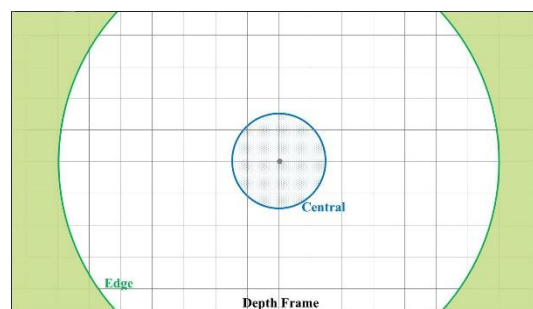


Figure 19. Central region and edge region of depth frame.

Figure 20 shows the accuracy improvement of the proposed method. The x axis in all figures is the ground truth distance in meters. The y axis in Figure 20a,b is the ranging error in meters. The y axis in Figure 20c,d is the relative ranging error to ground truth ratio in percentages. The markers linked with lines in the figures are the original observed data at different distances. The dot lines are the polynomial fitted trend for its corresponding data with the same colour. Figure 20a,b show the accuracy improvement of applying a calibration model for both central area and edge area of the depth frame. As shown in Figure 20a, for the central region, the uncalibrated range normally has around 30

cm ranging error at 8 m distance. After calibration, the central region will only have about 10 cm error at 8 m distance. The accuracy is improved by about 70% in the central region. The improvement for the edge area is more significant as the proposed model is developed from the actual mechanism of the distortion and provides the full frame coverage. As shown in Figure 20b, the uncalibrated edge region suffers a large error (86 cm) at 8 m distance. After calibration, the edge region calibrated depth reduces the error from 86 cm to 30 cm at 8 m distance, which is improved by about 65%. As for the relative error ratio (error to distance) that is shown in Figure 20c, the relative error of the central area is largely reduced through the entire working range from an overall 4% of the distance to about 1% of the distance. The improvement for the edge area shown in Figure 20d is more significant as the proposed model is developed from the actual mechanism of the distortion and provides the full frame coverage. The edge region has a large 11% relative error of the ranging distance before the calibration. It suffers a much more severe distortion compared to the central region's 4% relative error at the same distance. With the calibration model, the relative error in the edge region can be reduced from 10% of the distance to 3% of the distance. It is also indicated in Figure 20d that before calibration the relative error/distance ratio increases with the distance. After calibration, the calibration data can provide stable 2% to 3% ranging data in the edge region and 1% to 2% ranging data in the central region for the full range.

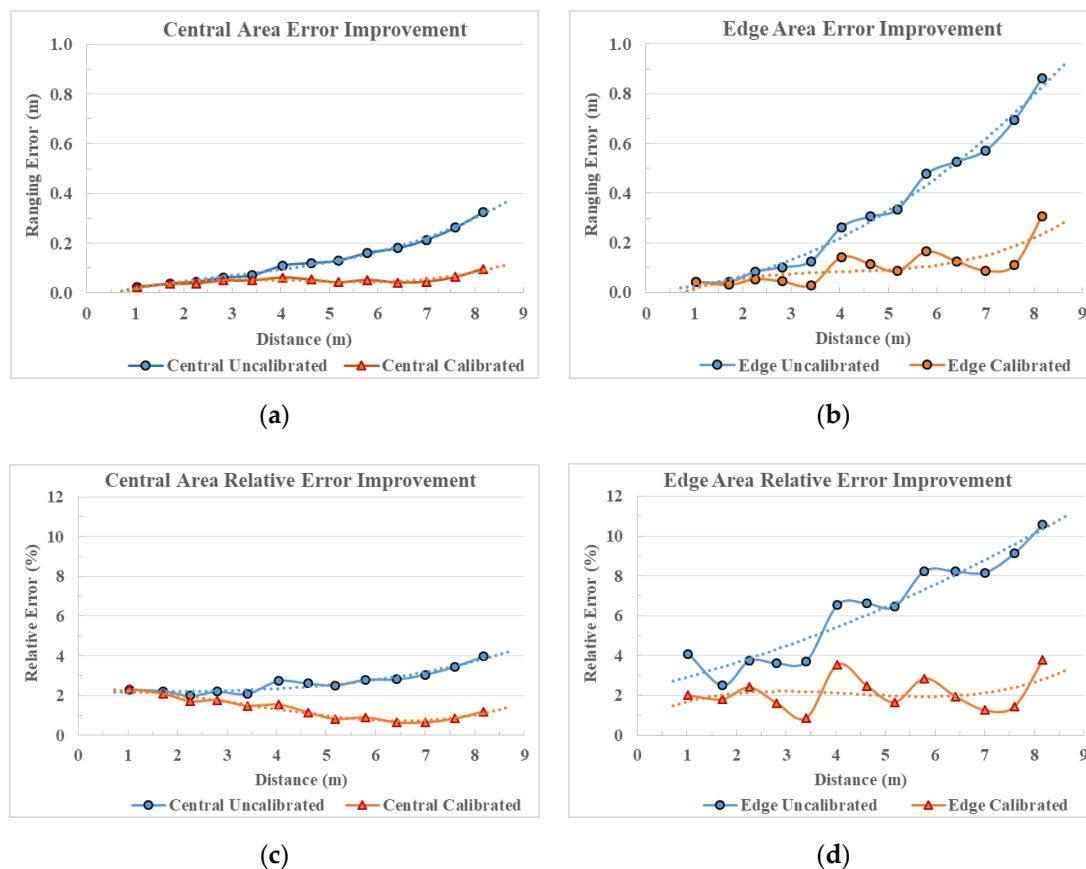


Figure 20. Accuracy and relative error ratio to distance for central and edge area with trend included. (a) Ranging error improvement in central area; (b) ranging error improvement in edge area; (c) relative error improvement in central area; and (d) relative error improvement in edge area.

5. Conclusions and Future Work

The RGBD sensor has a promising future to replace the high-cost 3D laser scanner and be applied in high-precision applications such as robotics, high-precision localization and mapping. The radial systematic error in the depth frame significantly limits its potential applications. Targeting this problem,

this paper presented range-independent disparity-based calibration method for the SLP RGBD sensor. By revealing the real cause and forming mechanism of the non-centrosymmetric depth distortion, the proposed calibration method targets the disparity unit rather than the depth. By applying the calibration model in disparity and calculating the calibrated depth based on the calibrated disparity, the calibration model is independent of distance. No additional distance-based calibration or parameters look up table is required for the proposed method. With only one model with 20 parameters, the proposed calibration method can provide a full range coverage for the SLP RGBD sensor. A new non-centrosymmetric distortion calibration model for the normalized disparity is proposed in this paper based on the discussion on the form of the mechanism of the SLP RGBD sensor distortion. The proposed non-centrosymmetric distortion model can significantly reduce the large offset in the edge area of the depth frame. Since the heavily distorted edge area now can be calibrated to a similar accuracy level to the central area, more usable and valid information can be extracted and used to benefit applications such as SLAM, robotic exploration and obstacle avoidance. A new two-step calibration procedure is also developed in this paper to calibrate the barrel distortion caused by the IR camera lens, pincushion caused by the IR projector lens, and the systematic error caused by the IR cone.

In the experimental results, the full frame and full-range coverage of the proposed calibration method is demonstrated. The comparison between the calibrated and uncalibrated point cloud clearly shows that the systematic errors in the measured point cloud have been removed by the calibration model. The significant offset in the edge area of long-range depth is reduced by the proposed model from 86 cm to 30 cm, which means that the relative error is reduced from 11% to 3% of the range distance. Overall, at far range the proposed calibration method can improve the depth accuracy by 70% in the central region of depth frame and 65% in the edge region.

Further work will study the long-term stability of calibration parameters for the consumer-grade RGBD sensor. Other potential systematic error sources, such as illumination condition, temperature, humidity, and air refractive index, will be investigated for a more comprehensive calibration model.

Author Contributions: Conceptualization, W.L. and Y.L.; Methodology, W.L. and Y.L.; Software, W.L., Y.L., W.D. and S.T.; Validation, W.L., Y.L. and Y.H.; Formal analysis, W.L., Y.L. and W.C.; Investigation, W.L, Y.L. and Y.H.; Data curation, Y.H.; Writing—original draft preparation, W.L.; Writing—review and editing, W.C. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Shenzhen Science and Technology Innovation Commission, grant number JCYJ20170818104822282, Hong Kong Research Grants Council (RGC) Competitive Earmarked Research Grant, grant number 152223/18E, and the Research Institute of Sustainable Urban Development, The Hong Kong Polytechnic University. The APC was funded by Shenzhen Science and Technology Innovation Commission.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Newcombe, R.A.; Izadi, S.; Hilliges, O.; Molyneaux, D.; Kim, D.; Davison, A.J.; Kohli, P.; Shotton, J.; Hodges, S.; Fitzgibbon, A.W. Kinectfusion: Real-time dense surface mapping and tracking. In Proceedings of the 2011 10th ISMAR, Basel, Switzerland, 26–29 October 2011; pp. 127–136.
2. Wang, K.; Zhang, G.; Bao, H. Robust 3D reconstruction with an RGB-D camera. *IEEE Trans. Image Process.* **2014**, *23*, 4893–4906. [[CrossRef](#)]
3. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [[CrossRef](#)]
4. Karaman, S.; Anders, A.; Boulet, M.; Connor, J.; Gregson, K.; Guerra, W.; Guldner, O.; Mohamoud, M.; Plancher, B.; Shin, R. Project-based, collaborative, algorithmic robotics for high school students: Programming self-driving race cars at MIT. In Proceedings of the 2017 IEEE Integrated STEM Education Conference (ISEC), Princeton, NJ, USA, 11 March 2017; pp. 195–203.
5. Khan, A.; Moideen, F.; Lopez, J.; Khoo, W.L.; Zhu, Z. KinDectect: Kinect detecting objects. In Proceedings of the International Conference on Computers for Handicapped Persons, Linz, Austria, 11–13 July 2012; pp. 588–595.

6. Wasenmüller, O.; Stricker, D. Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; pp. 34–45.
7. Occipital. Structure Sensor. Available online: <https://structure.io/> (accessed on 8 January 2019).
8. Asus. Use Xtion PRO Developer Solution to Make Motion-Sensing Applications and Games. Available online: https://www.asus.com/3D-Sensor/Xtion_PRO/ (accessed on 5 June 2019).
9. Introducing Intel®RealSense™ LiDAR Camera L515. Available online: <https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html> (accessed on 8 January 2019).
10. ZED. Meet ZED. Available online: <https://www.stereolabs.com/zed/> (accessed on 8 January 2019).
11. MIT. PrimeSense-MIT Technology Review. Available online: <http://www2.technologyreview.com/tr50/primesense/> (accessed on 28 October 2019).
12. Zennaro, S.; Munaro, M.; Milani, S.; Zanuttigh, P.; Bernardi, A.; Ghidoni, S.; Menegatti, E. Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications. In Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy, 29 June–3 July 2015; pp. 1–6.
13. Herrera, D.; Kannala, J.; Heikkilä, J. Joint depth and colour camera calibration with distortion correction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2058–2064. [[CrossRef](#)]
14. Konolige, K.; Mihelich, P. Technical Description of Kinect Calibration. 2011. Available online: http://www.wiki.ros.org/kinect_calibration/technical (accessed on 8 January 2019).
15. Smisek, J.; Jancosek, M.; Pajdla, T. 3D with Kinect. In *Consumer Depth Cameras for Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 3–25.
16. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
17. Zhang, C.; Zhang, Z. Calibration between depth and colour sensors for commodity depth cameras. In *Computer Vision and Machine Learning with RGB-D Sensors*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 47–64.
18. Canessa, A.; Chessa, M.; Gibaldi, A.; Sabatini, S.P.; Solari, F. Calibrated depth and colour cameras for accurate 3D interaction in a stereoscopic augmented reality environment. *J. Vis. Commun. Image Represent.* **2014**, *25*, 227–237. [[CrossRef](#)]
19. Darwish, W.; Tang, S.; Li, W.; Chen, W. A new calibration method for commercial RGB-D sensors. *Sensors* **2017**, *17*, 1204. [[CrossRef](#)] [[PubMed](#)]
20. Chow, J.C.; Lichti, D.D. Photogrammetric bundle adjustment with self-calibration of the PrimeSense 3D camera technology: Microsoft Kinect. *IEEE Access* **2013**, *1*, 465–474. [[CrossRef](#)]
21. Darwish, W.; Tang, S.; Li, W.; Chen, W. Full Parameter Calibration for Low Cost Depth Sensors. In Proceedings of the Melaha 2016 International Conference and Exhibition, Cairo, Egypt, 24–26 October 2016.
22. Darwish, W.; Li, W.; Tang, S.; Chen, W. Coarse to fine global RGB-D frames registration for precise indoor 3D model reconstruction. In Proceedings of the 2017 International Conference on Localization and GNSS (ICL-GNSS), Nottingham, UK, 27–29 June 2017; pp. 1–5.
23. Mallick, T.; Das, P.P.; Majumdar, A.K. Characterizations of noise in Kinect depth images: A review. *IEEE Sensors J.* **2014**, *14*, 1731–1740. [[CrossRef](#)]
24. Zhang, Z. Microsoft kinect sensor and its effect. *IEEE Multimed.* **2012**, *19*, 4–10. [[CrossRef](#)]
25. Zhang, Z. Flexible camera calibration by viewing a plane from unknown orientations. In Proceedings of the 7th ICCV, Kerkyra, Greece, 20–27 September 1999; pp. 666–673.
26. Moreno, D.; Taubin, G. Simple, accurate, and robust projector-camera calibration. In Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, Zurich, Switzerland, 13–15 October 2012; pp. 464–471.
27. Hübner, F.; Johnson, C.; Lartigue-Korinek, S.; Wang, G.; Buseck, P.; Hütch, M. Calibration of projector lens distortions. *J. Electron Microsc.* **2005**, *54*, 181–190. [[CrossRef](#)] [[PubMed](#)]
28. Kimura, M.; Mochimaru, M.; Kanade, T. Projector calibration using arbitrary planes and calibrated camera. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–2.
29. Qualisys. Miqus Technical Specifications. Available online: https://cdn-content.qualisys.com/2019/07/PI_Miqus_0.pdf (accessed on 8 January 2019).
30. Occipital. Structure Sensor Sepcification. Available online: <https://structure.io/structure-core/specs> (accessed on 8 January 2019).

31. Brown, D.C. Decentering distortion of lenses. *Photogramm. Eng. Remote Sens.* **1966**, *32*, 444–462.
32. Zhang, S.; Huang, P.S. Novel method for structured light system calibration. *Opt. Eng.* **2006**, *45*, 083601.
33. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
34. Lachat, E.; Macher, H.; Landes, T.; Grussenmeyer, P. Assessment and calibration of a RGB-D camera (Kinect v2 Sensor) towards a potential use for close-range 3D modeling. *Remote Sens.* **2015**, *7*, 13070–13097. [[CrossRef](#)]
35. Schmalz, C. Robust Single-Shot Structured Light 3D Scanning. Ph.D. Thesis, Friedrich–Alexander University Erlangen–Nürnberg, Eralngena and Nuremberg, Bavaria, Germany, 2012.
36. Gill, P.E.; Murray, W. Algorithms for the solution of the nonlinear least-squares problem. *SIAM J. Numer. Anal.* **1978**, *15*, 977–992. [[CrossRef](#)]
37. Mindru, F.; Moons, T.; Van Gool, L. Recognizing colour patterns irrespective of viewpoint and illumination. In Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), Fort Collins, CO, USA, 23–25 June 1999; pp. 368–373.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).