IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 19, 2018, accepted September 23, 2018, date of publication October 8, 2018, date of current version October 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2874205

A Smart Post-Rectification Algorithm Based on an ANN Considering Reflectivity and Distance for Indoor Scenario Reconstruction

JICHAO JIAO[1], (Member, IEEE), LIBIN YUAN[1], ZHONGLIANG DENG[1], (Member, IEEE), CHENG ZHANG[1], WEIHUA TANG[2], QI WU[1], AND JIAN JIAO[1]

[1]School of Electronic Engineering, Beijing University of Posts and Telecommunications , Beijing 100876, China
[2]China State Construction Engineering Corporation Ltd, Beijing 100029, China

Corresponding author: Jichao Jiao (jiaojichao@bupt.edu.cn)

This work was supported by the National Key Research and Development Program under Grant 2017YFB0503701.

ABSTRACT Indoor scene reconstruction is important for robot positioning and navigation in scenario reconstruction, especially in constructing a semantic map. In previous research, RGB-D cameras have been utilized to obtain a semantic map. However, because of indoor objects and depth sensors, the accuracy and precision of the depth values could be improved, which is a key factor in reconstructing indoor scenarios. Moreover, there is a relationship between reflectivity and depth accuracy. Therefore, to obtain depth information that is better than that obtained in our previous research, we present a smart post-rectification algorithm based on an artificial neural network (ANN). The algorithm improves the accuracy and precision of depth values by simultaneously considering reflectivity, distances, and different mechanisms of measuring depth. First, we analyze the RGB-D cameras' characteristics, including the pinhole camera model, lens distortions, and the types of error factors due to the types of RGB-D cameras used. Then, this paper proposes a smart post-rectification algorithm for depth images based on an ANN considering the depth error caused by reflectivity, the distance-related depth error, and different mechanisms for measuring depth. Finally, we perform experiments to evaluate the accuracy and precision of the proposed post-rectification approach by using different types of depth sensors. To evaluate the performance of our proposed algorithm, the proposed approach is applied to RGB-D SLAM, which is tested in different indoor environments. The experimental results show that applying our post-rectification algorithm to indoor scenario reconstruction can result in more accurate and more detailed 3-D reconstruction of objects than other state-of-the-art methods, highlighting the robustness and efficiency of our proposed algorithm.

INDEX TERMS Artificial neural network, depth sensor, RGB-D cameras, reflectivity, SLAM system.

I. INTRODUCTION

Precision three-dimensional (3D) indoor scenario reconstruction is an essential process for robot positioning [2], unmanned aerial vehicle navigation [3], [4] and semantic mapping [5], [6]. To date, a series of technologies have been developed for 3D indoor scene reconstruction. For example, 3D LiDAR is used in the 3D reconstruction of building facades [7], stereo cameras are used in real-time suited precise 3D environmental reconstruction [8], and RGB-D cameras are used in close-range 3D modeling [9]. Lee [10] used an expensive computational procedure to reconstruct 3D indoor scenes with low accuracy. Even though a 3D LiDAR could provide accuracy and robustness, it is too expensive for most researchers. In addition, the lack of important color information is a shortcoming of using a 3D LiDAR [11]. Recently, the RGB-D camera has become a better option than other equipment for the 3D reconstruction of indoor scenarios.

The RGB-D camera is a new type of sensing device that can capture RGB images and the corresponding depth image pixel-by-pixel. RGB-D cameras depend on either structured-light (SL) or time-of-flight (ToF) technology to collect depth data [12]. Recently, several types of RGB-D camera products, including the Microsoft Kinect v1 [13], Asus Xtion Pro [14], Occipital Structure Sensor [15], Intel RealSense [16] and Microsoft Kinect v2 [13], [17], have been produced. The Kinect v1, Asus Xtion Pro, Occipital Structure Sensor, and Intel RealSense adopt an SL depth

sensing mechanism [18], while the Kinect v2 employs ToF technology to sense depth values. Moreover, based on ToF technology, the Kinect v2 has other improved SL sensors with a higher color camera resolution and the ability to operate outdoors [19], [20]. Recently, due to improved accuracy and robustness, the Kinect v2 has become an increasingly popular RGB-D camera [1], [9], [19], [20]. In our previous work [1], we proposed a post-rectification method for depth images of Kinect v2, which is a ToF-based sensor used for 3D indoor reconstruction. Furthermore, based on our previous method [1], a more universal and smarter algorithm for RGB-D cameras is proposed in this paper.

There are great expectations that RGB-D systems will boost new 3D perception-based applications in the fields of robotics and visual/augmented reality.

Furthermore, to capture high-quality color images, either SL or ToF technology is introduced to capture depth in images. Therefore, using an RGB-D camera, many researchers have conducted related studies for 3D reconstruction to improve accuracy, precision, and robustness. To this end, the calibration of RGB-D cameras plays an important role. However, only a few off-the-shelf calibration approaches for RGB-D cameras and methods are available for processing color and depth in images that consider reflectivity as a key factor influencing accuracy and precision in 3D indoor scenario reconstruction when different types of sensors are used [21]– [27].

Therefore, our contributions in this paper are as follows:

(1) Introduction of an ANN for a fast and robust rectification model, rendering our proposed method smarter than other types of methods, including our method proposed in [1].

(2) Verification that the between among the reflectivity-related depth error, the distance-related depth error and the different mechanisms of measuring depth are nonlinear. A nonlinear relationship was identified in our previous work, which resulted in an improvement in the accuracy and precision of depth values.

This paper is organized as follows. Section II presents related studies investigating the properties and use of RGB-D cameras for 3D indoor reconstruction. Then, we provide a comprehensive analysis of the characteristics of RGB-D cameras in Section III. In Section IV, we propose a smart post-rectification approach for depth images based on reflectivity, distance and different mechanisms for measuring depth and ANN in detail. Then, experiments and experimental results are reported and discussed in Section V. Finally, Section VI concludes the paper.

II. RELATED WORK

In our previous research, we compared the performances of the Kinect v1 and v2 [12], [22], [28]. Because of their great performance, Kinect sensors are widely used in 3D reconstruction [9], [21], [29], [30] and mobile robot navigation [31].

Using different technologies to obtain depth information, many researchers have performed many studies to identify a reasonable Kinect sensor for reconstructing a 3D model. In [12], a detailed comparison of the two versions of Kinect sensors was performed. The authors provided a comprehensive analysis of the factors (including the reflectivity-related factor) that resulted in depth information errors. According to their research, the Kinect v2 performed better than the Kinect v1 in reducing the systematic error in the distance and being insensitive to illumination changes. Moreover, Gonzalez-Jorge *et al.* [32] tested the accuracies of different Kinect sensors at different distances and showed that the Kinect v2, which used ToF technology, can achieve better performance in depth accuracy and precision than the Kinect v1. Nevertheless, accuracy and precision can be further improved by considering reflectivity-related depth errors.

Lindner and Kolb [33] and Lindner *et al.* [34] proposed an approach for calibrating the intensity-related distance error of ToF cameras. A special planar checkerboard pattern with different stripes was used in their experiment, which inspired us to design our evaluation study. Wasenmüller and Stricker [28] proposed a principle for calculating depth using a ToF camera.

Rodríguez-Gonzálvez *et al.* [35] proposed a radiometric calibration function to display the relationship between the depth sensor of the Kinect v2 and reflectivity. This function can transform digital values into physical values. According to [1], the quality of this approach is appropriate for exploiting the radiometric possibilities of low-cost depth sensors used in agriculture and forestry. However, the authors only focused on the Kinect v2 and did not consider other RGB-D cameras used in indoor applications.

Yu *et al.* [36] proposed a shading-based shape refinement algorithm that uses a noisy, incomplete depth map from the Kinect to obtain a high-quality 3D surface reconstruction. However, for reflectance, the authors used mean-shift clustering to segment RGB images into small areas with a uniform albedo. This method is more complex than our method. Moreover, 3D surface reconstruction is only qualitative research without quantitative calculations. Therefore, we carry out both qualitative and quantitative studies. Han *et al.* [37] proposed a shading-based approach for shape refinement of an RGB-D image. However, this approach still requires explicit image segmentation for handling multi-albedo objects. The approach presented in our paper is simpler and smarter than previous approaches.

Kim *et al.* [38] refined depth sensing using a shading analysis. These authors assumed that neighboring pixels have a locally similar reflectance that includes the smoothness constraint of reflectance. However, such a hypothesis is slightly limiting. We do not make an assumption regarding reflectance in our paper to obtain a better result.

Yang *et al.* [24] proposed a novel framework to recover depth maps from low-quality measurements with various types of degradations, such as low resolution, noise, and missing depth in some areas. However, the authors did not consider reflectivity-related depth error or distance-related depth error.

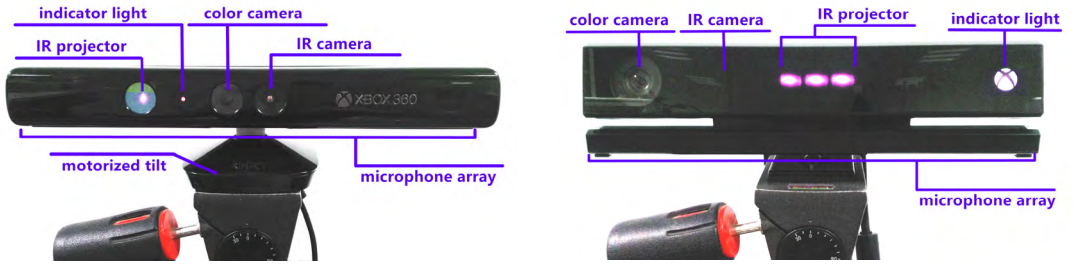


FIGURE 1. Two different structures of Microsoft Kinect: (a) Kinect v1 and (b) Kinect v2.

TABLE 1. Technical specifications of two different TYPEs of RGB-D cameras (Kinect v1 and Kinect v2).

Feature	Kinect v1	Kinect v2
Color Camera	640 × 480 pixels @ 30 fps	1920 × 1080 pixels @ 30 fps
Depth Camera	320 × 240 pixels @ 30 fps	512 × 424 pixels @ 30 fps
Infrared Operating Wavelength	830 nm	860 nm
Horizontal Field of View (depth)	57 degrees	70 degrees
Vertical Field of View (depth)	43 degrees	60 degrees
Operative Measuring Range	from 0.8 m to 4.0 m	from 0.5 m to 4.5 m
Depth Sensing Mechanism	structured-light (SL)	time-of-flight (ToF)
Tilt Motor	yes	no
USB Standard	2.0	3.0

Based on a multi-scale sparse representation, a data-driven depth map refinement method was presented by Kwon *et al.* [39]. This method requires the use of a corresponding training set for specific object classes and is not suitable for the 3D reconstruction of indoor scenes.

III. RGB-D SENSOR PRESENTATION

A. CHARACTERISTICS OF RGB-D CAMERAS

In addition to capturing color images, RGB-D cameras employ SL or ToF technology to provide depth images simultaneously.

Two RGB-D cameras are evaluated, i.e., the Kinect v1, which is a representative first-generation RGB-D camera based on SL technology, and the Kinect v2, which is a representative RGB-D camera based on ToF technology. The hardware structures of these two RGB-D cameras are shown in Fig. 1, and a comparison of their technical specifications is provided in Table 1 [30], [32], [40].

B. RGB-D CAMERA MODEL

In the RGB-D camera system, the RGB camera captures 2D images, and an infrared camera is used to acquire depth information. Generally, the pinhole model is used to convert a real-world scenario into a 2D image in these two types of cameras [12]. Thus, the system builds a mapping relationship between a location in the three-dimensional world and a two-dimensional image pixel.

Calibrating the RGB-D camera before use is necessary. Therefore, we can obtain the intrinsic and extrinsic parameters of the camera [29].

Before using the pinhole model, notably, two types of camera lens distortions occur when capturing images: radial distortion and tangential distortion [29].

Similarly, when calibrating an RGB-D camera, we can also obtain a set of distortion coefficients to be used for accurate color and depth data acquisition.

C. DEPTH SENSOR L

In an RGB-D camera with SL technology, an infrared light source projects a dot pattern onto a scene, and an offset infrared camera receives the pattern and estimates the depth value. However, the depth value is determined by measuring the phase difference between emitted and reflected light in an RGB-D camera with ToF technology. Regardless of whether a camera is equipped with SL or ToF technology, various error factors can affect the depth values of RGB-D cameras. TABLE 2 lists the factors that influence performance in detecting depth values when implementing Kinect sensors [12], [28], [30].

Several factors (except for the Flying Pixel) are well known to influence both SL- and ToF-based cameras during depth

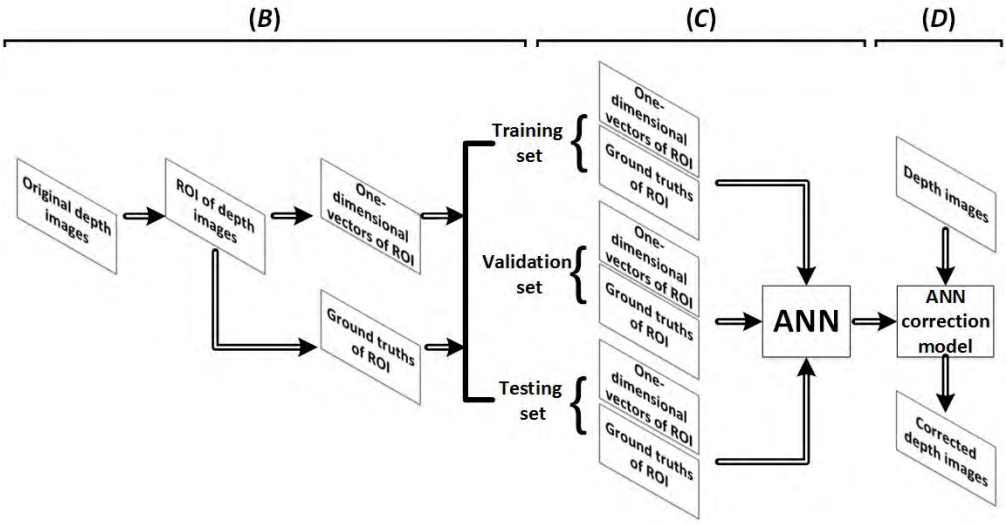


FIGURE 2. Overview of the smart post-rectification approach: (*B*) preparation of the stripe plane pattern (SPP) and data acquisition; (*C*) training of the ANN model; (*D*) correcting depth images through an ANN correction model.

TABLE 2. Error factors IN Kinect V1 and V2.

Error Factor \ Depth Sensor	v1	v2
Ambient Background Light [12]	√	√
Temperature Drift [12, 28, 30]	√	√
Systematic Error [12, 30]	√	√
Distance-Related Depth Error [28]	√	√
Depth Inhomogeneity [12, 30]	√	√
Multi-Path Effects [12, 28, 30]	√	√
Flying Pixel [12, 28, 30]	×	√
Reflectivity-Related Depth Error [12, 28]	√	√
Semitransparent and Scattering Media [12, 30]	√	√
Dynamic Scenery [30]	√	√

evaluation. In [12], two possible explanations are provided for reflectivity-related depth error: the multi-path of the effect and a nonlinear pixel response because of the low illumination change in indoor scenarios. In Section IV, a method for reducing the reflectivity-related depth error and distance-related depth error and the effect of different mechanisms of measuring depth using RGB-D cameras is proposed.

IV. METHOD

In this section, we propose a novel method for correcting the depth images of RGB-D cameras based on an artificial neural network (ANN). An overview is illustrated in Fig. 2.

A. CHARACTERISTICS OF THE ANN

Currently, with the rapid development of ANNs, we utilize an ANN to extract integrated features by considering multiple factors simultaneously. In this paper, we consider three different error factors, including different depth image sensing mechanisms, different distances between objects and RGB-D cameras and different reflectivity. These factors are considered simultaneously based on an ANN.

According to [42], a major benefit of ANNs is their flexibility in modeling the nonlinearity of independent variables. Linear regression techniques are very common in statistical data analysis because they can extract information based only on linear models, which can be a limitation in real data contexts. Based on the same selected variables, an ANN is employed to improve the prediction of the linear model, taking advantage of the nonlinear modeling capabilities. Therefore, this paper presents a smart post-rectification algorithm for building a nonlinear regression model using an ANN.

B. PREPARATION OF THE STRIPE PLANE PATTERN (SPP) AND DATA ACQUISITION

1) PREPARATION OF THE SPP

To build a model of the relationship between reflectivity and depth values captured by different types of sensors, we designed and utilized a striped plane pattern with six different gray levels. The gray values of the plane panel were divided into six levels; the reflectivity at each level is shown in Table 3.

2) DATA ACQUISITION

First, before data acquisition, we calibrated the Kinect v1 and v2 [23], [43]. We captured depth images at different distances within the effective scope of the measuring range. The region

TABLE 3. The expected offset of six TYPES of reflectivity [1].

Panel Level	Reflectivity
100%	0.0011
80%	0.1994
60%	0.4125
40%	0.6019
20%	0.7981
0%	0.9913

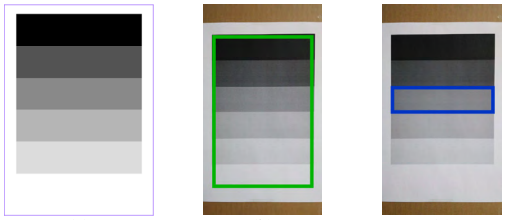


FIGURE 3. Striped plane pattern with different grayscales: (a) striped plane pattern, (b) green rectangle area, and (c) blue rectangle area.

of interest (ROI) in each original depth image to be studied is indicated by a green rectangle in Fig. 3b.

A calibrated RGB-D camera was fixed to a stable photographic tripod, and the SPP is fixed to another photographic tripod. The front panel of the RGB-D camera and the SPP were consistently parallel as shown in Fig. 4 (Kinect v2 as an example). The SPP was gradually moved away from the RGB-D camera within the effective scope of the measuring range at a step length of s m. Notably, s was usually no greater than 0.05 m. Furthermore, we captured L sets of depth images in the green rectangular area. The value L was determined by the size of the value of the operative measuring range and the step length. One set contained N depth images. The operative measuring range is based on the type of RGB-D camera and the practical application environment. Furthermore, smaller steps and greater numbers of depth images in one set are helpful for the method presented in this paper.

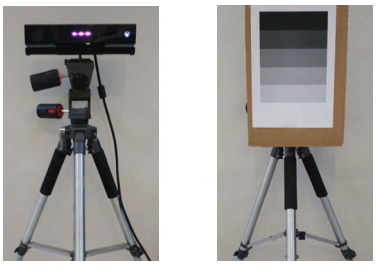


FIGURE 4. Setup of the method: RGB-D camera (Kinect v2 as an example) (a) and striped plane pattern (b).

Then, we obtained the original depth images. Subsequently, the ROI in each original depth image was processed into a one-dimensional vector. The ground truths of the ROIs were the corresponding distances. Therefore, we performed the entire data acquisition of the depth images with one-dimensional vectors of ROIs and their ground truths in pairs.

Finally, to train and achieve a better ANN-based correction model, the entire data acquisition of depth images was divided into the following 3 parts as shown in Fig. 2: a training set, a validation set, and a testing set.

C. TRAINING OF THE ANN MODEL

The training process was carried out as follows.

The ANN model used in our work was a multi-layer perceptron (MLP) [42], which is a feed-forward neural network used for mapping sets of input data onto a set of appropriate outputs. MLP is characterized by $L+2$ layers of neurons (input layer-1 layer, hidden layers-L layers, and output layer-1 layer) with nonlinear activation functions at the hidden layer units. To indicate the nonlinearity between the different influencing factors and reflectivity, a feed-forward MLP was used for the nonlinear mapping of the influencing factors (x) into a single predicted value y (shown in Fig. 5).

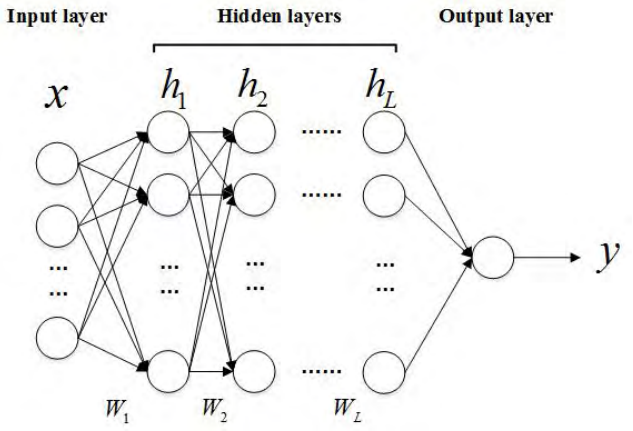


FIGURE 5. Nonlinear calculation for the ANN model.

In the MLP (as shown in Fig. 5), the input layer consisted of the one-dimensional vectors of ROIs in the original depth images and their ground truths; the hidden layers were characterized by hidden neurons with the rectified linear unit function; and the output layer was composed of only one output neuron (the nonlinear value y). The number of hidden neurons was determined through a trial-and-error process following the general principle of parsimony because no commonly accepted theory for determining the optimal number of neurons in hidden layers exists. In detail, we have designed and trained the model based on different number of neurons in the hidden layers. At last, we fix the network with scale as 6 Layers and [20, 50, 100, 50] neural units for hidden layers.

The input variable vector x was mapped to the neurons in the hidden layers as follows:

$$h_i = ReLU(W_i \cdot x + b_i), \quad (i = 1) \tag{1}$$

$$h_i = ReLU(W_i \cdot h_{i-1} + b_i), \quad (i = 2, 3, \ldots, L) \tag{2}$$

where h_i is the output value of layer i. L is the number of hidden layers, W_i is the weight matrix between the former

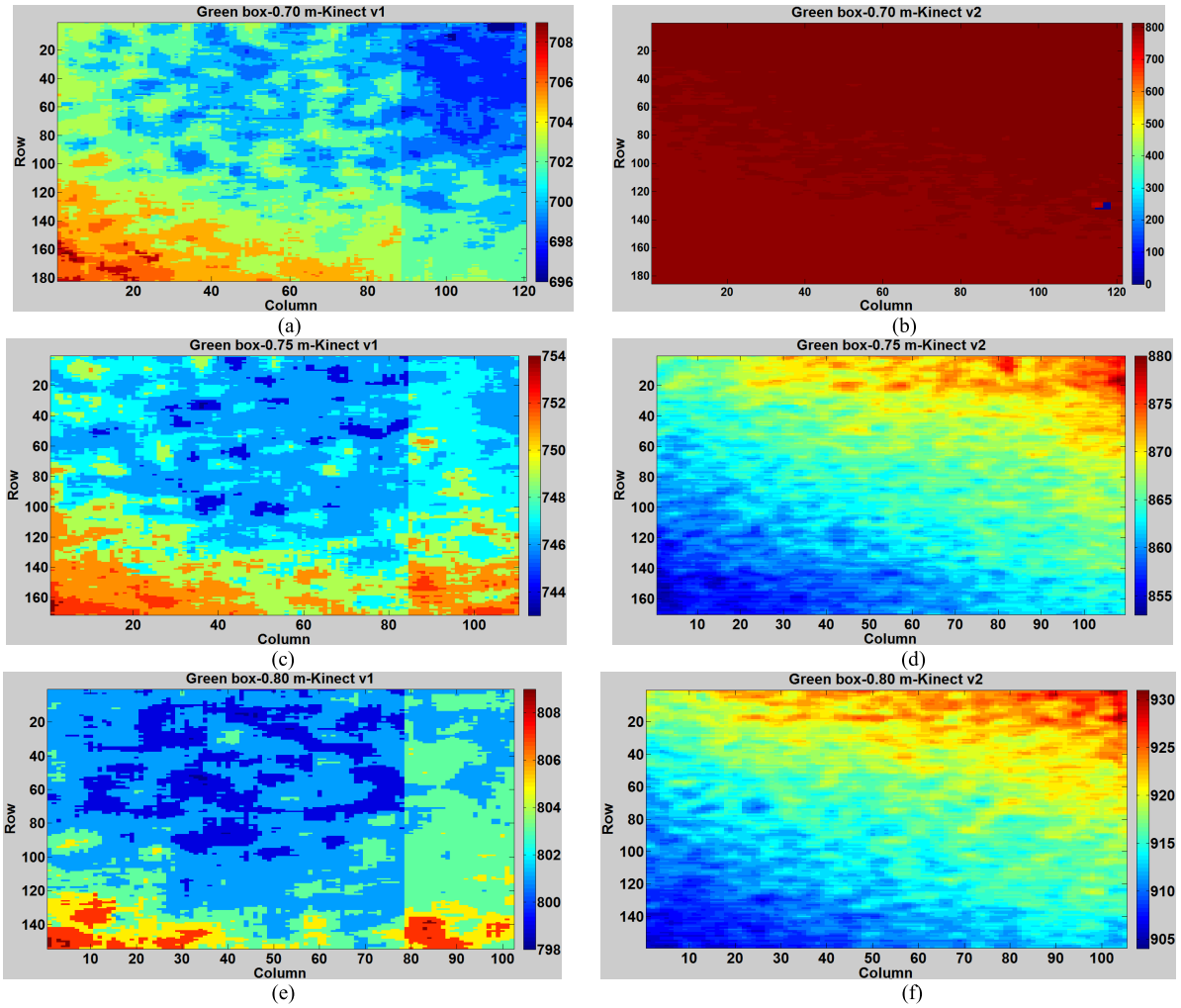


FIGURE 6. Visualization of depth images at different distances using Kinect v1 (left) and v2 (right).

layer *i-1*and the current layer i and b_i is the bias parameter vector of the former layer *i-1*and the current layer i. The value y represents the distance obtained from the depth images. The value y is the output of each sample which is obtained from a linear combination of the hidden neurons vector h_i as follows:

$$\begin{aligned} y &= f(x; W) \\ &= h_L(h_{L-1}(\cdots h_2(h_1(x; w_1); w_2); \cdots ; w_{L-1}); w_L) \end{aligned} \quad (3)$$

Finally, the cost function was calculated as follows:

$$Loss = \frac{1}{n}\sum_{j=1}^{n}[y_j - y_j^{gt}]^2 = \frac{1}{n}\sum_{j=1}^{n}[f(x_j; W) - y_j^{gt}]^2 \quad (4)$$

where *Loss* is the cost function of the training set, validation set or testing set; n is the number of samples; y_j is the output value of sample j, and y_j^{gt} is the ground truth of sample j.

Equation (4) shows the average error between the predicted value and the ground truth. The *AveError* is utilized for tuning the model.

$$AveError = \sqrt{Loss} \quad (5)$$

D. CORRECTING DEPTH IMAGES THROUGH AN ANN CORRECTION MODEL

After training the ANN, we established an ANN correction model. First, we captured the original depth images with an RGB-D camera based on either SL or ToF technology. Then, the original depth images were entered into the ANN correction model. Finally, the corrected depth images were obtained. Therefore, the corrected depth images with the corresponding color images can be used to obtain a better 3D reconstruction.

V. EXPERIMENTS AND RESULTS

We performed our experiments in our laboratory, which measures 60 m^2, using static illumination. We preheated the RGB-D camera for one hour to eliminate the effect of temperature drift [9], [12], [21]. Notably, we performed our experimental work in a setting similar to that reported in [1] to compare the ANN-based algorithm to our previously proposed method.

Because accuracy and precision [22], [28] decrease as the measurement range increases, the indoor environment

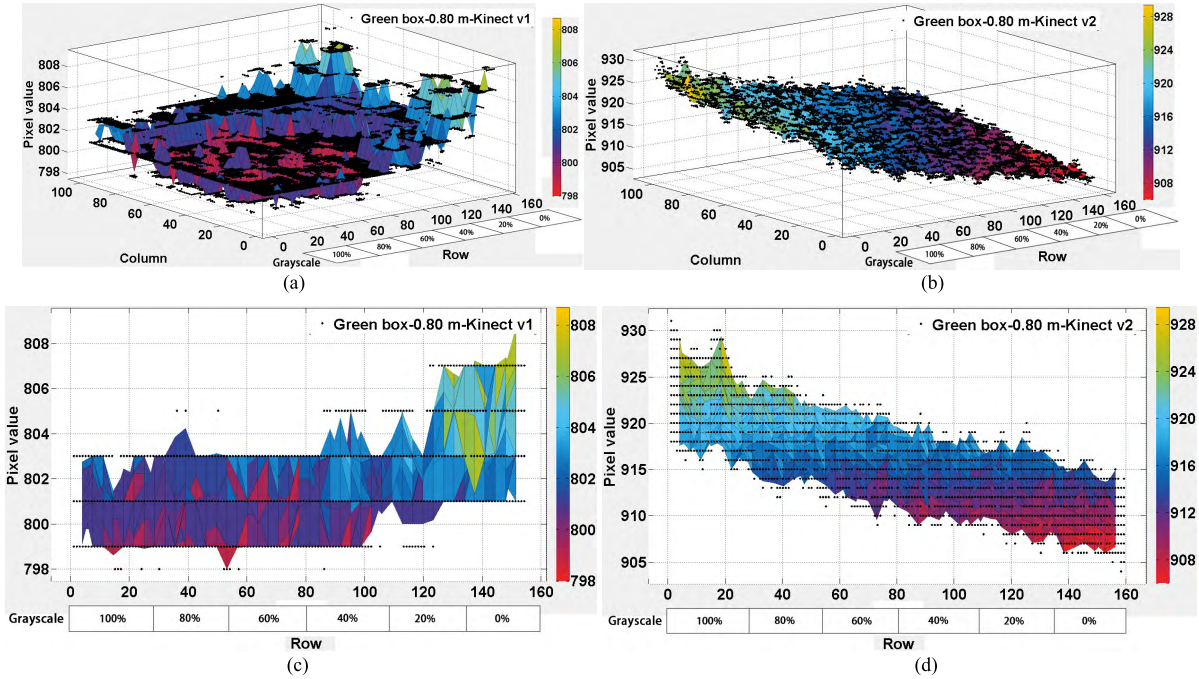


FIGURE 7. Green rectangle area at a distance of 0.80 m using Kinect v1 and Kinect v2: 3D view (top) and side view (bottom).

is usually very small. Therefore, the operative measurement range in our experiment was less than 2 meters. The tripod with the SPP was gradually moved away from the tripod with the RGB-D camera from 0.60 meters to 2 meters in steps of 0.05 meters. Ultimately, 1450 images were captured to construct a database in our experiment.

The green rectangle area's visualized depth images are presented in Figs. 6a–f at the preceding few distances. The color bar represents the pixel values of the depth images. As shown in Fig. 6b, the high-reflectivity areas occurred at the bottom of the rectangle box, and a pixel value of 0 (invalid null value) was easily obtained at a short distance when the Kinect v2 was utilized. However, no invalid null value (pixel value 0) was obtained in the depth image when using the Kinect v1. In summary, in our indoor environment, the minimum operative measuring distance using the Kinect v1 was shorter than the official value of 0.80m, while the minimum operative measuring distance using the Kinect v2 was between 0.70 m and 0.75 m [1].

Therefore, we captured the original depth images at distances of 0.75 m to 2.0 m with the same step width of 0.05 m. We considered the Kinect v1 and v2 simultaneously to avoid the effects of invalid null values and conducted reliable studies.

A. SAME DISTANCE BUT DIFFERENT REFLECTIVITY: REFLECTIVITY-RELATED DEPTH ERROR

Fig. 3b demonstrates that the different reflectivities in the green rectangle on the SPP significantly affected the measured depth value at the same distance. As shown in Fig. 7, as reflectivity increased, the depth value using the Kinect v1 increased, while the value obtained using the Kinect v2 was smaller at a distance of 0.80 m. The relationships between the different reflectivities and measured depth values at a distance of 0.80 m in the front of the depth sensor are illustrated in Fig. 6e and Fig. 6f. Furthermore, the bottom panel in Fig. 7 clearly indicates the side view at a distance of 0.80 m. Therefore, the depth values captured by the ToF-based RGB-D sensor (Kinect v2) and the SL-based sensor (Kinect v1) differed at the same distance under different reflectivity conditions.

B. SAME REFLECTIVITY BUT DIFFERENT DISTANCES: DISTANCE-RELATED DEPTH ERROR

At the same reflectivity, i.e., grayscale of 60% as shown in Fig. 3c, the relationship between the measured depth value and different distances was analyzed. Fig. 8 shows the experimental results. The expectation was evaluated using the fluctuation between the metrical depth and the ground truth. The standard deviation (**Std**) represents the standard deviation of the depth information difference between the corrected depth value and the ground fluctuation between the measured depth value and the ground truth. The depth accuracy of an RGB-D camera is evaluated using the expectation, and the depth precision is assessed by the **Std** [28], [32]. Therefore, as shown in Fig. 8, at the same reflectivity but under different distance conditions, the depth accuracy varied nonlinearly at different distances using both the Kinect v1 and v2. The same phenomenon occurred with respect to depth precision.

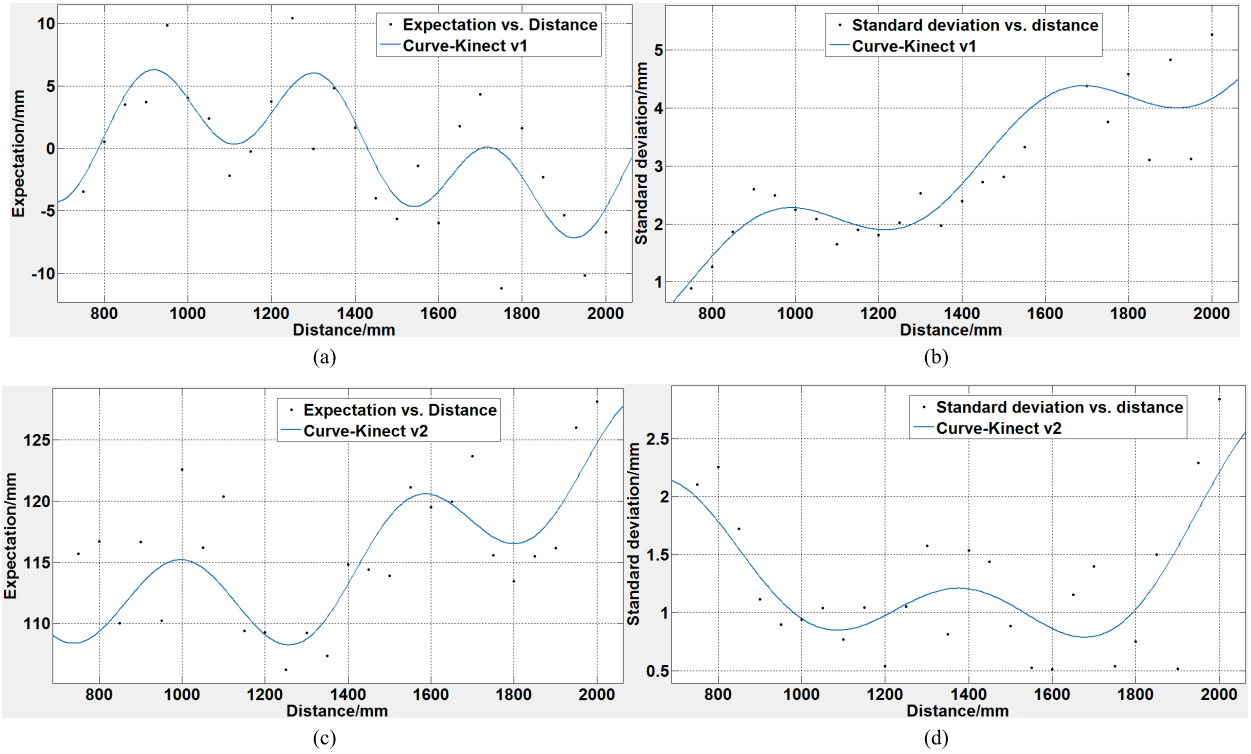


FIGURE 8. Expectation and standard deviation of the depth value difference at the same reflectivity and different distances using Kinect v1 (a, b) and Kinect v2 (c, d).

TABLE 4. The experimental results of the testing set for the Kinect v1.

Number / Value (mm) / Item	1	2	3	4	5	...	256	257	258	259	260
Ground truth	1150	750	1900	1950	1800	...	1600	1650	1750	1650	750
Original value	1158	741	1887	1935	1809	...	1587	1665	1736	1662	742
Predicted value	1152	746	1894	1939	1805	...	1594	1655	1740	1656	747

C. TRAINING THE ANN MODEL AND RECTIFYING DEPTH IMAGES USING KINECT V1 AND KINECT V2

Following Section IV, we obtained one-dimensional vectors of ROIs and the ground truths of the ROIs in the original depth images of the Kinect v1 and Kinect v2 at a distance of 0.75 m to 2.0 m. Then, we divided the data into the following 3 datasets: a training set, a validation set and a testing set (60%, 20%, and 20%, respectively).

The RMSprop algorithm was used to train our proposed ANN-based model by optimizing the multinomial logistic regression objective. Inspired by Simonyan and Zisserman [44], for the Kinect v1, the batch size was set to 128, and the learning rate was initially set to 10^{-3} but was decreased by a factor of 10 because the validation set accuracy stopped improving. Inspired by Simonyan and Zisserman [44], for the Kinect v2, the batch size was set to 256, and the learning rate was initially set to 10^{-2} but was decreased by a factor of 10 because the validation set accuracy stopped improving. The learning process was stopped after 390K iterations (300 epochs) for both the Kinect v1 and v2. The number of hidden layers for both the Kinect v1 and v2 was 3. This process was carried out to continuously adjust the various parameters to obtain suitable results. The training loss and validation loss are illustrated in Fig. 9. Because the size of the entire dataset used in this paper was 1300, according to the above distribution ratio, the testing set contained 260 data points. Finally, the experimental results of the testing set are shown in TABLE 4 and TABLE 5.

TABLE 5. The experimental results of the testing set for the Kinect v2.

Number / Value (mm) / Item	1	2	3	4	5	...	256	257	258	259	260
Ground truth	1300	1550	800	1800	1150	...	950	1550	1350	1300	850
Original value	1410	1671	917	1914	1259	...	1059	1671	1457	1409	959
Predicted value	1303	1544	807	1790	1155	...	960	1545	1347	1304	857

TABLE 6. The expectation and standard deviation of the difference between the original value/predicted value and the ground truth.

		Expectation (mm)	Standard deviation (mm)
Kinect v1	Before rectification	-2.2846	9.8712
	After rectification	-1.0692	5.8556
Kinect v2	Before rectification	-115.5654	5.6384
	After rectification	0.9962	5.4298

TABLE 7. Results of the comparison of our previous method and our proposed method.

Method	Accuracy (mm)		Precision (mm)		Average correction time(ms/frame)	
	v1	v2	v1	v2	v1	v2
The method in [32]	25	5	12	8	29	29
Our previous method	-27.00	3.00	15.00	6.00	38	38
Our proposed method	-1.0692	0.9962	5.8556	5.4298	33	33

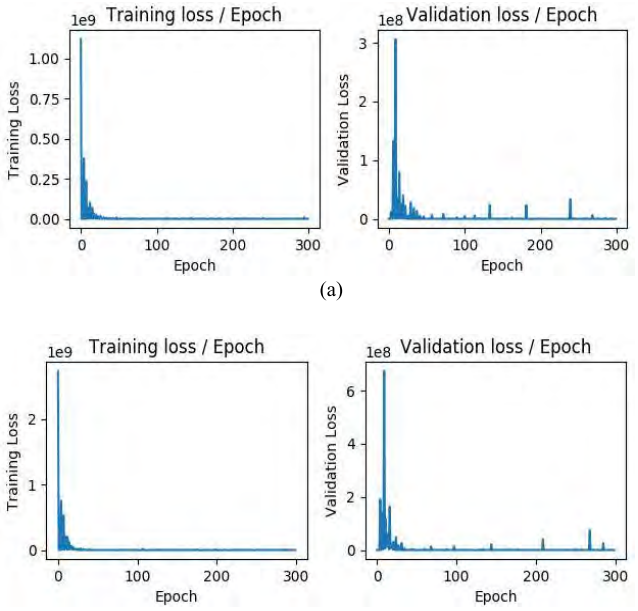


FIGURE 9. The relationship between the training loss/validation loss and the epochs for v1 (a) and v2 (b).

By analyzing the data in TABLE 4 and TABLE 5, we calculated the expectation and standard deviation of the difference between the original value/predicted value and the ground truth to evaluate the depth accuracy and depth precision at distances of 0.75 m to 2.00 m. Therefore, the effects of reflectivity and distance errors were considered in the experimental results shown in TABLE 6. Using this smart post-rectification-based ANN, for the Kinect v1, the depth accuracy increased by 1.2 millimeters, and the depth precision increased by 4 millimeters. Simultaneously, for the Kinect v2, the depth accuracy increased by 114.6 millimeters, and the depth precision increased by 0.2 millimeters. Furthermore, using the Kinect v1, the depth accuracy was 1.0692 mm, and the depth precision was 5.8556 mm, which are both higher than the depth accuracy (25 mm) and precision (12 mm) reported in [32]. Moreover, using the Kinect v2, the depth accuracy was 0.9962 mm, and the depth precision was 5.4298 mm, which are both better than the depth accuracy (5 mm) and precision (8 mm) reported in [32]. In addition, the depth accuracy and precision reported in [32] were obtained by a state-of-the-art rectification approach. Furthermore, we compared the results of our previous method with those of our proposed method and found that our proposed method surpasses our previous method in accuracy and precision. The results are shown in TABLE 7.

In addition, we performed a series of experiments to measure the actual correction time of our two methods and the method in [32] on a computer with Intel Core i5

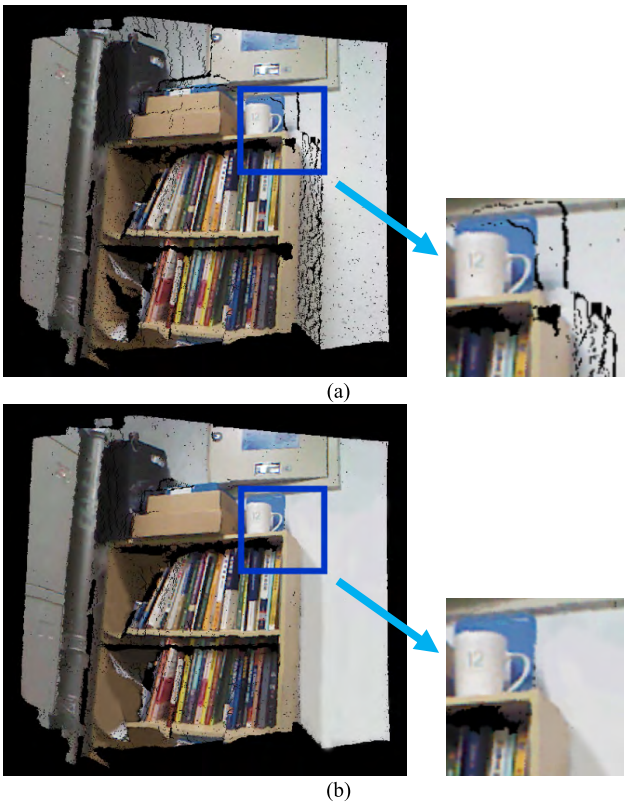


FIGURE 10. A point cloud of an indoor scene captured by Kinect v1 before (a) and after (b) correction using our proposed method.

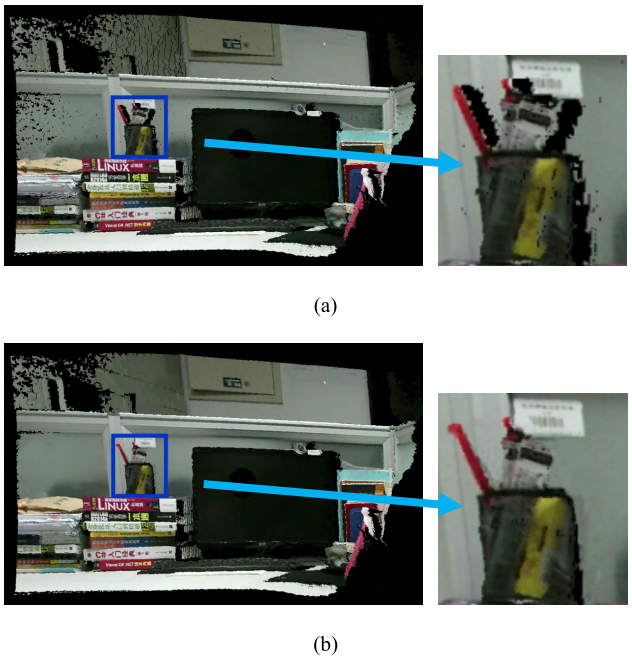


FIGURE 11. A point cloud of an indoor scene captured by Kinect v2 before (a) and after (b) correction using our proposed method.

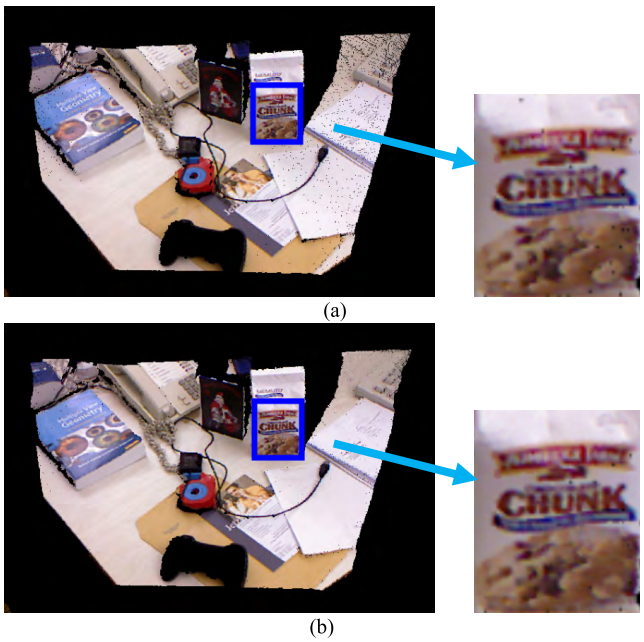


FIGURE 12. A point cloud of an indoor scene captured by Kinect v2 before (a) and after (b) correction using our proposed method.

3.2 GHz CPU, 16 GB of RAM and NVIDIA GeForce GTX 1060 6 GB (as shown in TABLE 7). Although the average correction time per frame of our proposed method is more than that of the method in [32], the accuracy and precision are better than the method in [32]. And the average correction time per frame, the accuracy and precision of our proposed method are much better than those of our previous method.

For RGB-D cameras, the depth values of depth images are simultaneously related to reflectivity, distances and different mechanisms of measuring depth. Therefore, the general methods can not get depth values of high accuracy and precision. However, the model trained by ANN can adapt to the scene well and get more accurate results than the previous methods.

D. APPLICATION TO RGB-D SLAM SYSTEM

To prove the effectiveness of the smart post-rectification algorithm based on the ANN, we implemented this method using RGB-D SLAM as proposed by Endres *et al.* [45] for indoor scenarios. We captured color and depth images using the following two different categories of cameras in our indoor scenarios: an SL camera (e.g., Kinect v1) and a ToF camera (e.g., Kinect v2). Based on the offline RGB-D SLAM approach, we perform 3D reconstructions of indoor scenes using uncorrected and corrected depth images with corresponding color images. Using this smart post-rectification method for depth images, which is described in detail in Section IV and Section V above, better visual effect 3D reconstructions were achieved, as illustrated in Fig. 10 and Fig. 11.

Furthermore, the smart post-rectification algorithm was applied to the RGB-D SLAM Dataset (captured by Kinect v1) [46] and the George Mason University Kitchen Dataset (captured by Kinect v2) [47]. Fig. 11 and Fig. 12 present the 3D point clouds before and after rectification using the smart

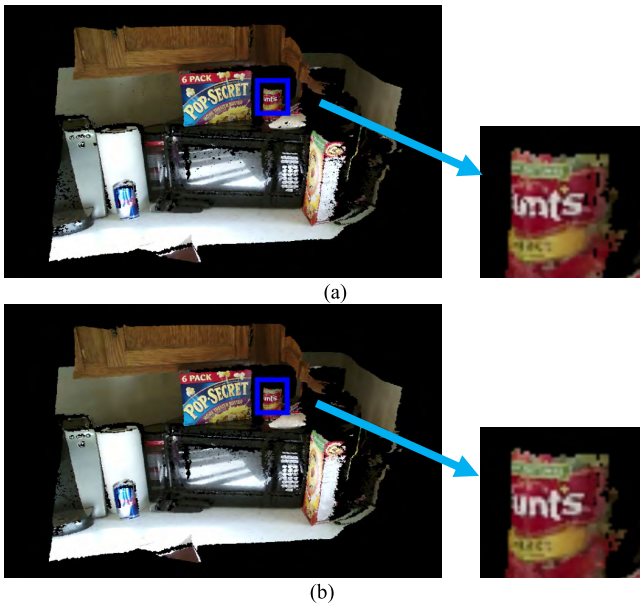


FIGURE 13. A point cloud of the GMU Kitchen Dataset before (a) and after (b) correction using our proposed method.

post-rectification algorithm. Obviously, because more accurate depth data were obtained after rectification, we could obtain a 3D reconstruction that could provide more important scenario information than that obtained without our proposed smart post-rectification algorithm.

VI. CONCLUSION

This paper systemically presents the relationship between the measured distances of depth images and reflectivity/distance in indoor scenarios using RGB-D cameras. Therefore, we propose a smart post-rectification algorithm for depth images based on an ANN considering the reflectivity-related depth error and distance-related depth error for indoor scenario 3D reconstruction. As demonstrated by the experimental results, using the Kinect v1, the depth accuracy is 1.0692 mm, and the depth precision is 5.8556 mm. Moreover, using the Kinect v2, the depth accuracy is 0.9962 mm, and the depth precision is 5.4298 mm. Therefore, the depth accuracy and depth precision using the Kinect v1 and Kinect v2 are better than the results reported in [1] and [32]. Finally, more accurate and precise depth images were utilized to obtain a better visual effect 3D reconstruction of indoor environments.

Applying the smart post-rectification algorithm to the reconstruction of 3D indoor scenarios in real time using an RGB-D camera will be considered in our further studies.

REFERENCES

[1] J. Jiao, L. Yuan, W. Tang, Z. Deng, and Q. Wu, "A post-rectification approach of depth images of Kinect v2 for 3D reconstruction of indoor scenes," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 11, p. 349, 2017.
[2] P. Koch *et al.*, "Multi-robot localization and mapping based on signed distance functions," *J. Intell. Robotic Syst.*, vol. 83, pp. 409–428, Sep. 2016.
[3] P. J. Zarco-Tejada, R. Diaz-Varela, V. Angileri, and P. Loudjani, "Tree height quantification using very high resolution imagery acquired from an unmanned aerial vehicle (UAV) and automatic 3D photo-reconstruction methods," *Eur. J. Agronomy*, vol. 55, pp. 89–99, Apr. 2014.
[4] A. S. Huang *et al.*, "Visual odometry and mapping for autonomous flight using an RGB-D camera," in *Robotics Research*. Cham, Switzerland: Springer, 2017, pp. 235–252.
[5] J. McCormac, A. Handa, A. Davison, and S. Leutenegger. (2016). "SemanticFusion: Dense 3D semantic mapping with convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1609.05130
[6] C. Zhao, L. Sun, and R. Stolkin. (2017). "A fully end-to-end deep learning approach for real-time simultaneous 3D reconstruction and material recognition." [Online]. Available: https://arxiv.org/abs/1703.04699
[7] L. Yang, Y. Sheng, and B. Wang, "3D reconstruction of building facade with fused data of terrestrial LiDAR data and optical image," *Optik–Int. J. Light Electron Opt.*, vol. 127, pp. 2165–2168, Feb. 2016.
[8] K.-D. Kuhnert and M. Stommel, "Fusion of stereo-camera and PMD-camera data for real-time suited precise 3D environment reconstruction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2006, pp. 4780–4785.
[9] E. Lachat, H. Macher, M.-A. Mittet, T. Landes, and P. Grussenmeyer, "First experiences with Kinect v2 sensor for close range 3D modelling," in *Proc. Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. XL-5/W4, pp. 93–100, Feb. 2015.
[10] D. Lee. *Optimizing Point Cloud Production From Stereo Photos by Tuning the Block Matcher*. Accessed: Jan. 18, 2018. [Online]. Available online: https://erget.wordpress.com/2014/05/02/producing-3d-point-clouds-from-stereo-photos-tuningthe-block-matcher-for-best-results/
[11] B. Wu *et al.*, "A graph-based approach for 3D building model reconstruction from airborne LiDAR point clouds," *Remote Sens.*, vol. 9, no. 1, p. 92, 2017.
[12] H. Sarbolandi, D. Lefloch, and A. Kolb, "Kinect range sensing: Structured-light versus time-of-flight Kinect," *Comput. Vis. Image Understand.*, vol. 139, pp. 1–20, Oct. 2015.
[13] Microsoft. Kinect. Accessed: Nov. 13, 2017. [Online]. Available: https://en.wikipedia.org/wiki/Kinect
[14] Asus. Asus Xtion Pro. Accessed: Nov. 15, 2017. [Online]. Available: https://www.asus.com/3D-Sensor/Xtion_PRO/
[15] Occipital. *Occipital Structure Sensor*. Accessed: Nov. 22, 2017. [Online]. Available: https://structure.io/
[16] Intel. *Intel RealSense*. Accessed: Dec. 10, 2017. [Online]. Available: https://software.intel.com/realsense
[17] Microsoft. *Kinect for Xbox One*. Accessed: Dec. 16, 2017. [Online]. Available: https://www.xbox.com/en-US/xbox-one/accessories/kinect
[18] P. Zanuttigh, G. Marin, C. Dal Mutto, F. Dominio, L. Minto, and G. M. Cortelazzo, "Operating principles of structured light depth cameras," in *Time-of-Flight and Structured Light Depth Cameras*. Cham, Switzerland: Springer, 2016, pp. 43–79.
[19] M. G. Diaz, F. Tombari, P. Rodriguez-Gonzalvez, and D. Gonzalez-Aguilera, "Analysis and evaluation between the first and the second generation of RGB-D sensors," *IEEE Sensors J.*, vol. 15, no. 11, pp. 6507–6516, Nov. 2015.
[20] T. Butkiewicz, "Low-cost coastal mapping using Kinect v2 time-of-flight cameras," in *Proc. Oceans-St. John's*, Sep. 2014, pp. 1–9.
[21] E. Lachat, H. Macher, T. Landes, and P. Grussenmeyer, "Assessment and calibration of a RGB-D camera (Kinect v2 Sensor) towards a potential use for close-range 3D modeling," *Remote Sens.*, vol. 7, no. 10, pp. 13070–13097, 2015.
[22] 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.
[23] T. Wiedemeyer. *Tools for Using the Kinect One (Kinect v2) in ROS*. Accssed: Dec. 23, 2017. [Online]. Available: https://github.com/code-iai/iai_kinect2
[24] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014.
[25] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.
[26] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, "Robust L2E estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.
[27] K. Sun, L. Liu, and W. Tao, "Progressive match expansion via coherent subspace constraint," *Inf. Sci.*, vols. 367–368, pp. 848–861, Nov. 2016.
[28] O. Wasenmüller and D. Stricker, "Comparison of Kinect v1 and v2 depth images in terms of accuracy and precision," in *Proc. Asian Conf. Comput. Vis.*, 2016, pp. 34–45.

[29] F. J. Lawin, "Depth data processing and 3D reconstruction using the Kinect v2," M.S. thesis, Dept. Elect. Eng., Fac. Sci. Eng., Linköping Univ., Linköping, Sweden, 2015.
[30] L. Valgma, "3D reconstruction using Kinect v2 camera," M.S. thesis, Comput. Eng. Curriculum, Fac. Sci. Technol., Inst. Technol., Univ. Tartu, Tartu, Estonia, 2016.
[31] P. Fankhauser, M. Bloesch, D. Rodriguez, R. Kaestner, M. Hutter, and R. Siegwart, "Kinect v2 for mobile robot navigation: Evaluation and modeling," in *Proc. Int. Conf. Adv. Robot. (ICAR)*, Jul. 2015, pp. 388–394.
[32] H. Gonzalez-Jorge *et al.*, "Metrological comparison between Kinect I and Kinect II sensors," *Measurement*, vol. 70, pp. 21–26, Jun. 2015.
[33] M. Lindner and A. Kolb, "Calibration of the intensity-related distance error of the PMD ToF-camera," *Proc. SPIE*, vol. 6764, p. 67640W, Sep. 2007.
[34] M. Lindner, I. Schiller, A. Kolb, and R. Koch, "Time-of-flight sensor calibration for accurate range sensing," *Comput. Vis. Image Understand.*, vol. 114, pp. 1318–1328, Dec. 2010.
[35] P. Rodríguez-Gonzálvez, D. González-Aguilera, H. González-Jorge, and D. Hernández-López, "Low-cost reflectance-based method for the radiometric calibration of Kinect 2," *IEEE Sensors J.*, vol. 16, no. 7, pp. 1975–1985, Apr. 2016.
[36] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin, "Shading-based shape refinement of RGB-D images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1415–1422.
[37] Y. Han, J.-Y. Lee, and I. S. Kweon, "High quality shape from a single RGB-D image under uncalibrated natural illumination," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1617–1624.
[38] K. Kim, A. Torii, and M. Okutomi, "Joint estimation of depth, reflectance and illumination for depth refinement," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2015, pp. 1–9.
[39] H. Kwon, Y.-W. Tai, and S. Lin, "Data-driven depth map refinement via multi-scale sparse representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 159–167.
[40] S. Zennaro *et al.*, "Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jun./Jul. 2015, pp. 1–6.
[41] O. Wasenmüller and D. Stricker, "Comparison of Kinect v1 and v2 depth images in terms of accuracy and precision," in *Proc. Asian Conf. Comput. Vis. Workshop (ACCV Workshop)*. Cham, Switzerland: Springer, 2016, pp. 34–45.
[42] A. Landi, P. Piaggi, M. Laurino, and D. Menicucci, "Artificial neural networks for nonlinear regression and classification," in *Proc. 10th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, Nov./Dec. 2010, pp. 115–120.
[43] *Calibrating the Kinect Depth Camera to the Built-in RGB Camera*. Accessed: Feb. 13, 2018. [Online]. Available: http://wiki.ros.org/openni_launch/Tutorials/ExtrinsicCalibration
[44] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[45] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3-D mapping with an RGB-D camera," *IEEE Trans. Robot.*, vol. 30, no. 1, pp. 177–187, Feb. 2014.
[46] J. Sturm. *RGB-D SLAM Dataset and Benchmark*. Accessed: Feb. 27, 2018. [Online]. Available: https://vision.in.tum.de/data/datasets/rgbd-dataset
[47] G. Georgakis, M. A. Reza, A. Mousavian, P.-H. Le, and J. Košecká, "Multiview RGB-D dataset for object instance detection," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 426–434.

JICHAO JIAO was born in Jining, China, in 1983. He received the B.S. degree in electronic information engineering from Yanshan University, China, in 2007, and the Ph.D. degree in signal and information processing from the Beijing Institute of Technology in 2013.

Since 2013, he has been with the Beijing University of Posts and Telecommunications. Since 2017, he has been an Associate Professor. His research activities include computer science, vision-based navigation, and indoor positioning.

LIBIN YUAN received the B.Eng. degree in communication engineering from the Hebei University of Science and Technology, China, in 2013. From 2013 to 2015, he was an RF Assistant Engineer with Beijing Hong Yang Jie Xun Technology Co., Ltd, Beijing.

He is currently pursuing the master's degree with the Laboratory of Intelligent Communication, Navigation and Micro/Nano-Systems, Beijing University of Posts and Telecommunications. His research interests focus on computer vision and simultaneous localization and mapping.

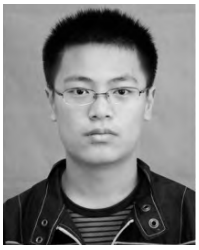
ZHONGLIANG DENG received the M.Sc. degree in manufacturing engineering from Beihang University and the Ph.D. degree in mechanical manufacture from Tsinghua University, China. He is currently a Professor and a Doctoral Supervisor with the School of Electronic Engineering, Beijing University of Posts and Telecommunications, where he is also the Director of Research with the Laboratory of Intelligent Communication, Navigation and Micro/Nano-Systems. He is also the Executive Vice President of BUPT. His research interests include indoor and outdoor seamless positioning, GNSS, satellite communications, MEMS, and multimedia.

CHENG ZHANG received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2016, where he is currently pursuing the master's degree with the Laboratory of Intelligent Communication, Navigation and Micro/Nano-systems. His research mainly focuses on semantic segmentation in 3-D dataset of indoor environment and transfer learning based on small sample dataset.

WEIHUA TANG was born in Yongzhou, China, in 1984. She received the B.S. and M.S. degrees in computer science from the Beijing Institute of Technology in 2009.

Since 2009, she has been an Assistant Researcher with China State Construction Engineering Corporation Ltd.

Her research interests include computer vision and image processing.



QI WU received the B.Eng. degree in electronic information engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016. He is currently with the Laboratory of Intelligent Communication, Navigation and Micro/Nano-systems, Beijing University of Posts and Telecommunications. His research topic is simultaneous localization and mapping, and his interest is loop detection for visited place.

JIAN JIAO received the B.Eng. degree in communication engineering from Liaoning Technical University, China, in 2016. From 2016 to 2017, he was an RD Assistant Engineer with Ling Da Technology Co., Ltd, Beijing.

He is currently pursuing the master's degree with the Laboratory of Intelligent Communication, Navigation and Micro/Nano-Systems, Beijing University of Posts and Telecommunications. His research interests focus on computer vision and simultaneous localization and mapping.