Research Article
A Structured Light RGB-D Camera System for Accurate Depth Measurement

Van Luan Tran [iD][1] **and Huei-Yung Lin**[1,2]

[1]*Department of Electrical Engineering, National Chung Cheng University, 168 University Rd., Chiayi 621, Taiwan*
[2]*Advanced Institute of Manufacturing with High-Tech Innovation, National Chung Cheng University, 168 University Rd., Chiayi 621, Taiwan*

Correspondence should be addressed to Van Luan Tran; tranvanluan07118@gmail.com

The ability to reliably measure the depth of the object surface is very important in a range of high-value industries. With the development of 3D vision techniques, RGB-D cameras have been widely used to perform the 6D pose estimation of target objects for a robotic manipulator. Many applications require accurate shape measurements of the objects for 3D template matching. In this work, we develop an RGB-D camera based on the structured light technique with gray-code coding. The intrinsic and extrinsic parameters of the camera system are determined by a calibration process. 3D reconstruction of the object surface is based on the ray triangulation principle. We construct an RGB-D sensing system with an industrial camera and a digital light projector. In the experiments, real-world objects are used to test the feasibility of the proposed technique. The evaluation carried out using planar objects has demonstrated the accuracy of our RGB-D depth measurement system.

1. Introduction

In recent years, 3D imaging has received a great value in industrial and consumer applications. Machine vision systems developed with 3D imaging allow faster and more accurate measurement of components at manufacturing whereabouts. Nowadays, RGB-D cameras, such as Microsoft Kinect and Asus Xtion, are very popular due to the ability to provide the depth information directly. However, they have the limitation on accuracy and thus are not suitable for the applications that require accurate shape measurements [1–3]. As a result, the development of real-time RGB-D cameras still receives much attention from researchers and practitioners. The objective is to provide highly accurate RGB-D sensing techniques with more effective implementation approaches in terms of the density of acquired point clouds, time consumption, working environment, noise level, etc.

3D reconstruction based on the structured light technique has been investigated in the past few decades due to its popularity in the manufacturing applications. Structured light systems are suitable solutions for structured light scanning, 3D reconstruction, and 3D sensing with accurate shape measurements [4, 5]. Structured light refers to the process of projecting predesigned known patterns on the scene and capturing the images to calculate the depth for 3D surface reconstruction. It is an important contribution to the development of 3D measurement systems. The patterns projected on the scene can be generated by a projector or other devices [6], and the relationship between the light source and the camera is a crucial factor. The accuracy of 3D reconstruction depends on the correctness of the calibration, which provides the relative pose between the camera and the light source projector.

In recent literature, several works presented the structured light systems for 3D reconstruction and proposed different approaches to deal with the related problems [7–10]. Scharstein *et al.* [11] proposed a method for acquiring high-complexity stereo image pairs with pixel-accurate correspondence information using structured light. Some previous works such as [12–15] described various methods to perform 3D reconstruction and obtained some satisfactory results. However, those techniques require to use precalibrated
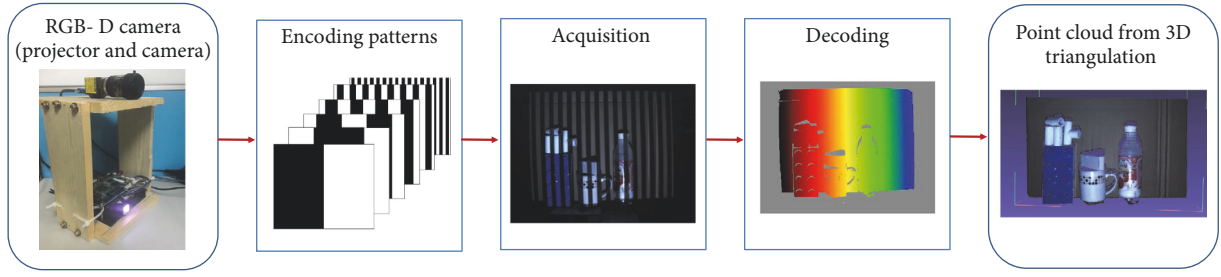
FIGURE 1: The overview of our RGB-D camera system with the structured light technique. The encoding pattern is a gray-code pattern. The acquisition is the images captured by the camera for each pattern in the sequence. The decoding is a coded map.

cameras to find the 3D world coordinates of the projected pattern. Thus, they highly depend on the accuracy of camera calibration and may transfer the error to the projector calibration. In [16], Huang and Tang described a method to perform fast 3D reconstruction using one-shot spatial structured light. Although the method can provide relatively accurate results, the evaluation and analysis were not carried out comprehensively. Some restrictions are also shown in their experiments when performing the tests on complex object surfaces. Cui and Dai [17] proposed a simple and efficient 3D reconstruction algorithm using structured light from 3D computer vision. However, their approach has some limitations on measuring inclined objects, and the 3D information cannot be recovered for the shadow areas.

In this work, we develop an RGB-D camera system based on the structured light technique. A system flowchart is shown in Figure 1. The encoding method is based on the gray-code coding [5], and the 3D reconstruction is achieved by the ray triangulation principle with the estimation of intersection points. The accuracy and density of the obtained point clouds are both high, and therefore it is suitable for the applications such as accurate shape measurements, 3D object recognition, and pose estimation for robotic manipulation.

This article is organized as follows: Section 2 presents a general overview of the structured light system and an accurate calibration method to derive the parameters of the camera-projector system. Section 3 contributes a method to create encoded patterns and decode the captured images from the camera and presents a ray triangulation principle for 3D computation by point intersection. Section 4 provides some experimental results, including the experimental setup, results with several different objects, and the evaluation of the accuracy of the object reconstruction. Finally, Section 5 gives the conclusion.

2. Background

2.1. Structured Light Technique.
Currently, the development of structured light systems is in high demand. The structured light technique is based on the principle described in Figure 2. In general, the process of a structured light system can be divided into three basic steps:

(i) **Encoding.** The encoding of the information into a sequence of patterns is performed in the temporal domain. A sequence of structured light patterns
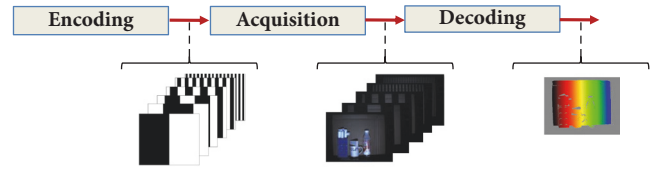

FIGURE 2: The overview of three basic steps in the process of a structured light system [1].

depends on the number of required patterns, parameters of the system, and the resolution of the projector and the camera.

(ii) **Acquisition.** The sequence of patterns is projected on the scene by a data projector, and a camera is used to continuously capture the images.

(iii) **Decoding.** The captured pattern-coded images are processed with the recognition of projected patterns to find the corresponding points associated with the projector and the camera.

In the implementation, there might be some additional steps depending on the solution of the system designer. It often follows a procedure to create range images, point clouds, or mesh models and possibly integrates several decoded coordinate maps, calibration, and triangulation principle. The calibration is to determine the intrinsic and extrinsic parameters of the camera and the projector, and the reconstruction is usually based on the ray triangulation principle by computing the intersection point.

2.2. Calibration.
Calibration is an important issue which greatly affects the accuracy of the results [18]. In the proposed technique, firstly, we find the parameters of the system using the calibration method by Moreno and Taubin [6]. It is a simple and accurate method to calibrate the projector and camera systems. In this method, the projected corner locations are estimated with subpixel precision using local homographies to each corner in the images as illustrated in Figure 3. It includes three main steps as follows:

(i) The camera calibration step to determine the intrinsic parameters of the camera. Camera calibration includes collecting a sequence of images of a planar
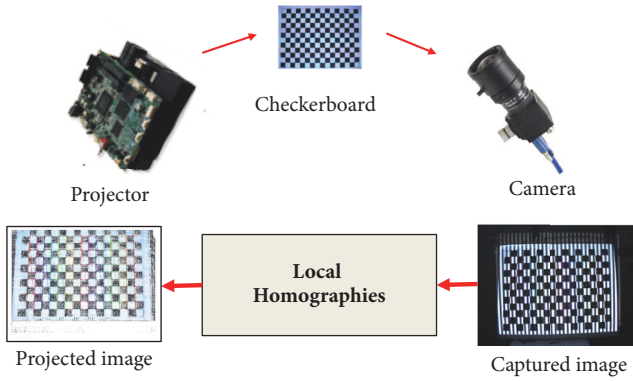
FIGURE 3: An illustration of the structured light system calibration. The captured image was selected in a set of the calibrated images, which are the projector project the patterns on the checkerboard and captured by the camera. The projected image is estimated with subpixel precision using local homographies to each corner in the captured images [1].

checkerboard pattern. The intrinsic camera calibration is derived by estimating the parameters using the perspective camera model [19]. We find the coordinates in the camera image plane for all of the checkerboard corners captured with different pattern orientations. We use OpenCV's findChessboardCorners() function [20] to automatically find the checkerboard corner locations. They are then refined to approach the subpixel accuracy. Finally, OpenCV's calibrateCamera() function is used to derive the calibrated camera parameters.

(ii) The projector calibration step is to determine the intrinsic parameters of the projector. The mathematical model of our projector can be described the same as the camera. But the projector cannot capture the images from its viewpoint to find checkerboard corners. In this situation, we know a relation between the projector and image pixels extracted from the structured light sequences. Thus, we can estimate the checkerboard corner locations in the projector pixel coordinates based on a local homography [6] as an illustration in Figure 3.

(iii) The stereo system calibration step is to derive the extrinsic parameters of the system, which consist of the rotation matrix and the translation vector. We use OpenCV's stereoCalibrate() function with the previously found checkerboard corner coordinates and their projections. The stereo parameters are a rotation matrix R and a translation vector T relating the projector-camera pair.

3. RGB-D Sensing Based Using Structured Light

3.1. Encoding and Decoding Patterns.
The gray-code pattern [4] is a sequence of images with black and white stripes created for encoding the scene from the camera viewpoint.
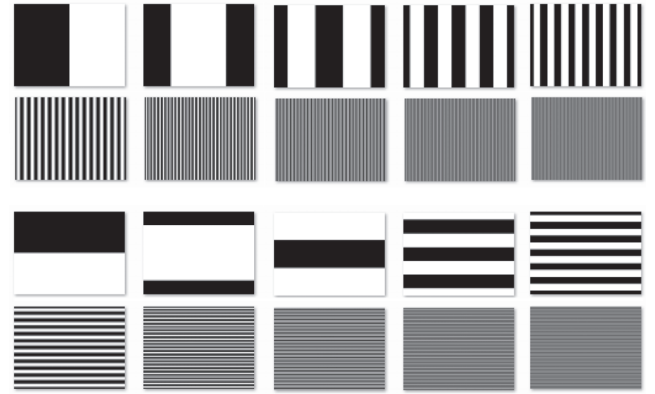

FIGURE 4: The gray-code patterns for the RGB-D camera used by the structure light technique [1].
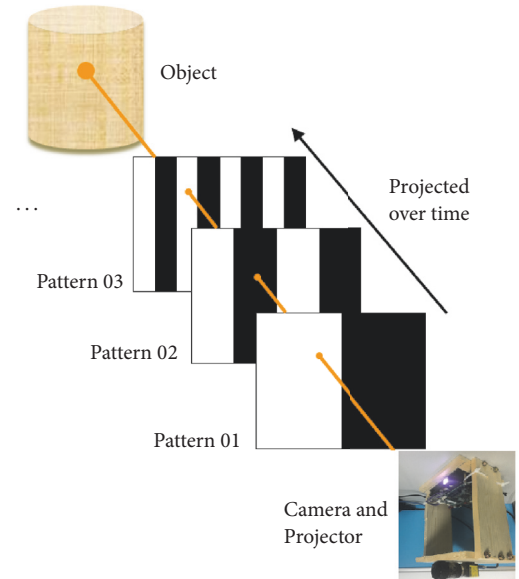

FIGURE 5: The acquisition of the projected patterns on an object. The gray-code pattern is projected by the projector and the scene is captured by the camera [1].

The pattern sequence has two types, one is the horizontal stripe and the other is the vertical stripe, as illustrated in Figure 4. All patterns are projected to a scene or an object as shown in Figure 5. The horizontal patterns consist of 10 pattern images which represent 10-bit values for each pixel. The first pattern is half black and half white, which represents the most significant bit, and the rest patterns are the images that switch between black and white in the columns. After combining all of the 10 horizontal patterns into one image, each column has the same 10 bits with the columns in the same image.

Structured light encoding depends on the resolution of the projector. The information is encoded into a sequence of patterns performed in the temporal domain. Commonly used approaches include gray-code coding and binary-code coding. Gray codes can be calculated by first computing the
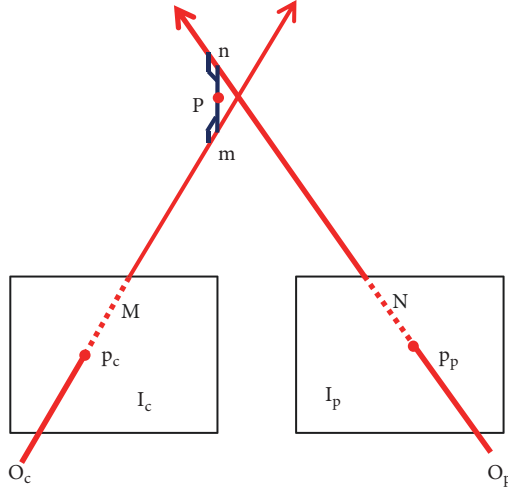
Figure 6: The intersection point of two rays from the pixels in the camera and projector coordinates [1].

binary representation of a number and then converting it using the following process: copy the most significant bit as it is, and replace the remaining bits (taking one bit at a time) with the result of an XOR operation of the current bit, with the previous bit of higher significance in the binary form.

For the binary-code coding, only two illumination levels are used and encoded as 0 and 1. The gray-code coding is an alternative to the binary representation, with only one bit change at a time between any two adjacent numbers. If there is an error reading on any changed bit, the value will never be off by more than one unit. In our system, we use a projector with the resolution of 1024×768 and decode the pattern with 10 bits ($2^{10} = 1024$ and $2^{10} -$ offset $= 768$), where the number of vertical patterns is $\lceil \log(1024)/\log(2.0) \rceil = 10$ and the number of horizontal patterns is $\lceil \log(768)/\log(2.0) \rceil = 10$.

The camera captures the images of the projected patterns, and the coding step is to decode each pixel in the captured images into their corresponding decimal number presenting the column and row. It will be used to create a coded map as shown in Figure 1, which presents the corresponding point between the projector and the camera.

3.2. 3D Reconstruction. With a robust projector-camera calibration step, we define the location and orientation of the camera and projector with respect to the world coordinate frame. In the pattern encoding and decoding step, we determine one pixel in image I_c and its corresponding pixel in image I_p. Our reconstruction is based on the ray triangulation principle by the estimation of intersection points [4]. In order to compute the direction vector, two points in a ray are needed. The first point is the camera's center of projection, which is determined based on the extrinsic parameters of the structured light system. The second point is the point corresponding to the pixel from which the ray passes through. One ray passes through O_c and p_c of the left image (I_c), and the other ray passes through O_p and p_p of the right image (I_p), as shown in Figure 6. Here, O is the origin of the system and p is a pixel in the image. The 3D point cloud is obtained from

the intersection point P, which is a midpoint of the shortest segment between the rays.

To estimate the intersection point, we consider two rays M and N in 3D space passing through points p_c and p_p with direction vectors $\overrightarrow{x}$ and $\overrightarrow{y}$, respectively. Let the two closest points on the lines be m and n, as defined in (1) and (2), where g and k are scalar values.

$$m = p_c + g\overrightarrow{x} \tag{1}$$

$$n = p_p + k\overrightarrow{y} \tag{2}$$

The mn segment connecting rays M and N is perpendicular to the rays, and therefore the dot product of the vectors is equal to 0 as follows:

$$(m - n)\overrightarrow{x} = 0 \tag{3}$$

$$(m - n)\overrightarrow{y} = 0 \tag{4}$$

With (1) and (2), (3) and (4) are represented by

$$\overrightarrow{r} \cdot \overrightarrow{x} + g\overrightarrow{x} \cdot \overrightarrow{x} - k\overrightarrow{y} \cdot \overrightarrow{x} = 0 \tag{5}$$

$$\overrightarrow{r} \cdot \overrightarrow{y} + g\overrightarrow{y} \cdot \overrightarrow{x} - k\overrightarrow{y} \cdot \overrightarrow{y} = 0 \tag{6}$$

where

$$\overrightarrow{r} = p_c - p_p \tag{7}$$

From (5) and (6), the scalar values are calculated by

$$g = \frac{\left(\overrightarrow{r} \cdot \overrightarrow{x}\right)\left(\overrightarrow{y} \cdot \overrightarrow{y}\right) - \left(\overrightarrow{y} \cdot \overrightarrow{x}\right)\left(\overrightarrow{r} \cdot \overrightarrow{y}\right)}{\left(\overrightarrow{y} \cdot \overrightarrow{x}\right)\left(\overrightarrow{y} \cdot \overrightarrow{x}\right) - \left(\overrightarrow{y} \cdot \overrightarrow{y}\right)\left(\overrightarrow{x} \cdot \overrightarrow{x}\right)} \tag{8}$$

$$k = \frac{\left(\overrightarrow{y} \cdot \overrightarrow{x}\right)\left(\overrightarrow{r} \cdot \overrightarrow{x}\right) - \left(\overrightarrow{x} \cdot \overrightarrow{x}\right)\left(\overrightarrow{r} \cdot \overrightarrow{y}\right)}{\left(\overrightarrow{y} \cdot \overrightarrow{x}\right)\left(\overrightarrow{y} \cdot \overrightarrow{x}\right) - \left(\overrightarrow{y} \cdot \overrightarrow{y}\right)\left(\overrightarrow{x} \cdot \overrightarrow{x}\right)} \tag{9}$$

The midpoint of the shortest segment are then estimated by

$$P = \frac{p_c + g\overrightarrow{x} + p_p + k\overrightarrow{y}}{2} \tag{10}$$

4. Experiments

In the structured light system, the quality of the captured images is important for obtaining a good pattern database to perform the calibration, decoding, and reconstruction. Hence, the resolution of the camera is usually higher than the resolution of the projector. Then, the projection field of view is adjusted inside the field of view of the camera. In the experiments, we use a Flea3 FL3-U3-32S2C from Point Grey Research with the image resolution of 2080×1552. The digital light projector is a DLP Light Crafter 4500 projector from Texas Instrument with the resolution of 1024×678. Their focus length, resolution, zoom, and direction were selected prior to calibration accordingly to the target of the system. All devices are connected to a host computer. After the system is calibrated, part of the system cannot be moved. We have to
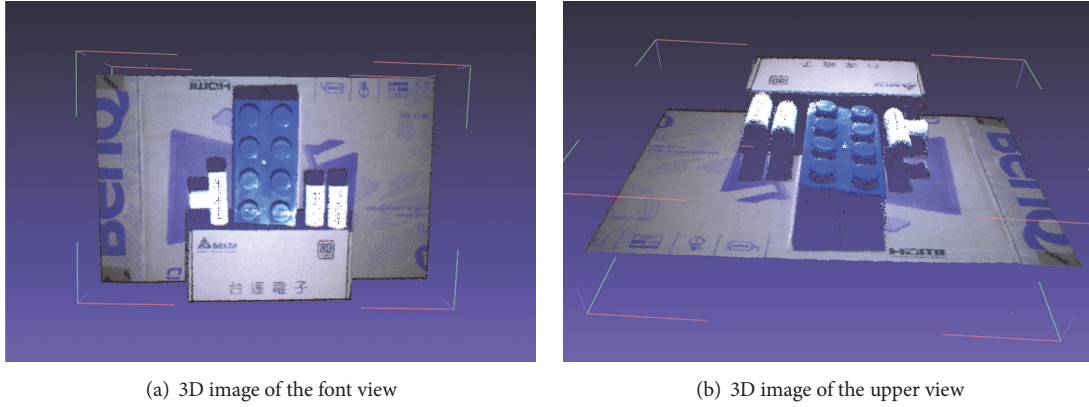
(a) 3D image of the font view

(b) 3D image of the upper view

FIGURE 7: 3D reconstruction result with some objects (a lego box, three plastic pipes, and a big box), these objects are low reflecting light.

(a) 3D image of the font view
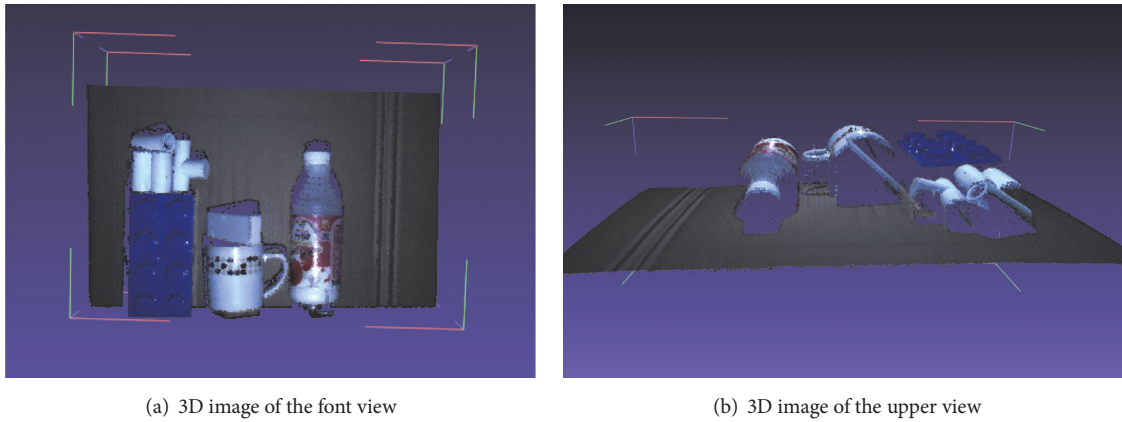
(b) 3D image of the upper view

FIGURE 8: 3D reconstruction result with multiple objects (a cup, bottle, small box, lego box, and some plastic pipes), the bottle and cup are with high reflecting light.

keep the distance and orientation between the projector and the cameras intact; otherwise it will be essential to perform a recalibration.

The settings of the camera and projector should be adapted according to the lighting in the scene. Other lighting sources projected directly in the scene should be rejected. If not, the calibrated and reconstructed results will be affected. The system was calibrated with 12 sets of acquired projected checkerboard patterns. An acquisition set includes the images captured by the camera for each pattern in the sequence. After the system is calibrated as exposited in Section 2.2, the calibration result is stored in a .yml file.

For reconstruction, our system includes three main steps. Firstly, our system loads the calibration parameters and projects one acquisition set of patterns on objects. Secondly, decoding the captured pattern-coded images provides a coded map to store the corresponding points between the projector and the camera. Finally, with calibrated parameters and the coded map, we apply the ray triangulation principle to get the 3D point that will be rendered simultaneously with one color image to create an XYZRGB point cloud. In Figures 7 and 8, we present the 3D reconstruction results of several objects. The results successfully measure objects with reflecting light for some of the projected colors. After

performing the reconstruction, the 3D information of the reconstructed objects is saved in a .txt file.

The evaluation of the proposed technique is performed by measuring the dimension of a reconstructed checkerboard pattern and its corner points. The checkerboard has the dimension of $399 \times 285 \, mm^2$ and each small square has the size of $28.5 \times 28.5 \, mm^2$ as demonstrated in Figure 9. After performing the 3D reconstruction for this checkerboard, we can use 3ds Max Design or Meshlab software to examine, as shown in Figure 10. A distance measurement tool is used to measure the dimension of the checkerboard. The accuracy is presented in Table 1, with the errors of the corner points. This table reports that our system can measure the objects with high accuracy. Compared with the algorithms proposed by Moreno *et al.* [6] with Max. error of 0.8546(%). They use a method to estimate the image coordinates of 3D points in the projector image plane and perform the calibration on both projector and camera. With Max. error of 0.1240(%), our proposed method provides the better 3D reconstruction results.

5. Conclusion

In this work, we have developed an RGB-D camera system based on the structured light technique. It contains a camera

TABLE 1: Measuring the accuracy of a reconstructed checkerboard (CB) as shown in Figure 10.

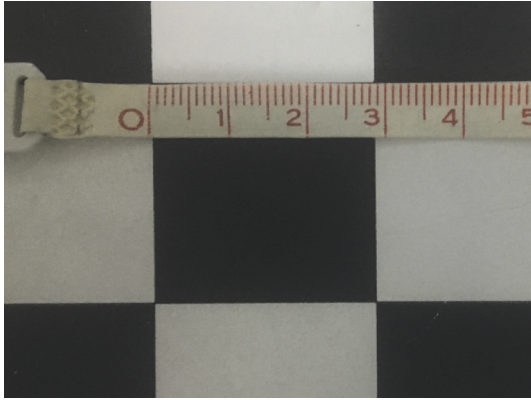| Name	Real CB (mm)	Reconstructed CB (mm)	Error (%)
Width of top checkerboard (AB)	399.000	399.382	0.0957
Width of bottom checkerboard (CD)	399.000	399.495	0.1240
Height of left checkerboard (AC)	285.000	284.746	0.0891
Height of right checkerboard (BD)	285.000	285.298	0.1045
Max. error			0.1240

FIGURE 9: The real dimension of each small square has the size of $28.5 \times 28.5 \, mm^2$ in the checkerboard reconstructed as shown in Figure 10.
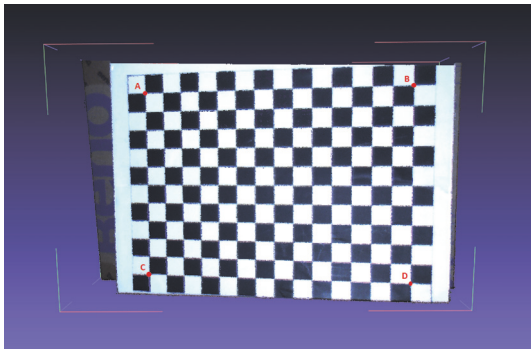
FIGURE 10: The 3D reconstruction of a checkerboard shown with Meshlab software. We used a measuring tool of the Meshlab to measure it.

and a projector to perform accurate shape measurements with high-density point cloud outputs. 3D reconstruction with multiple objects and performance evaluation of the system are carried out in the real-world environment. Our method has high accuracy as presented in the experimental results. In the experiment, we tested the system with different objects to check the surface of reconstruction and accuracy evaluation. The results have demonstrated that the proposed technique is feasible for dense 3D measurement applications.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request (https://github.com/luantran07/data-for-a-structured-light-rgb-d-camera-system).

Disclosure

This publication is an extended version of 2017 International Conference on System Science and Engineering (ICSSE) [1].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] V.-L. Tran and H.-Y. Lin, "Accurate RGB-D camera based on structured light techniques," in *Proceedings of the 2017 International Conference on System Science and Engineering, ICSSE 2017*, pp. 235–238, Viet Nam, July 2017.

[2] O. Wasenmüller and D. Stricker, "Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision," in *Computer Vision – ACCV 2016 Workshops*, vol. 10117 of *Lecture Notes in Computer Science*, pp. 34–45, Springer International Publishing, Cham, 2017.

[3] J. Smisek, M. Jancosek, and T. Pajdla, "3D with Kinect," in *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV '11)*, pp. 1154–1160, IEEE, Barcelona, Spain, November 2011.

[4] Q. Gu, K. Herakleous, and C. Poullis, "3DUNDERWORLD-SLS: An open-source structured-light scanning system for rapid geometry acquisition," *Computer Vision and Pattern Recognition*, 2016.

[5] D. Moreno, W. Y. Hwang, and G. Taubin, "Rapid Hand Shape Reconstruction with Chebyshev Phase Shifting," in *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*, pp. 157–165, Stanford, CA, USA, October 2016.

[6] D. Moreno and G. Taubin, "Simple, accurate, and robust projector-camera calibration," in *Proceedings of the 2nd Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2012*, pp. 464–471, Switzerland, October 2012.

[7] C.-Y. Chen, P.-S. Huang, S.-W. Huang, J.-H. Zhang, and B. R. Chang, "Structured light 3D face scanning system," in *Proceedings of the 2nd IEEE International Conference on Consumer Electronics - Taiwan, ICCE-TW 2015*, pp. 344-345, Taiwan, June 2015.

[8] M. Massot-Campos, G. Oliver-Codina, H. Kemal, Y. Petillot, and F. Bonin-Font, "Structured light and stereo vision for underwater 3D reconstruction," in *Proceedings of the MTS/IEEE OCEANS 2015 - Genova*, Italy, May 2015.

[9] M. Piccirilli, G. Doretto, A. Ross, and D. Adjeroh, "A Mobile Structured Light System for 3D Face Acquisition," *IEEE Sensors Journal*, vol. 16, no. 7, pp. 1854-1855, 2016.

[10] D. Lanman, *Grey code patterns for 3D reconstruction by structure light technique*, AIT Computer Vision Wiki, Brown University, 2014.

[11] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. I/202, USA, June 2003.

[12] J. Wang, C. Zhang, W. Zhu, Z. Zhang, Z. Xiong, and P. A. Chou, "3D scene reconstruction by multiple structured-light based commodity depth cameras," in *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2012*, pp. 5429–5432, Japan, March 2012.

[13] C. Bourgeois-Republique, A. Dipanda, and A. Koch, "A Structured Light System Encoding for an Uncalibrated 3D Reconstruction Based on Evolutionary Algorithms," in *Proceedings of the 2013 International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 124–129, Kyoto, Japan, December 2013.

[14] S. Huang, L. Xie, Z. Wang, Z. Zhang, F. Gao, and X. Jiang, "Accurate projector calibration method by using an optical coaxial camera," *Applied Optics*, vol. 54, no. 4, pp. 789–795, 2015.

[15] K. Herakleous and C. Poullis, "Stripe boundary codes for real-time structured-light range scanning of moving objects," in *Proceedings of the Eighth IEEE International Conference on Computer Vision*, vol. 2, pp. 359–366, Vancouver, Canada.

[16] B. Huang and Y. Tang, "Fast 3D reconstruction using one-shot spatial structured light," in *Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2014*, pp. 531–536, USA, October 2014.

[17] H. Cui, N. Dai, W. Liao, and X. Cheng, "An accurate reconstruction model using structured light of 3-D computer vision," in *Proceedings of the 7th World Congress on Intelligent Control and Automation, WCICA'08*, pp. 5095–5099, China, June 2008.

[18] M. Fu, Y. Leng, and H. Zhang, "A calibration method for structured light systems based on a virtual camera," in *Proceedings of the 8th International Congress on Image and Signal Processing, CISP 2015*, pp. 57–63, China, October 2015.

[19] S. Zhang and P. S. Huang, "Novel method for structured light system calibration," *Optical Engineering*, vol. 45, no. 8, Article ID 083 601083 6018, 2006.

[20] G. Bradski, "The OpenCV Library," *Dr Dobbs Journal of Software Tools*, 2000.