Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Miniature 3D Depth Camera for Real-time Reconstruction

Zilong Zhao[1,2], Feifei Gu[2,3,*], Pengju Xie[2], Huazhao Cao[2], and Zhan Song[2,3]

[1]*College of Computer Science and Electronic Engineering*
Hunan University
Changsha, Hunan Province, China
zerahhah@hnu.edu.cn

[2]*Shenzhen Key Laboratory of Minimally Invasive Surgical Robotics and System, Shenzhen Institutes of Advanced Technology*
Chinese Academy of Sciences
Shenzhen, Guangdong Province, China.
{pj.xie, hz.cao1}@siat.ac.cn

[3]*The Chinese University of Hong Kong*
Hong Kong, China
{ff.gu, zhan.song}@siat.ac.cn

***Abstract* - This paper developed a miniature 3D vision system for dynamic depth recovery. Our system is constituted by a DOE projector and two cameras. A single pattern with pseudo-random speckles was projected onto the surface of the target object and the coupling pattern was imaged by the stereo cameras. An optimal camera calibration algorithm is adopted to guarantee the measurement accuracy. The speckle pattern projected through a diffractive optical element (DOE) of size 8mm×8mm. The baseline between the stereo cameras is as short as 50mm. The system has the advantages of simple structure, small size, and low cost. It can be easily integrated into mobile devices such as tablet computers and mobile phones. In comparison with several other mature products, the proposed 3D depth camera can achieve 3D reconstruction with higher accuracy.**

Index Terms* - *Speckle, structured light, optimized SGM, diffraction optical element, 3D dynamic reconstruction.

I. INTRODUCTION

With the rapid improvement of mobile devices' computing power and the development of integration technology, 3D reconstruction technology has been gradually applied to the field of robotics, human-computer interaction, face recognition, face payment, and AR/VR, etc[1-3]. The 3D vision technique carrying on intelligent terminals is gradually developed towards integration, miniaturization, and high accuracy, etc. Currently, 3D dynamic reconstruction techniques mainly include the time of flight (TOF) [4], binocular stereo vision technique [5] and structured light technique (SLT) [6,7]. The TOF technique has the advantages of high integration and fast response time, however, it has the disadvantages of low depth resolution and more noise when working in close range. The binocular stereo vision system is referred to be a passive vision system, which recovers 3D point clouds via the triangulation structure constituted by two cameras. It requires first to detect and match feature points from the captured stereo images, which therefore can only conduct sparse reconstruction. The structured light technique (SLT) also works on the principle of triangulation, except that an active light projection source is added. Compared with the binocular stereo vision technique, it addresses the stereo matching problem straightforwardly and enhances the performance in reconstructing textureless regions.

II. RELATED WORKS

The SLT has a trade-off between acquisition time and spatial resolution. Multi-shot temporally-coded methods can achieve 3D reconstruction with high density and accuracy, yet it is not applicable to dynamic reconstruction. On the contrary, single-shot spatially-coded methods are applicable to dynamic reconstruction but can only achieve sparse reconstruction. Except for spatially-coded methods of a general case, optical projection using speckles is another structure light method that has been applied to 3D dense and dynamic reconstruction. The statistically random coding method has a distinct advantage in achieving high-density 3D reconstruction. This method generates locally unique features per pixel on the structured pattern, and corresponding counterparts for every camera pixel can be uniquely located by analyzing the statistical pattern. Due to its specific advantages in single-shot-imaging and high-density, the statistically random coding method has been successfully employed on consumer-level 3D depth sensors (e.g., Microsoft Kinect V1, Intel RealSense D400 series, Apple iPhone X, etc). Du et al.[8] conducted deep researches on the design of infrared speckles and the density distribution of speckles. Silberman et al.[9] accomplished semantic segmentation of indoor scenes using the Kinect v1, which works on the principle of structured light and infrared sensors, as shown in Fig.1(a). Wu et al.[10] designed a 3D motion-sensing interaction system based on speckles, as shown in Fig.1(b). A fast method for recovering depth information using the zero-mean normalized cross-correlation operator (ZNCC) was proposed. Tang et al. [11] introduced the basic principle of binocular stereo vision and speckle-based 3D reconstruction, as displayed in Fig.1(c). The whole-pixel corresponding point search method and the sub-pixel interpolation algorithm based on Newton-Lafferson iteration are emphatically analyzed, and the sub-pixel 3D reconstruction of speckle is realized. In [12], the commercial DOEs (Diffractive

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

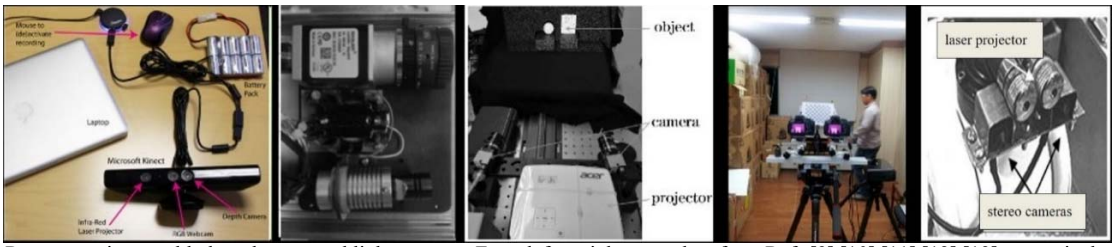


Fig. 1 Representative speckle-based structured light systems. From left to right was taken from Refs [9] [10] [11] [12] [13] respectively.

Optical Elements) are used to project speckles, and the SLR camera is used to acquire images, as shown in Fig.1 (d). Dense depth images can be obtained. The speckle-based 3D reconstruction system in [13] was built to provide depth information to robotics in robot obstacle avoidance and autonomous navigation. The related system was displayed in Fig.1 (e).

However, all the vision systems mentioned above are relatively large, and the system cost is relatively high. Therefore they cannot be applied in consumer-level mobile intelligent terminals. Besides, recent works [15-18] find that stereo methods have implicit assumptions that lead to a bias towards reconstructing front-parallel surfaces. Our method alleviates this kind of bias. In this paper, a miniature three-dimensional system with a short baseline is designed, which combines active structured light to compensate for the inherent shortcomings of passive binocular.To improve the accuracy of the 3D reconstruction, a high-precision camera calibration method was utilized and an improved stereo matching method was proposed. The paper is organized as follows. Section III describes the camera calibration, stereo rectification, and feature matching algorithms adopted in the proposed 3D depth sensor. The experimental results are given in Section IV. Conclusions are presented in Section V.

III. Proposed Algorithm

As mentioned above, most structured light systems rely on expensive and high-powered projection devices. To reduce system cost, a DOE projector was utilized in our structured light system. The schematic diagram of our proposed system is displayed by Fig. 2, including a binocular system and a DOE projector. The general procedures to conduct 3D reconstruction include stereo calibration of left and right cameras, image rectification, stereo matching, and 3D reconstruction. Among them, the measurement accuracy of the speckle-based structured light system depends mainly on two crucial steps: system calibration and stereo matching, which is a big concern of this paper. The detailed procedures of our algorithm were described hereinbelow.

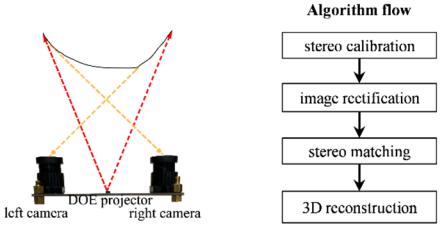


Fig. 2 Schematic diagram of the proposed miniature measurement device.

A. System calibration

High-accuracy measurement in the speckle-based structured light is heavily dependent on the accurate calibration of the stereo cameras. The calibration procedure generally involves the recovery of intrinsic parameters (i.e., the principle point, the focal length and the lens distortions of both cameras) and extrinsic parameters (i.e., the rigid transformation between the two cameras). Zhang's method [19] has been one of the most widely-used methods to calibrate a stereo rig by viewing a planar pattern from different orientations. This method is based on the forward imaging process (FIP) and optimizes stereo parameters by minimizing the re-projection pixel errors on both image planes. However, 3D reconstruction is based on the backward projection process (BPP). Minimum re-projection pixel errors do not strictly equal to minimum reconstruction errors. In our previous work [20, 21], we developed an optimal calibration method based on BPP, which optimized stereo parameters by minimizing the total 3D reconstruction errors of both cameras based on BPP. In this way, both calibration and reconstruction accuracy can be improved. Detailed illustrations can refer to [20, 21].

B. Image rectification

To accelerate the process of feature matching between the stereo images, epipolar rectification should be conducted. In this case, all candidate matching points in the right images keep in the same row as the left image. There are also a series of effective algorithms for stereo rectification, which in this paper the one from [14] was adopted.

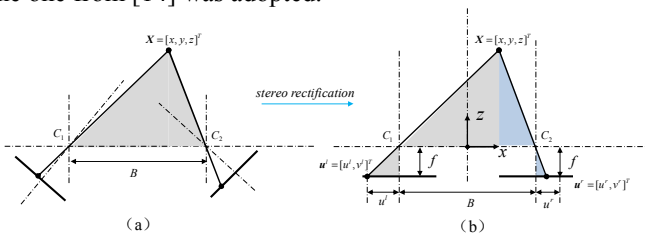


Fig. 3 Epipolar rectification of stereo images: triangular structure (a) before rectification (b) after rectification

After rectification, the stereo vision system can be thought of as a single camera translated along the X-axis of its reference system. The optical centers $C_1$and C_2 keep the same as the old cameras, whereas the new rotation differs from the old ones. New intrinsic parameters are the same for both cameras. Assume the pre-rectification projection matrics of each camera are $P_{o1}$and P_{o2}, and their rectified counterparts are $P_{n1}$and P_{n2}re-

spectively. The new projection matrics can be achieved by rotating the camera around $C_1$and C_2. The process to achieve the new projection matrics equals the process of stereo rectification.

Assume the new rotation matrix is $R_n = [r_1, r_2, r_3]$, then the way to determine it is as follows:

1) the new X-axis is parallel to the baseline, thus $r_1 = (C_1 - C_2)/(\|C_1 - C_2\|)$;

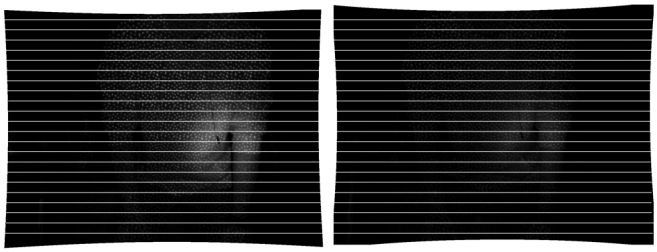
2) the new Y-axis is orthogonal to the X-axis and the old Z-axis (denoted by $k = [0,0,1]^T$): $r_2 = k \wedge r_1$;

3) The new Z-axis is orthogonal to both X and Y axes: $r_3 = r_1 \wedge r_2$.

Then the new projection matrics $P_{n1}$and P_{n2}can be determined by,

$$\begin{cases} P_{n1} = K_n[R_n| - R_nC_1] \\ P_{n2} = K_n[R_n| - R_nC_2] \end{cases} \quad (1)$$

After obtaining the new projection matrics, stereo images can be rectified by making use of this new mapping relationship. For non-integer pixel positions, bilinear interpolation operation is conducted to obtain smooth images. One example of stereo rectification in our system was displayed in Fig.4 as below.



(a) rectified left image (b) rectified right image

Fig. 4 One example of image rectification in our system

C. *Stereo matching based on an optimal SGM*

After the operation of image rectification, the stereo matching task can be conducted along the same rows between the stereo images. Compared with other local and global matching methods, the SGM algorithm achieves a good balance between high-quality depth maps and running efficiency [22]. The problem of stereo matching is to find optimal disparity map D that can minimize the energy $E(D)$. In the process of cost calculation, additional constraints are added in $E(D)$.to support smoothness, as shown in Eq. (2).

$$E(D) = \sum_p (Cost\,(p,d) + \sum_{q \in N_p} V(d_p, d_q)) \quad (2)$$

In Eq. (2), the first term is the sum of all pixel matching costs for the disparities of D. The second term adds constant penalties for all pixels q in the neighborhood N_p of p. This extra term penalizes changes in neighboring disparities. One characteristic of SGM is aggregating matching costs in 1D from all directions equally. Along each direction r, the cost of the pixel p at disparity d is defined recursively as,

$$\begin{aligned} L_r(p,d) = Cost(p,d) + \min(&L_r(p-r,d) + V(d_p, d_{p-r}), \\ &L_r(p-r,d-1) + V(d_p, d_{p-r}), \\ &L_r(p-r,d+1) + V(d_p, d_{p-r}), \\ &\min_i L_r(p-r,i) + V(d_p, d_{p-r})) \end{aligned} \quad (3)$$

The first term is the calculated cost at the disparity d. The second term is the lowest cost of the previous pixel of p-r alone the path. The symbol $V(d_p, d_{p-r})$ is a penalty term that penalizes changes in neighboring disparities. Its definition can refer to Eq. (4). The second term in Eq. (2) can be specifically interpreted as: If the disparity on pixel p-r keeps the same with the disparity on p, no penalty is added. If the disparity on pixel q changes only 1 pixel, the penalty factor P_1 is utilized. For larger disparity changes (that is, >1pixel), the penalty factor P_2 is added. Generally, pixels with larger disparity changes have little effect on the final results, so P_2 is generally set to be a large value.

$$V(d_p, d_{p-r}) = \begin{cases} 0, if\ d_p - d_{p-r} = 0 \\ P_1, if\ |d_p - d_{p-r}| = 1 \\ P_2, if\ |d_p - d_{p-r}| > 1 \end{cases} \quad (4)$$

Recent works [15-18] find that SGM has the same assumption with local matching methods that all pixels in the matching window have constant disparities. This implicit assumption does not hold for slanted surfaces and leads to a bias towards reconstructing frontoparallel surfaces. We experimentally find that the penalty term $V(d_p, d_{p-r})$ reinforced the above-mentioned assumption to a great extent. Fig.5 illustrates this problem. Take a slanted surface and a curved surface as examples. Ideal disparity estimation routes are displayed in Fig.5 (b) as a comparison. However, in traditional SGM algorithm, the case of keeping constant disparities was encouraged by giving a smaller penalty ($V(d_p, d_{p-r}) = 0$) to the state of $d_p = d_{p-r}$. This inevitably leads to inaccurate disparity results as P and Q in Fig.5 (a). In fact, the penalty mechanism in Eq. (3) works only well for frontoparallel surfaces ($\alpha = 90^o$in Fig.5 (a)) but does not hold for slanted surfaces or curved surfaces. In practice, it will lead to a ladder effect in reconstructing non-frontoparallel surfaces, refer to Fig.6.

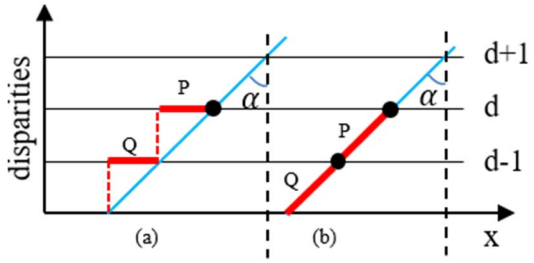
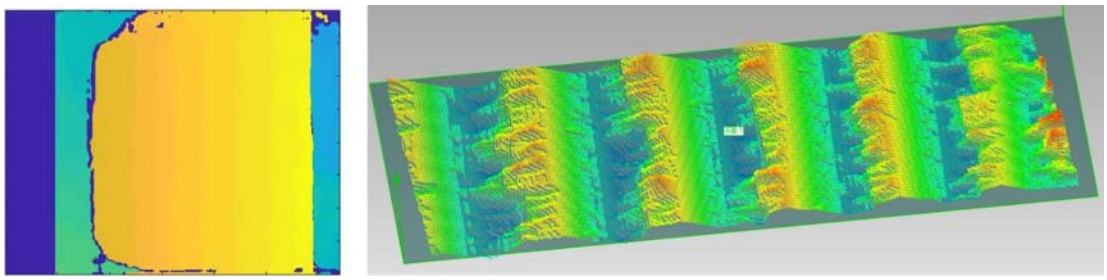


Fig. 5 Disparity searching in 1D. The points of the blue surface shall be reconstructed. The disparity search path was shown by red bars. (a) Disparity estimation route with SGM. (b) Ideal disparity estimation route.

To alleviate this kind of bias, a new penalty term is defined as in Eq. (5). In this case, equal penalties were utilized for unchanged disparities and disparities that change only 1 pixel.

$$V'(d_p, d_{p-r}) = \begin{cases} 0, if\ |d_p - d_{p-r}| \le 1 \\ P_2, if\ |d_p - d_{p-r}| > 1 \end{cases} \quad (5)$$

By the above improvement, the ladder phenomenon can be alleviated effectively, as shown in Fig.7. The improved SGM plays an important role in improving the 3D measurement accuracy of speckle-based SLT.



(a) depth map of a slanted plane (b) reconstructed result of (a)

Fig. 6 Ladder effect of SGM for slanted surfaces.

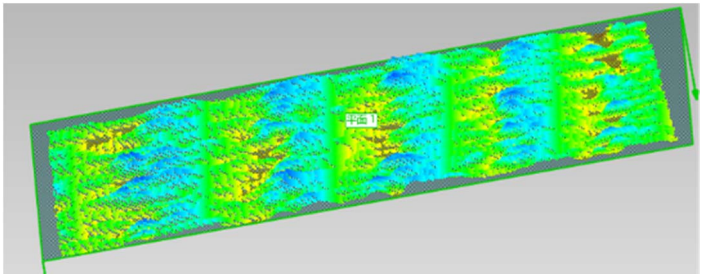


Fig. 7 The proposed optimal SGM for slanted surfaces.

D. 3D reconstruction

Based on the above Sections, optimal stereo parameters can be obtained and stereo images can be rectified. Based on the proposed stereo matching algorithm, the optimal depth map can be achieved. Then, 3D information can be achieved based on the rectified stereo parameters and depth information, as shown in Eq. (6).

$$\begin{cases} x = z(u^L - u_0)/f \\ y = z(v^L - v_0)/f \\ z = Bf/(d - (u_0^L - u_0^R)) \end{cases} \tag{6}$$

Here $[u_0^L, v_0^L]$ denotes the image center coordinates of the rectified left camera, and $[u_0^R, v_0^R]$ denotes the image center coordinates of the rectified right camera. $[u^L, v^L]$ represents every pixel coordinates on the depth map and d is the value of estimated depth at $[u^L, v^L]$. The symbol of B denotes the baseline of the stereo vision system.

IV. Experimental Results

A. Experimental setup

A prototype of our system is shown in Fig. 8, which consists of one-two cameras with a resolution of 1280×960 and a DOE projector. The size of each camera is 12mm×12mm and the baseline is 50mm. The DOE projector is of size 8×8 mm, with 12000 pseudorandom speckle dots etched on it. As can be seen in Fig.8, two other common 3D depth cameras, the DS460 of Percipio Company and the Intel RealSense D435 were installed below our system to provide comparison data.

The speckle pattern was put in Fig.9 (a), and its projection effect was put in Fig.9 (b). The DOE device is attached to a laser source that works at 850 nm with a power of 0.2W. The projected pattern was designed by using pseudo-random speckle points. The depth of focus of the DOE projector is 0.03m~1.6m, and the efficient working distance of our system is 0.03m~1.60m. To sum up, the whole size of our miniature 3D camera sensor is below 60mm and the hardware cost is below 150 dollars, which makes it easy to integrate into intelligent terminals.

Fig. 8 Experimental setups.

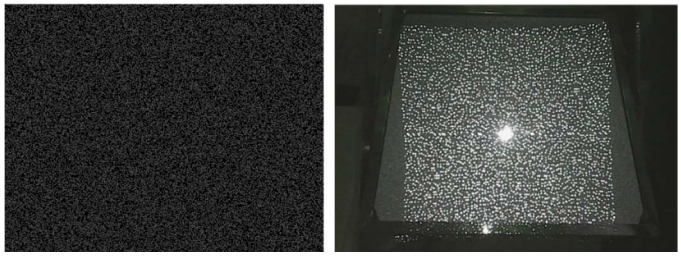
B. Accuracy comparison with other systems

To test the reconstruction accuracy of our system, two other common 3D depth cameras, the DS460 of Percipio Company and the Intel RealSense D435 were utilized as a comparison. Without loss of generality, a standard plane and a sphere were utilized as the target objects, as displayed in Fig.10. The plane is of size 20cm×20cm and its machining accuracy can be up to ±5μm. The sphere is with a diameter of 15cm and its machining accuracy can be up to ±20μm.

The 3D-point set P is computed by our method, and $dist(i, \Pi_p)$ denotes the distances of the point $p_i \in P$ and the fitting plane Π_p. Plane fitting error is obtained by summing

$dist(i, \Pi_p)$, as shown in Eq. (7). Sphere fitting error is the same as above.

$$\begin{cases} e_{plane} = \sum_i dist(i, \Pi_p) \\ e_{sphere} = \sum_i dist(i, \Pi_s) \end{cases} \quad (7)$$



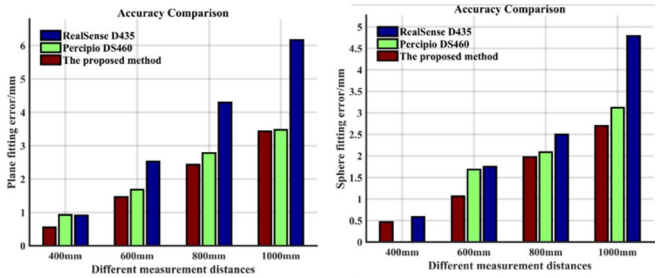
(a) The designed speckle pattern (b) speckle pattern projected by DOE

Fig. 9 The speckle pattern designed in this paper.

(a) standard plane (b) standard sphere

Fig. 10 Accuracy comparison between DS460, D435, and our system

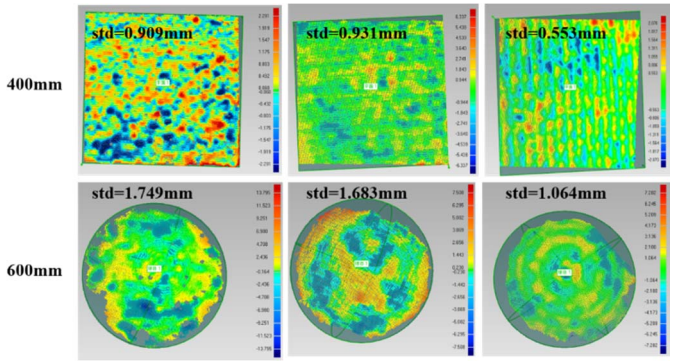
Results at different measurement distances (0.4m, 0.6m, 0.8m, and 1.0m) are put together and compared as below in Fig.11.



(a) Plane fitting errors (b) Sphere fitting errors

Fig. 11 Accuracy comparison between DS460, D435 and our system

Note that DS460 does not work for close range, so part of the sphere at 0.4m cannot be measured. Corresponding data were missing in Fig.11 (b). Results showed that no matter for the 2D plane or 3D sphere, our proposed system achieves better accuracy than the other two systems at different distances. To save some space, the results of the plane data at 0.4m and the sphere data at 0.6m were taken out and displayed in Fig.11. At a close range of 0.4m, the superiority of the proposed system is more obvious. Its measurement error of the plane is 0.553mm. Yet the error of Intel RealSense D435 and Percipio DS460 is 0.909mm and 0.931mm respectively, which is 64.4% and 68.4% higher than our system respectively. For the sphere data, the measurement error of the proposed system is 1.064mm, yet the error of Intel RealSense D435 and Percipio DS460 is 1.749mm and 1.683mm respectively, which is 64.4% and 58.2% higher than our system respectively. It is safe to say that our system has an obvious advantage in measurement accuracy.



(a) RealSense D435 (b) Percipio D460 (c) the proposed system

Fig. 12 Accuracy comparison with other systems

C. *Reconstruction of complex surfaces and dynamic objects*

In this part, experiments are conducted to test our system by reconstructing surfaces with rich textures and dynamic objects. First is a skirt with rich color textures. Originally captured image was put in Fig.13(a). Corresponding estimated depth map based on our improved SGM was put in Fig.13 (b). Other results, such as point clouds based on triangulation, point cloud with textures and 3D model are displayed in Fig.14. It can be seen that our system works well on complex surfaces.

(a) original image (b) estimated depth map

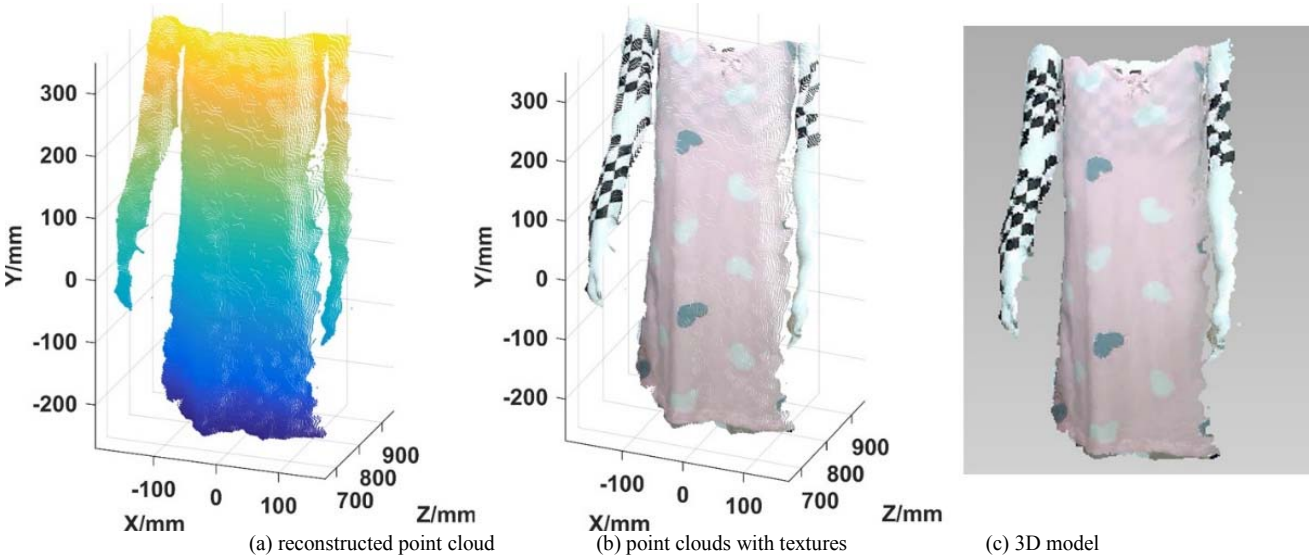
Fig. 13 Depth estimation results of surfaces with rich textures by using the proposed system.

Dynamic objects, a waving arm, was also reconstructed by using our system. To save some space, only three gestures of the waving arm were taken out and put in Fig.15. The results of the estimated depth map are also displayed in Fig15. Corresponding results, of reconstructed point clouds, based on triangulation, point cloud with textures and 3D model are displayed in Fig.16. It shows that rather good results can be achieved.

V. CONCLUSIONS

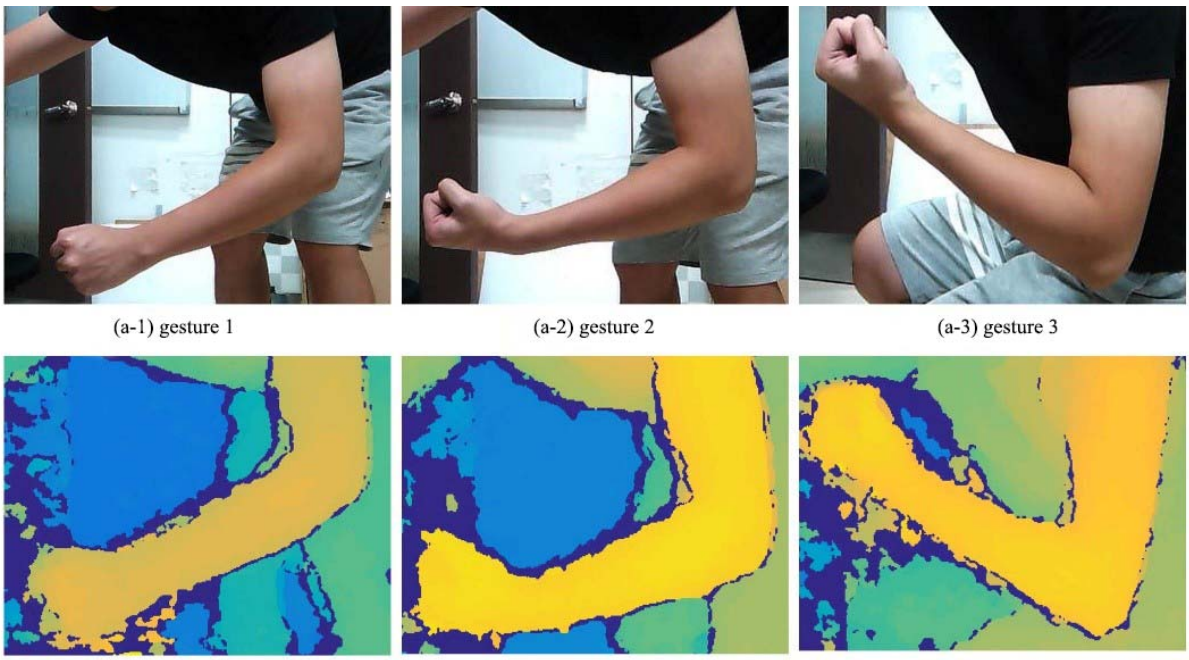
This paper proposed a structured light system based on pseudo-random speckles, which is constituted by two cameras and a micro DOE projector. A single pattern with pseudo-random speckles was projected onto the surface of the target object and the coupling pattern was imaged by the stereo cameras. An

optimal camera calibration algorithm is adopted to guarantee the reconstruction accuracy. The diffractive optical element (DOE) is of size 8mm×8mm, and the baseline between stereo cameras is as short as 50mm. The system has the advantages of simple structure, small size, and low cost. It can be easily



(a) reconstructed point cloud (b) point clouds with textures (c) 3D model

Fig. 14 Reconstruction result of a skirt by using the proposed system. From left to right are reconstructed point clouds, point clouds with textures and 3D model respectively.



(a-1) gesture 1 (a-2) gesture 2 (a-3) gesture 3

(b-1) depth map of gesture 1 (b-2) depth map of gesture 2 (b-3) depth map of gesture 3

Fig. 15 Depth estimation results of a waving arm by using the proposed system

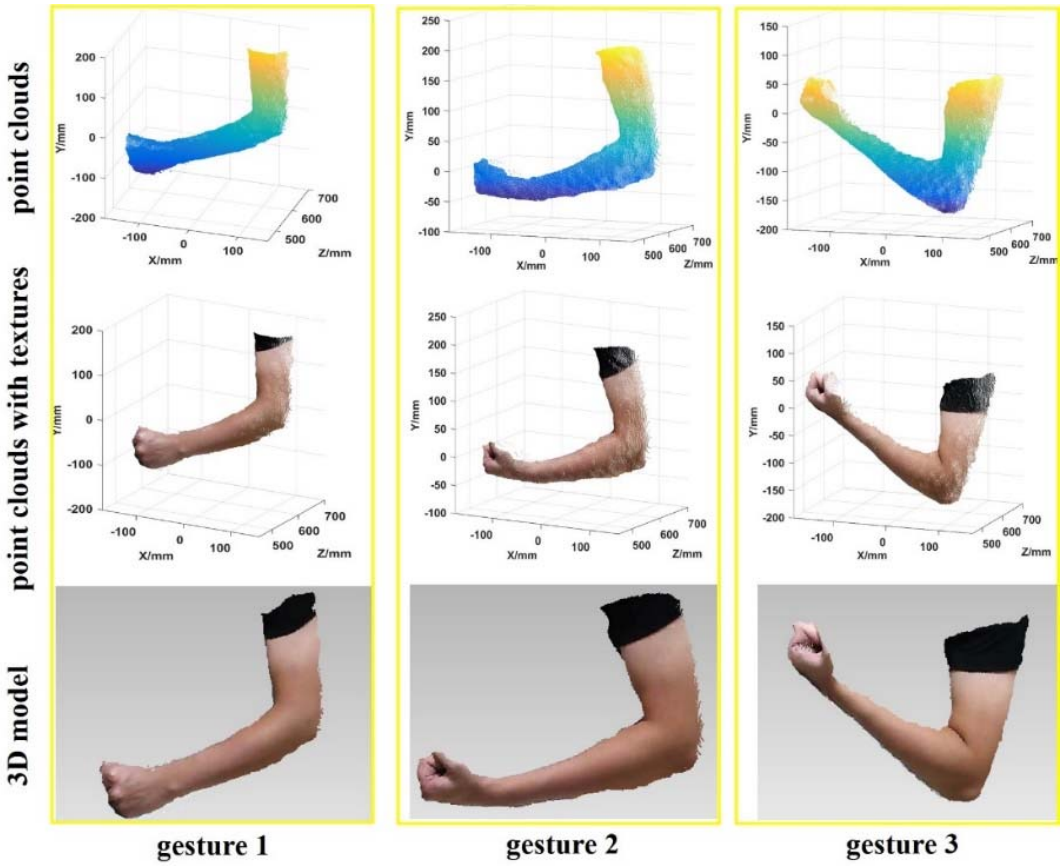


Fig. 16 Reconstruction result of a waving arm by using the proposed system. From top to bottom are reconstructed point clouds, point clouds with textures and 3D model respectively.

integrated into mobile devices such as tablet computers and mobile phones. In comparison with several other mature products, the proposed 3D vision system can achieve 3D reconstruction with higher accuracy. In our next step, accuracy improvement in the case of short baselines will be deeply researched.

ACKNOWLEDGMENT

This work was supported in part by the National Key R&D Program of China (No. 2017YFB1103602), the NSFC-Guangdong (No.2018A030310668, 2017A030310474), the Shenzhen Science Plan (ZDSYS201707271637577, KQJSCX20170731165108047, JCYJ20170413152535587), and the NSFC-China (Nos. 51705513, U16132133).

REFERENCES

[1] Chen, C., Li, J., Luo, J., Xie, S., Pu, H., Cui, Z., and Gu, J., Analysis, and simulation of the kinematics of 5-DOF nuclear power station robot manipulator. In 2014 IEEE International Conference on Robotics and Biomimetics, 2014:2025-2030, IEEE.

[2] Stančić, I., Musić, J., and Zanchi, V., Improved structured light 3D scanner with application to anthropometric parameter estimation. Measurement, 2013, 46(1): 716-726.

[3] Conradt, J., On-board real-time optic-flow for miniature event-based vision sensors. In 2015 IEEE International Conference on Robotics and Biomimetics, 2015:1858-1863, IEEE.

[4] Park, J., Kim, H., Tai, Y. W., Brown, M. S., and Kweon, I., High-quality depth map upsampling for 3d-tof cameras. IEEE International Conference on Computer Vision, 2011: 1623-1630.

[5] Khamis, S., Fanello, S., Rhemann, C., Kowdle, A., Valentin, J., and Izadi, S., Stereonet: Guided hierarchical refinement for real-time edge-aware depth prediction. In Proceedings of the European Conference on Computer Vision, 2018: 573-590.

[6] V Van der Jeught, S., and Dirckx, J. J., Real-time structured light profilometry: a review. Optics and Lasers in Engineering, 2016, 87: P. 18-31.

[7] Boyer, K. L., and Kak, A. C., Color-encoded structured light for rapid active ranging. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1987, (1): 14-28.

[8] Du, P. Q., Shih, H. F., Chen, J. S., and Wang, Y. S., Design and verification of diffractive optical elements for speckle generation of 3-D range sensors. Optical Review, 2016, 23(6): 1017-1025.

[9] Silberman, N., and Fergus, R., Indoor scene segmentation using a structured light sensor. In 2011 IEEE international conference on computer vision workshops, 2011, 601-608, IEEE.

[10]Wu Q., Liu S., Huang X., Han L., Guo T. and Wang M., 3D motion sensing interaction system based on speckles. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(7): 1105-1114.

[11] Tang Q., Liu X., Wu Y., Liu X., Liu M., Li Q., and Peng X., Analysis of Speckle Characteristics in Digital Speckle 3D Reconstruction. Chinese Journal of Lasers, 2018, 45(10):180-187.

[12] Choi, S. M., Jeong, J. C., and Chang, J., Dense 3D depth map with DOE pattern. In 2012 9th International Conference on Ubiquitous Robots and Ambient Intelligence, 2012, 34-37. IEEE.

[13]Lin Y., Lv N., Lou X., and Dong M., Robot vision system for 3D reconstruction in low texture environment, Optics and Precision Engineering, 2015, 23(2): 540G549.

[14] Fusiello, A., Trucco, E., and Verri, A., A compact algorithm for rectification of stereo pairs. Machine Vision and Applications, 2000, 12(1): 16-22.

[15]Bleyer M, Rhemann C, Rother C. PatchMatch Stereo-Stereo Matching with Slanted Support Windows. Bmvc. 2011, 11: 1-11.

[16] Scharstein D , Taniai T , Sinha S N . Semi-Global Stereo Matching with Surface Orientation Priors. International Conference on 3D Vision (3DV). IEEE Computer Society, 2017.
[17] Sinha S N, Scharstein D, Szeliski R. Efficient high-resolution stereo matching using local plane sweeps. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1582-1589.
[18] Hane C , Ladicky L , Pollefeys M . Direction matters: Depth estimation with a surface normal classifier. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.
[19] Zhang Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000, 22(11): 1330-1334.
[20] Gu F., Zhao H., Ma Y and Bu P. Camera calibration based on the back projection process. Measurement Science and Technology, 2015. 26(12): 125004.
[21] Gu, F., Zhao, H., Ma, Y., Bu, P., and Zhao, Z., Calibration of stereo rigs based on the backward projection process. Measurement science and technology, 2016, 27(8): 085007.
[22] Hirschmuller, H., Stereo processing by semiglobal matching and mutual information. IEEE Transactions on pattern analysis and machine intelligence, 2007, 30(2): 328-341.