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

Highly Reflective Surface Measurement Based On Dual Stereo Monocular Structured Light System Fusion*

Daochuan Wang*
School of Mechanical Engineering and Automation
Harbin Institute of Technology, Shenzhen
Shenzhen, China
wangdaochuan@stu.hit.edu.cn

Kejing He*, Congying Sui and Congyi Lyu
Department of Mechanical and Automation Engineering
The Chinese University of Hong Kong
HKSAR, China
kjhe@mae.cuhk.edu.hk

Yinping Lai and Huiwen Guo
Smarteye Technology Co., Ltd.
Shenzhen, China
laiyinping@smarteyetech.com

Yun-Hui Liu, *IEEE Fellow*
Department of Mechanical and Automation Engineering
The Chinese University of Hong Kong
HKSAR, China
yhliu@mae.cuhk.edu.hk

***Abstract*—3D surface measurement plays an important role in robotics. In this paper, we provide a highly reflective surface measurement method based on combination of dual stereo monocular structured light system fusion. A predetermined rigid transformation matrix is utilized to improve fusion quality for the fusion area. Moreover, a final decision map is created to avoid the reconstruction of repeating regions of left and right system for ensuring the accuracy of the non-fusion area to achieve monocular system accuracy. The validity and accuracy of our algorithm are verified by designed experiments. Furthermore, various complex objects are measured including a metal surface, a phone case and a motor component, which show the proposed method has the capability to solve the highly reflective surface measurement problem and the occlusion problem.**

***Index Terms*—3D measurement, reflective surface, structured light, dual stereo monocular system, fusion**

I. Introduction

With the development of robotic automatic picking and sorting, autonomous assembly and industrial inspection technology, the industry is increasingly demanding high-precision 3D information for object recognition, surface detection and quality control[1-4]. Even in common tasks such as robotic automatic picking and sorting and industrial inspection, it has been recognized that high-precision 3D measurements must be used to enhance existing 2D methods to another level. However, there are many objects with highly reflective surfaces in practical applications, which poses additional challenges to 3D measurement technology.

The 3D reconstruction of highly reflective surface is a challenging issue in computer vision. Its goal is to obtain 3D information of the points on the surface of the highly reflective object and accumulate these points to form a reconstructed surface. In the case where the target surface reflectance is high, the projected pattern in the captured image is usually greatly degraded. In the overexposed and underexposed image areas, the projected image information is not well extracted, which seriously affects the accuracy of the 3D measurement. Traditionally, the method that is often used is to spray a thin anti-reflection coating on the surface of the measured object to change the reflection property of the surface into a diffuse reflection surface, and then we can measure the target. This method requires spraying and cleaning the coating, which reduces the measurement efficiency; in addition, the thickness and uniformity of the coating depend on the experience and skill of the operator, it is easy to bring measurement error [5]. More importantly, this treatment complicates the measurement process and may also cause the target surface to be corroded by the coating. Therefore, this method makes the existing structured light 3D reconstruction technology not practical in many applications.

In recent years, many scholars have conducted in-depth research on how to use optical methods to directly reconstruct the 3D surfaces of highly reflective objects and many solutions have been proposed. They are mainly divided into two types active and passive 3D reconstruction methods. In terms of active 3D reconstruction, there have been many research results in recent years. Y.Yoshinori et al. proposed a method for measuring polarization extinction [6], but this method eliminates strong reflections and also makes the non-reflective area underexposed which makes the non-reflective unable to be measured. Song Zhang et al. [7] proposed a high dynamic range scanning technology based on the three-step phase shift method. By changing the aperture value or exposure time, a series of stripe images with successively decreasing exposures are acquired, and then the brightest and unsaturated pixel gray value is selected pixel by pixel to synthesize the final fringe image, which is used for phase resolution. However, this method is difficult to obtain a complete high reflective surface 3D information, and it can not meet the real-time requirements. Zhan Song et al.[8] proposed a coding strategy based on stripe

*This work is supported in part by the Shenzhen Peacock Plan Team Grant(KQTD20140630150243062), and in part by the HK RGC under T42-409/18-R and 14202918, in part by project 4750352 of the CUHK-SJTU Joint Research Fund, in part by the VC Fund 4930745 of the CUHK T Stone Robotics Institute, CUHK. D. Wang and K. He contributed equally. Corresponding author: Yun-Hui Liu, Email:yhliu@mae.cuhk.edu.hk

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

edges, which enhances the robustness of the system to the reflection characteristics of highly reflective objects. However, this method can not be conducted in fast speed. Sui-Kei Tin et al.[9] presented a two-layer LCD setup as active illumination source for recovering the shape of mirror-type objects, and developed single view approach for estimating the initial scale factor and a multiview joint optimization approach to refine the scale factors. However, the reflected images of the method only covers a limited area on the surface of the object, and can only perform 3D reconstruction for small and medium-sized objects, and the device of the method is relatively large in volume. Ding Liu et al.[10] propose a frequency-based 3D reconstruction method, which incorporates the frequency-based matting method. Each pixel of the captured image is analyzed along the time axis and the corresponding signal is transformed to the frequency-domain using the Discrete Fourier Transform. However, this method is not good for reconstructing objects with high curvature or high brightness, and cannot measure the object in real-time.

In this work, we propose a highly reflective surface measurement method based on combination of dual stereo monocular structured light system fusion. We consider the dual-camera structured light system to be composed of two camera-projector monocular structured light systems to obtain different viewing angles for scanning. The 3D data of the right camera is used to supplement the 3D data of the left camera that is lost due to the high reflective surface being too bright, so as to realize the fusion of the left and right camera 3D data. A predetermined rigid transformation matrix is utilized to improve fusion quality for the fusion area. Moreover, a final decision map is created to avoid the reconstruction of repeating regions of left and right system for ensuring the accuracy of the non-fusion area to achieve monocular system accuracy.

The remaining part of this paper is organized as follows. Section II briefly introduces the principle and process of the dual monocular coordinate system fusion. Section III describes the principle of stereo imaging. Section IV shows our experiments and gives some experimental results. Section V concludes the paper.

II. Dual Monocular Coordinate System Fusion

To calibrate the monocular system, the projector is treated as the inverse camera including pinhole model with distortion coefficient [11]. Then the camera and the projector in the monocular system are calibrated through Zhang's method [12].

The difficulty of fusion is to put the two monocular systems unified into one coordinate system. Ideally, the two monocular systems are built in the same projector world coordinate. However, due to the calibration errors, the obtained coordinate systems are not completely coincident with each other. In order to eliminate the pose deviation between the two systems, in this paper, we propose a novel method which is to predetermine a rigid transformation matrix to align the two coordinate systems as shown in Fig. 1. The rigid transformation matrix $\widehat{H}$ can be found by minimizing :

$$\widehat{H} = \arg\min_{H} \sum_{\forall p} \| p - Hq \|^2 \tag{1}$$

$$H = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \in R^{4\times 4}$$

$$p = [x_{lp}, y_{lp}, z_{lp}, 1]^T$$

$$q = [x_{rp}, y_{rp}, z_{rp}, 1]^T$$

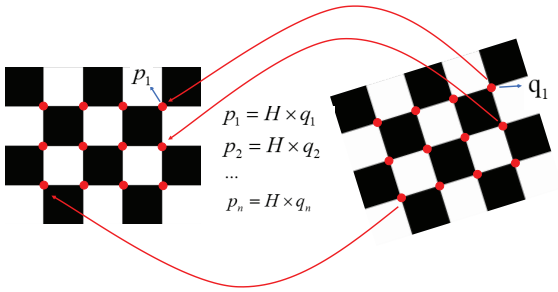

Fig. 1. Rigid transformation matrix $\widehat{H}$ are estimated to align right monocular system to left monocular system.

A rigid transformation matrix $\widehat{H}$ is obtained as a transformation from right coordinate system to left coordinate system to correct the pose deviation of two systems. The 3D point q is the checkerboard corner in the projector world coordinate of right system and 3D point p represents the corresponding checkerboard corner in the projector world coordinate of left system. For the subsequent 3D reconstruction points, we apply this rigid transformation matrix $\widehat{H}$ to the right monocular system:

$$p_r = \widehat{H} \times q_r \tag{2}$$

where q_r is the reconstructed 3D point in the projector world coordinate of right system and p_r is the 3D point in the projector world coordinate transformed from right system to left system.

An experiment is designed to verify the importance of this operation. Evaluations of the improvement in accuracy are presented later.

III. Stereo Imaging Principle

For a preliminary understanding of our approach, we provide a schematic diagram shown in Fig. 2. In the traditional dual stereo monocular system fusion method, two reconstructed point clouds are combined together directly as the final result, which causes many problems. First, recovering the repeating regions of left and right system increases the measurement time. Second, data overlap in the repeating areas of dual monocular system affects the experimental result accuracy. Third, direct fusion without any refinement causes low accuracy in the fusion area. To overcome the three problems above, our overall idea is that based on the result of left monocular system, we utilize the result of right monocular system to fill the unreconstructed area of left system. After

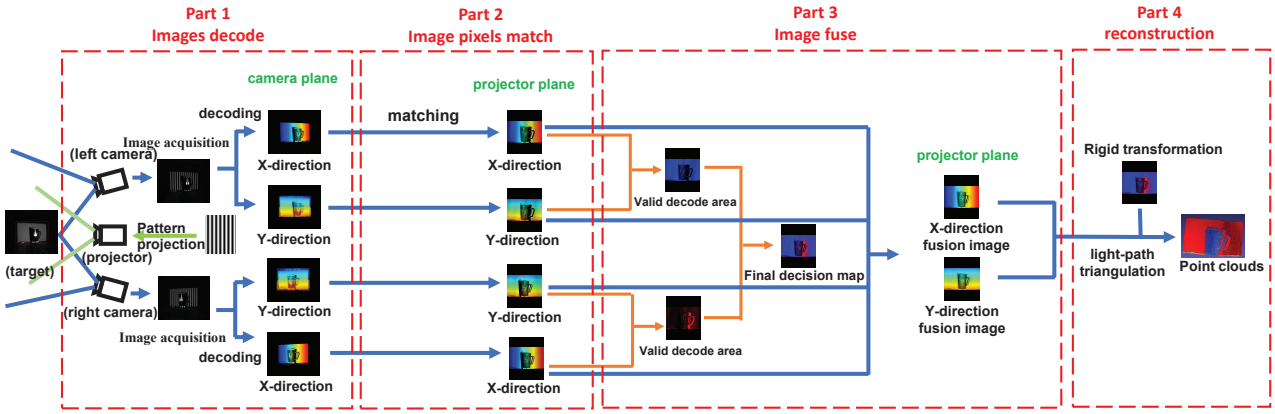

Fig. 2. Principle of the dual stereo monocular system fusion.

the reconstruction, a rigid transformation method is applied to the fusion area for improving the fusion quality. Based on this idea, the whole algorithm is divided into 4 steps as follows :

1) After image acquisition, The images received from the two monocular systems are decoded separately by contrast gray code method, which is introduced in III A.
2) Sub-pixel adaptation transformation method is applied to reach the goal of matching the corresponding pixels between camera image and projector image, which is introduced in III B.
3) Based on a certain principle presented in III C, the undecoded matching region of the left system is filled by the decoded matching area of right monocular system to form the matching fusion image.
4) the light-path triangulation [14] is used to reconstruct the three-dimensional information and the rigid transformation is applied for the filling area to improve the fusion accuracy, which is presented in III D.

A. Contrast Gray Code Method

Gray code method is a classical technique to encode the object and has the advantage of robustness [13] as shown in Fig. 3. Since gray code method is binary decoded according to the gray value of the received images, different parts of the surface have different reflectivity causing we can't use a single threshold to decode the received images. Therefore, both the patterns and their inverse are cast onto the objects in our method. In this way, surface with weak or high response to the light intensity of patterns could be both evaluated by intensity contrast, which is defined as below:

$$IC = \frac{I - I_{inv}}{I + I_{inv}} \tag{3}$$

where IC, I and I_{inv} are the intensity contrast, received intensity of the pattern and received intensity of its inverse pattern. Similar as the gray code method, m pairs (one pair means one pattern and one inverse of the pattern) of patterns can encode 2^m pixels of the image. Since the resolution of the

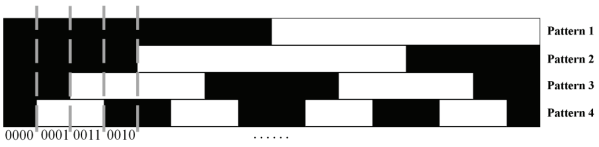

Fig. 3. Gray code structured light patterns.

projector image is 912×1140, we project 10 pairs of patterns to encode the x-direction pixels and y-direction pixels of the image separately. Therefore, totally 40 patterns are cast onto the objects. The received images are decoded one pixel by one pixel as follows :

$$P^c(u,v)|x_{IC} = \sum_{i=1}^{m=10} 2^{i-1} \times sgn(IC_i^x(u,v) - thr) \tag{4}$$

$$P^c(u,v)|y_{IC} = \sum_{i=1}^{m=10} 2^{i-1} \times sgn(IC_i^y(u,v) - thr) \tag{5}$$

$$sgn(x) = \begin{cases} 0 & x > 0 \\ 1 & x <= 0 \end{cases}$$

where (u, v) represents the coordinate of a point on the camera image. $IC_i^x(u,v)$ and $IC_i^y(u,v)$ are the intensity contrast of (u, v) calculated by the i_{th} pair of images in the x-direction and the y-direction, respectively. $P^c(u,v)|x_{IC}$ and $P^c(u,v)|x_{IC}$ represent the contrast gray code of (u, v) in x-direction and y-direction, respectively. Last, $m = 10$ means we receive 10 pairs of images for x-direction and y-direction, separately.

B. Sub-pixel Adapted Transformation

After received images are decoded one pixel by one pixel, we can acquire the x-direction decoded image and the y-direction decoded image in the camera plane. Since the

contrast gray code of every pixel in the projector plane is known in advance, the matching relations of pixels between the camera and the projector can be obtained through applying look-up-table method to contrast gray code. It is worth noting that the camera resolution is 2448×2048 and projector resolution is 912×1140. Since the camera resolution is much larger than projector resolution, several pixels on the camera plane may match the same pixel on the projector plane. For precise registration, the sub-pixel adapted transformation is determined as follows:

$$P(x,y)\,|\mathrm{u} = \frac{\sum_{i=1}^{n} P_i^c(\mathrm{u}_i, v_i)|u_i}{n} \tag{6}$$

$$P(x,y)\,|v = \frac{\sum_{i=1}^{n} P_i^c(u_i, v_i)|v_i}{n} \tag{7}$$

Where (x, y) represents the coordinate of a point on the projected image and (u, v) represents the coordinate of a point on the camera image. $P(x, y)|u$ and $P(x, y)|v$ are horizontal and vertical coordinate of the sub-pixel matching point on the camera image of (x, y). $P_i^c(u_i, v_i)|u_i$ and $P_i^c(u_i, v_i)|v_i$ stand for horizontal and vertical coordinate of the i_{th} matching point on the camera image of (x, y). n stand for the number of matching pixels to (x, y).

C. Decision Map Fusion

Fig. 4(a)(b) are the decoded images on the projector plane from left system and right system. It can be observed that, shiny areas on the object cannot be well reconstructed by using only monocular system. When left and right system are evaluated together, we can acquire their own valid matching area which means their respective matching successful area between the projector image plane and camera image plane. The right system decoded image region is used for supplementation to the undecoded image region of left system based on their own valid matching area. A decoded image of larger effective area (Fig. 4(c)) is obtained in this way accompanied by a final decision map diagram (Fig. 4(f)) to distinguish the valid area from which system. According to the final decision map, the repeating decoded regions of left and right system is reconstructed by left system once as the non-fusion area to achieve monocular system accuracy. For the undecoded region of left system, the decoded region of right system is filled as the fusion area to reconstruct it which is the red area shown in Fig. 4(f). Therefore, from the final decision map, we can distinguish that the corresponding decoded fusion image area is from which system.

D. 3D Reconstruction with Rigid Transformation

Trough light-path triangulation [13], we can transform decoded matching point to 3D info in the calibrated 3D coordinates which is the projector world coordinate. Then the predetermined rigid transformation matrix $\widehat{H}$ is applied to correct posture deviation for the reconstruction points of

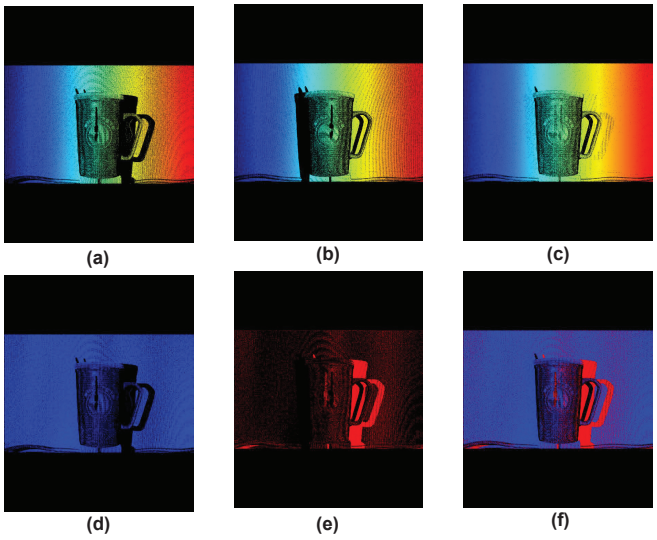

Fig. 4. Final decision map of a metal cup. (a) X-direction decoded image of the left system. (b) X-direction decoded image of the right system. (c) X-direction decoded fusion image. (d) Valid matching area in left system (blue area represents left system valid area). (e) Valid matching area in right system (red area represents right system valid area). (f) Final decision map (blue area represents left system valid area and red area represents right system valid area).

right system which is red color area shown in Fig. 4(f) as follows :

$$p_r = \widehat{H} \times q_{red} \tag{8}$$

where q_{red} represents 3D point reconstructed from red area in the decision map .

IV. Experiments

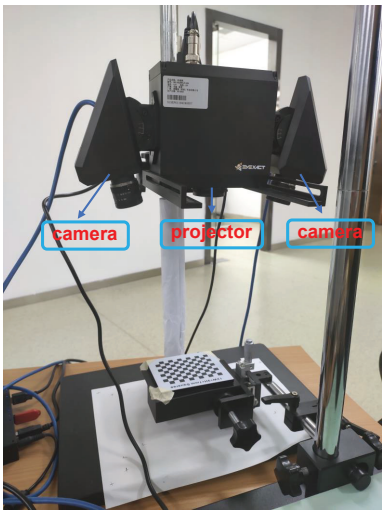

Fig. 5. The experiment system utilized in this paper.

To validate the performance of the proposed method, various experiments are carried out on experiment system shown

in Fig. 5. The system consists of two cameras (MER-502-79U3M) and one projector (DLP Lightcrafter 4500). The resolutions of the cameras and the projector are 2448×2048 and 912×1140, respectively. The system takes about 2.5 s to acquire all projected images. After image acquisition, the algorithm is completed to reconstruct approximately 1 million points by a program written in C++ within 0.5 s on a standard PC platform (CPU: Intel Core i7-7700 3.6 GHz, 8 GB RAM).

A. Stereo Monocular System Error Analysis

In this experiment, we measure the depth of a plane whose motion is known to evaluate the stereo monocular system accuracy. A plane is mounted on the spiral micrometer with the motion accuracy of 0.01 mm. This plane is moved 11 times toward the projector with steps of 0.1mm, as shown in Fig. 6. After each motion, our system reconstruct the plane 5 times within a certain time range. Totally 60 planes are reconstructed.

The depth of the plane center is measured after every motion. The average displacement deviations of the whole motion steps are presented in Fig. 7(a). Furthermore, the root mean square error of the five measurement results of the same plane position is acquired, as shown in Fig. 7(b). The results measured are listed in Table I. The result shows the maximum absolute error of 0.010mm and the maximum root mean square error of 0.006mm. Notice that two types of error are very close even lower than the motion accuracy of spiral micrometer, indicating that the depth measurement accuracy and precision are both approaching the motion accuracy of 0.01mm.

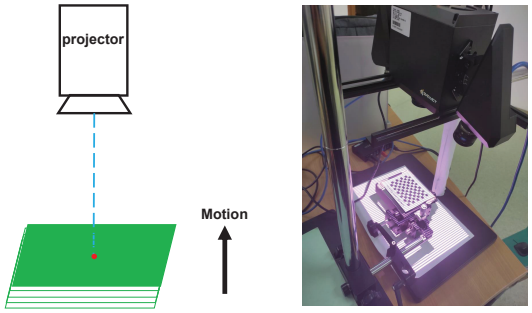

Fig. 6. Plane was raised 11 times toward the projector with steps of 0.1 mm.

TABLE I
PLANE DISPLACEMENT MEASURED AT EVERY MOTION STEP
(ACCURATE TO THE MICROMETER LEVEL)

No.	0	1	2	3	4	5
Distance(mm)		0.997	0.990	0.998	1.008	0.998
RMSE(mm)	0.003	0.002	0.006	0.003	0.002	0.002
No.	6	7	8	9	10	11
Distance(mm)	1.006	1.010	0.997	0.994	1.003	0.994
RMSE(mm)	0.002	0.003	0.001	0.004	0.005	0.003

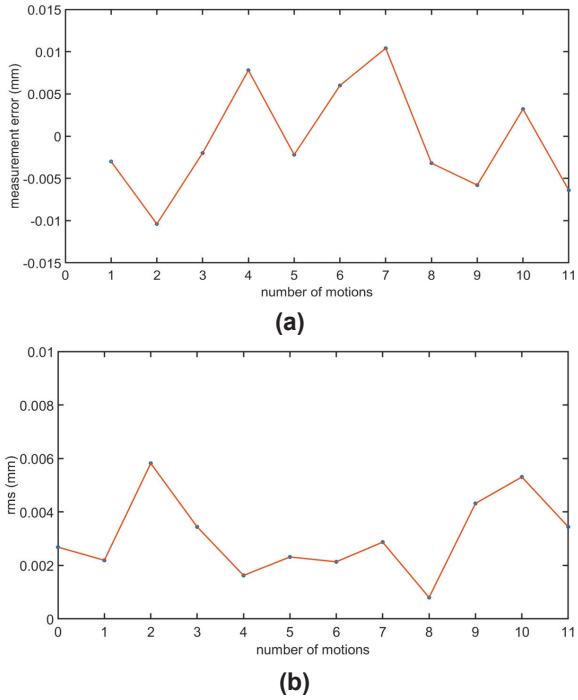

Fig. 7. Error plot of the measured plane displacement at every motion. (a) Depth difference compared with the motion step distance. (b) Mean square error of each measurement.

B. Fusion Error Analysis with Rigid Transformation

To illustrate the importance of rigid transformation method for improving the accuracy of the fusion area, an experiment is designed to compare the fusion errors of measurement with rigid transformation method and without rigid transformation method. Fig. 8(a) and 8(b) show the captured images of a double-layered cylindrical block from the left camera and right camera, respectively. Processed by algorithm part 1 and part 2, The left system decoding matching diagram and the right decoding matching diagram is acquired. Then the point clouds of the left system and right system are reconstructed in the projector world coordinate separately. Since the matching images are in the same projector coordinate plane, the 3D coordinate distance difference of each corresponding pixel under the projector plane of the dual monocular system is obtained through calculation. Fig. 8(c) and Fig. 8(d) show the distance difference of dual monocular system without applying rigid transformation method and distance difference of dual monocular system with applying rigid transformation method, respectively. The distributions of distance difference through two methods are separately counted and the corresponding histogram is shown in Fig. 9. The specific distance difference results of dual monocular system are presented in Table II.

The Table II demonstrates that the method with rigid transformation method can improve the average fusion area error from 0.798 mm to 0.271 mm, which means the rigid transformation method takes significant effect on the accuracy

of the results of the fusion region.

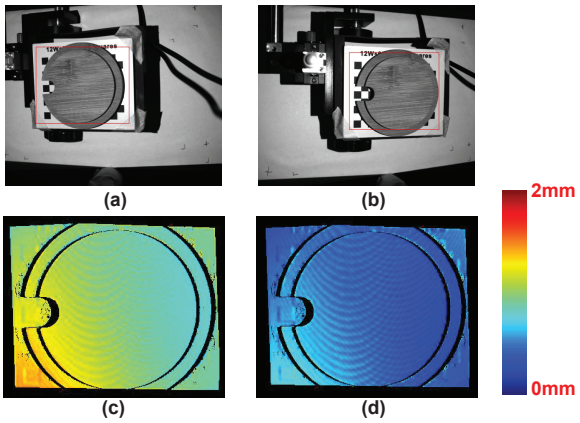

Fig. 8. Fusion area errors analysis. (a) Region of interest in left image (red square interior). (b) Region of interest in right image (red square interior). (c) 3D distance difference of dual monocular system without applying rigid transformation method. (d) 3D distance difference of dual monocular system with applying rigid transformation method.

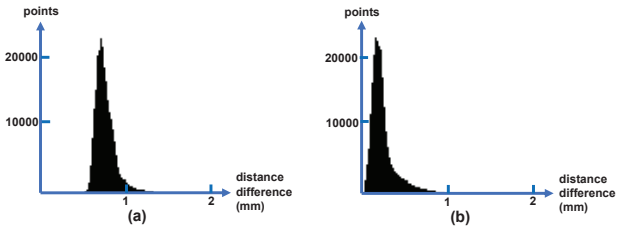

Fig. 9. Distribution histograms of distance difference through two methods. (a) Distribution histogram of distance difference through the algorithm excluding rigid transformation method. (b) Distribution histogram of distance difference through the algorithm including rigid transformation method.

TABLE II
DISTANCE DIFFERENCE RESULTS OF DUAL MONOCULAR SYSTEM

Method	I	II
Average difference (mm)	0.798	0.271
Maximum difference (mm)	1.615	0.994
Minimum difference (mm)	0.462	0.021
Points	209851	209851

Method I represents the algorithm excluding rigid transformation method. Method II represents the algorithm including rigid transformation method.

C. Reconstruction of Complex Objects

More experiments are carried out to verify the performance of the proposed experiment system. Fig.10(a)(b) are captured images from a perforated reflective metal surface. Fig. 10(c) is the reconstructed surface of the metal surface under the same camera exposure time. A more complex reflective surface of the phone case model is measured as shown in Fig. 10(d)(e). The point clouds are recovered through our method as shown in Fig. 10(f). A motor component shown in Fig. 10(i) is reconstructed to validate the algorithm ability of solving occlusion problem. Due to the hierarchical structure of this model, Fig. 10(g)(h) shows the middle part of the model is partial missing from the captured image of left camera and right camera, causing that the middle part of the object cannot be completely recovered through classical binocular 3D reconstruction. By combining the dual monocular system results, the intermediate area can be obtained as shown in Fig. 10(i).

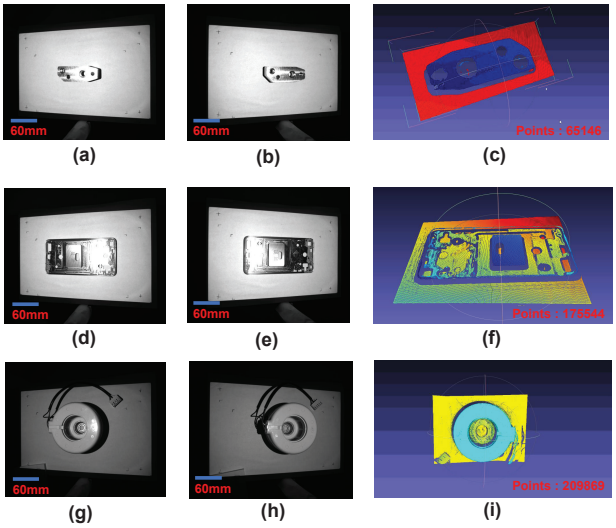

Fig. 10. Demonstration of 3D surface reconstruction of complex objects. (a)(b) The captured images of the reflective metal surface from left camera and right camera. (c) The reconstructed surface of the reflective metal surface. (d)(e) The captured images of the reflective phone case from left camera and right camera. (f) The reconstructed surface of the phone case. (g)(h) The captured images of the motor component from left camera and right camera. (i) The reconstructed surface of the motor component.

V. CONCLUSION

In this work, a dual stereo monocular system fusion method is proposed to solve the reflective surface measurement problem, which has merits of high reconstruction accuracy and robustness to noises. Moreover, a predetermined rigid transformation matrix obtained in the calibration step is utilized to improve fusion quality for fusion area. In addition, a final decision map is proposed to be a reference to avoid recovering the repeating regions of left and right system for ensuring the accuracy of the non-fusion area to achieve monocular system accuracy. Various experiments are conducted to verify the performances of the proposed scheme. The experiment of plane displacement measurement shows the accuracy of the non-fused area can reach 0.01mm, and the RMSE is less than 0.006mm. The experiment of fusion error analysis shows the rigid transformation method can decrease the fusion area error from 0.798mm to 0.271mm under a certain working distance. Experiments on the complex objects validate the algorithm has

the ability to reconstruct reflective surface and solve occlusion problem.

VI. Acknowledgement

The authors would like to thank Smarteye Technology Co., Ltd. for the experimental device support and the anonymous reviewers for their valuable comments.

References

[1] Z. Song, H. Jiang, H. Lin and S. Tang, "A high dynamic range structured light means for the 3D measurement of specular surface," Optics and Lasers in Engineering, vol. 95, pp. 8-16, 2017.

[2] M. Vzquez-Arellano, H. Griepentrog, D. Reiser and D. Paraforos, "3-D Imaging Systems for Agricultural Applications-A Review," Sensors, vol.16, 2016.

[3] G. Sansoni, M. Trebeschi and F. Docchio, "State-of-The-Art and Applications of 3D Imaging Sensors in Industry, Cultural Heritage, Medicine, and Criminal Investigation," Sensors, vol. 9, pp. 568-601, 2009.

[4] F. Chen, G.M. Brown and M. Song, "Overview of three-dimensional shape measurement using optical methods," Optical Engineering, vol. 39, pp. 10-22, Dec. 1999.

[5] D. Palouek et al., "Effect of matte coating on 3D optical measurement accuracy," Optical Materials, Feb. 2015.

[6] Y. Yoshinori et al., "Shape measurement of glossy objects by range finder with polarization optical system," Gazo Den shi Gakkai Kenkyukai Koen Yoko 200, pp. 43-50, Jan. 2003.

[7] S. Zhang and S.-T. Yau, "High dynamic range scanning technique," Optical Engineering, vol. 48, 2009.

[8] Z. Song, R. Chung, X.-T. Zhang, "An Accurate and Robust Strip-Edge-Based Structured Light Means for Shiny Surface Micromeasurement in 3-D," IEEE Transactions on Industrial Electronics, vol. 60, no. 3, pp. 1023-1032, March 2013.

[9] S.-K. Tin, J. Ye, M. Nezamabadi and C. Chen, "3D Reconstruction of Mirror-type Objects using Efficient Ray Coding," 2016 IEEE International Conference on Computational Photography, May 2016.

[10] D. Liu, X. Chen and Y.-H. Yang, "Frequency-Based 3D Reconstruction of Transparent and Specular Objects," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[11] Moreno, Daniel, and Gabriel Taubin, "Simple, accurate, and robust projector-camera calibration," 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission. IEEE, 2012.

[12] Z. Zhang, A Flexible New Technique for Camera Calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 13301334, Nov. 2000.

[13] Inokuchi and Seiji, "Range imaging system for 3-D object recognition," ICPR, pp. 806-808, 1984.

[14] Chari, Visesh, and Peter Sturm, "A theory of refractive photo-light-path triangulation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013.