

3D Surface Reconstruction Using A Two-Step Stereo Matching Method Assisted with Five Projected Patterns

Congying Sui¹, Kejing He¹, Congyi Lyu², Zerui Wang¹, and Yun-Hui Liu¹

Abstract—Three-dimensional vision plays an important role in robotics. In this paper, we present a 3D surface reconstruction scheme based on combination of stereo matching and pattern projection. A two-step matching scheme is proposed to establish reliable correspondence between stereo images with high computation efficiency and accuracy. The first step (coarse matching) can quickly find the correlation candidates, and the second step (precise matching) is responsible for determining the most precise correspondence within the candidates. Two phase maps serve as codewords and are utilized in the two-step stereo matching, respectively. The phase maps are derived from phase-shifting patterns to provide robustness to the background noises. Only five patterns are required, which reduces the image acquisition time. Moreover, the precision is further enhanced by applying a correspondence refinement algorithm. The precision and accuracy are validated by experiments on standard objects. Furthermore, various experiments are conducted to verify the capability of the proposed method, which includes the complex object reconstruction, the high-resolution reconstruction, and the occlusion avoidance. The real-time experimental results are also provided.

I. INTRODUCTION

3D machine vision is an important subject for robotics since it provides the 3D perception of the unknown surroundings [1]. The 3D surface information helps robots to improve the reliability in localization and recognition [2]. Due to its advantages, 3D machine vision has attracted extensive applications in automation systems, such as robotic soldering [3], mobile robots navigation [4], object grasping [5], [6], object manipulation [7], [8], and robot-assisted surgery [9]–[12], etc. The 3D surface reconstruction techniques can be classified into two categories: the passive methods and the active methods [13]. For passive ways, images acquired from original environment are utilized for 3D surface reconstruction. One representative example is the stereo vision [14], where correspondence is established between images captured from different perspectives. The major limitation of the passive methods is that the matching procedure highly relies on the features of the surface [13]. When there are no sufficient textures, such methods present low matching accuracy. On the other hand, the active methods avoid the problem of correspondence determination by actively projecting coded structured light onto the target. The patterns

This work is supported 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, and in part by the VC Fund 4930745 of the CUHK T Stone Robotics Institute, CUHK. (Corresponding author: Zerui Wang, Email: zrwang@mae.cuhk.edu.hk)

¹T Stone Robotics Institute and Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong, China

²The Smarteye Tech, Shenzhen, China.

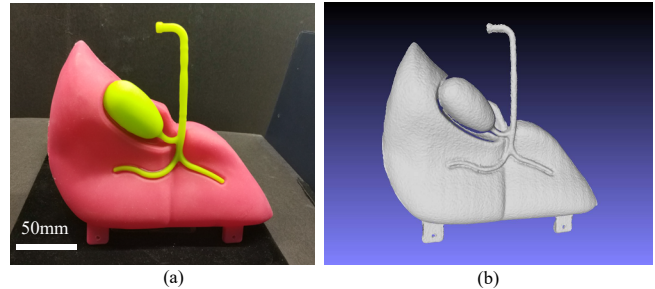


Fig. 1. High-precision 3D surface reconstruction of a liver model. (a) Photo of the liver model. (b) The reconstructed surface using the method proposed in this paper.

assign unique codewords to the pixels that facilitate accurate correlation establishment on homogeneous-texture surfaces. In the single-shot methods, a single pattern is projected on the object, and unique local feature is formed by each pixel and its spatial surroundings. The pattern is specially designed, such as the statistical random speckle used by Kinect [15], De Bruijn sequences [16], and M-arrays [17], etc. The single-shot methods have the advantage of less image acquisition time, which leads to fast 3D surface reconstruction. However, the spatial resolution is restricted by the density of the features in the pattern. Besides, since this technique relies on the surrounding pixels, the reconstruction accuracy is low at the discontinuous surfaces or edges. In addition, this technique is less robust to the noises such as environment light and background textures [13]. As a contrast, in multiple-shot methods, such as gray code [18], and n-ary codes [19], etc., a sequence of patterns is successively projected on the target. The codeword of a pixel is constituted by its illumination values across the sequence. But capturing the multiple patterns severely restricts the reconstruction speed. As one of the most widely used 3D imaging techniques, phase-shifting algorithm can realize unique codification with less pattern requirement [13]. However, phase unwrapping is required to remove ambiguity in the phase-shifting method. For instance, [20] obtained the continuous phase map by analyzing the wrapped phase itself, which is computationally expensive and not suitable for discontinuous surface. Some authors ([21] [22]) retrieved the phase by projecting more patterns, like the gray code and multi-frequency fringes, which was time-consuming. Recently, methods without phase unwrapping [23] [24] have been proposed by introducing multiple cameras, which transfers the complex phase recovery problem into the simple correspondence establishment problem

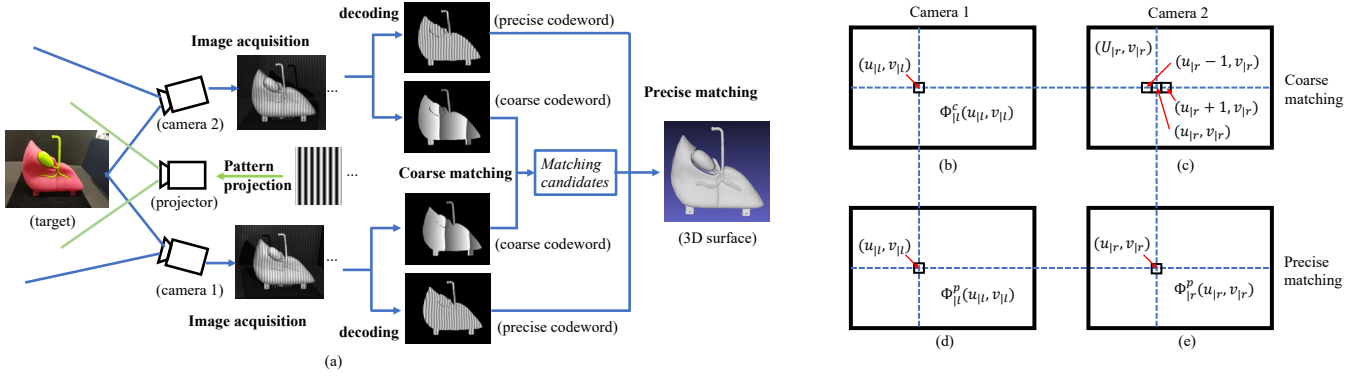


Fig. 2. Principle of the two-step matching method. (a) Schematic of the two-step matching method. (b) A pixel (u_l, v_l) on the left coarse codeword map with a coarse codeword $\phi_{|l}^c(u_l, v_l)$. (c) The corresponding candidates (U_r, v_r) on the right coarse codeword map. (d) A pixel (u_l, v_l) on the left precise codeword map with a precise codeword $\phi_{|l}^p(u_l, v_l)$. (e) The best corresponding point (u_r, v_r) on the right precise codeword map whose precise codeword $\phi_{|r}^p(u_r, v_r)$ is the closest to $\phi_{|l}^p(u_l, v_l)$ among the candidates.

in stereo vision. However, the stereo matching methods they utilized is time-consuming.

To solve the aforementioned problems in 3D surface reconstruction, in this paper, we develop a novel scheme based on active stereo matching. In the scheme, a two-step matching method is proposed to enhance the algorithm efficiency and achieve accurate reconstruction. Dual-frequency sinusoidal phase-shifting fringes are utilized to encode each pixel. The codification method has high encoding accuracy and robustness to noises. Totally five patterns are used to generate the codewords, which is much less than the number of patterns used in the common codification methods [25] [13]. Another advantage of the proposed scheme is that phase unwrapping and projector calibration are not required. Fig. 1 demonstrates the high-precision 3D surface reconstruction of a liver model by using the proposed method.

The remaining part of this paper is organized as follows. Section II describes the principle of the proposed 3D surface reconstruction scheme. In Section III, experiments are conducted to validate the performance of the proposed method. Finally, Section IV gives the conclusion.

II. METHODOLOGY

A. Scheme of The Two-step Matching

Fig. 2(a) shows the schematic of the proposed 3D surface reconstruction method. A sequence of patterns is successively projected onto the target by the projector. The distorted pattern images are captured by the two cameras from different perspectives simultaneously. The images are rectified with the method mentioned in [26] so that each pair of corresponding pixels from two cameras locates in the same epipolar line. Through pattern decoding which is introduced in the following subsection, two codeword maps are obtained from the acquired images. The first codeword is called *coarse codeword*, denoted as ϕ^c , and the second codeword is called *precise codeword*, denoted by ϕ^p , as described in Fig. 2(a). The coarse codeword encodes the whole image uniquely with low encoding accuracy, and the precise codeword encodes

local pixels with high encoding accuracy. For a pixel whose coordinate on the image is (u_k, v_k) , its coarse codeword and precise codeword are represented by $\phi_{|k}^c(u_k, v_k)$ and $\phi_{|k}^p(u_k, v_k)$, respectively. The subscript $k \in \{l, r\}$ indicates that the images are from the left camera or the right camera, respectively. For the sake of convenience, the coordinate (u_k, v_k) is also used to represent the pixel.

The two established codeword maps are then utilized in a two-step matching, i.e. *coarse matching* and *precise matching*. The process is as follows.

First, correspondence is built between left coarse codeword map $\phi_{|l}^c$ and right coarse codeword map $\phi_{|r}^c$. For each pixel (u_l, v_l) on the left codeword map $\phi_{|l}^c$, a pixel (u_r, v_r) from the right camera is regarded as one of the *corresponding candidates* of the pixel (u_l, v_l) if it satisfies $|\phi_{|l}^c(u_l, v_l) - \phi_{|r}^c(u_r, v_r)| < \xi$, where ξ is a user defined small constant. Here we denote the set of all the corresponding candidates as (U_r, v_r) , as shown in Fig. 2(c). Since the coarse codeword encodes the whole image uniquely, at most one group of connected pixels can be found as the corresponding candidates for each pixel from left camera. This step is called *coarse matching*, as shown Fig. 2(b) and (c).

Second, among the correspondence candidates (U_r, v_r) , pixel with the minimum value of $|\phi_{|l}^p(u_l, v_l) - \phi_{|r}^p(u_r, v_r)|$ is regarded as the best corresponding pixel. This step is called *precise matching*, as shown in Fig. 2(d) and (e).

The advantages of the proposed two-step matching are from the following aspects:

- 1) Fast speed and high accuracy: Instead of searching for precise correspondence across the whole image, the proposed method builds rough correspondence quickly, and finds precise correlation in local pixels, which reduces the matching time and achieves high accuracy reconstruction.
- 2) Low algorithm complexity: The matching criterion is realized by simple addition and subtraction, which reduces the algorithm complexity.

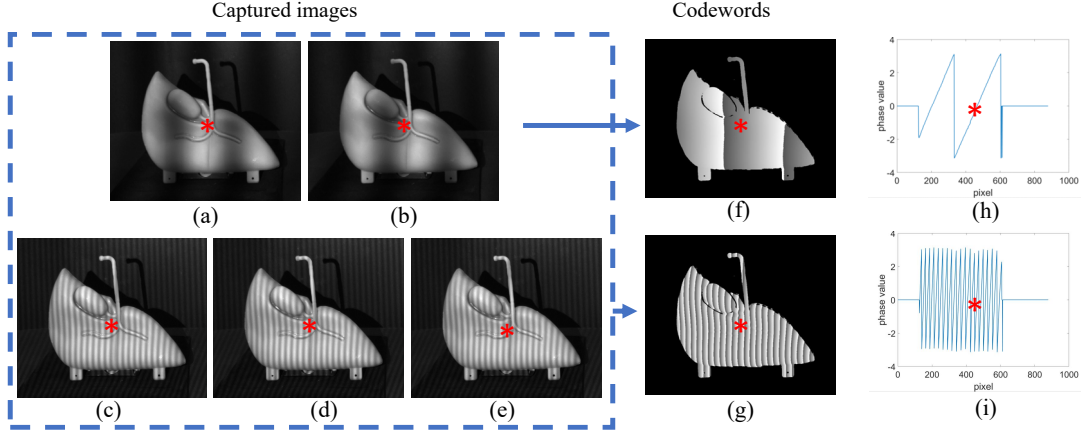


Fig. 3. The captured images and generation of codeword maps. (a)-(e) The acquired images. (f) Coarse codeword map generated from images (a)-(b). (g) Precise codeword map obtained from images (c)-(e). (h) Phases of a row of the codeword map (f). (i) Phases of a row of the codeword map (g). An arbitrary point (marked with red star) has two codewords: the coarse codeword can be obtained from (h), the precise codeword can be obtained from (i).

To further enhance the accuracy, robustness and speed of 3D surface reconstruction, a special codification strategy should be developed. The requirements of the desired codification strategy are listed as follows.

- 1) Unique codeword should be assigned to each pixel.
- 2) The codification strategy should be robust to the background noises such as light and textures.
- 3) The number of patterns should be minimized to reduce the image acquisition time.

Considering the aforementioned requirements, a codification strategy using two-frequency phase-shifting fringes is proposed.

B. Two-phase Codification

Phase-shifting fringes are widely used in pattern codification strategies because merely three patterns can uniquely encode the image. In addition, the phase-shifting fringes have high encoding accuracy and robustness to the background noises [13]. In this work, we utilize phase-shifting patterns of two frequencies ω^c and ω^p ($\omega^p \gg \omega^c$) to generate the coarse codeword and precise codeword, respectively. The projected patterns are described by

$$F_m^c(x, y) = A_0 + B_0 \cos(2\pi\omega^c x + \delta_m^c), (m = 1, 2), \quad (1)$$

$$F_n^p(x, y) = A_0 + B_0 \cos(2\pi\omega^p x + \delta_n^p), (n = 1, 2, 3), \quad (2)$$

where F^c and F^p are the fringes for the coarse codeword and the precise codeword, respectively. $\delta_m^c = m \cdot \pi/2$ and $\delta_n^p = n \cdot 2\pi/3$ are the phase shifts of the two fringe sequences, respectively. (x, y) represents the coordinate of a point on the projected image. A_0 and B_0 are constant numbers.

After projected on the object, the patterns are distorted, and the corresponding captured images can be described by:

$$I_m^c(u, v)|_k = A(u, v)|_k + B^c(u, v)|_k \cos[\phi^c(u, v)|_k + \delta_m^c], \quad (3)$$

$$I_n^p(u, v)|_k = A(u, v)|_k + B^p(u, v)|_k \cos[\phi^p(u, v)|_k + \delta_n^p], \quad (4)$$

where I_m^c ($m = 1, 2$) are images for coarse codification and I_n^p ($n = 1, 2, 3$) are for precise codification, as shown in

Fig. 3(a)-(e), respectively. A is the texture map, which is considered to remain unchanged during the acquisition of five images. B^c and B^p are the intensity modulation. (u, v) is the coordinate of an arbitrary pixel on the captured image. ϕ^c and ϕ^p are phase maps, which serve as the desired coarse codeword and precise codeword, respectively. ϕ^c and ϕ^p can be derived from Eq. (3) and Eq. (4), respectively:

$$\phi^p(u, v)|_k = -\tan^{-1} \frac{\sum_{n=1}^3 I_n^p(u, v)|_k \sin \delta_n}{\sum_{n=1}^3 I_n^p(u, v)|_k \cos \delta_n}. \quad (5)$$

$$\phi^c(u, v)|_k = \tan^{-1} \frac{I_1^c(u, v)|_k - A(u, v)|_k}{I_2^c(u, v)|_k - A(u, v)|_k}. \quad (6)$$

where $A(u, v)$ is given by:

$$A(u, v)|_k = \frac{I_1^p(u, v)|_k + I_2^p(u, v)|_k + I_3^p(u, v)|_k}{3}, \quad (7)$$

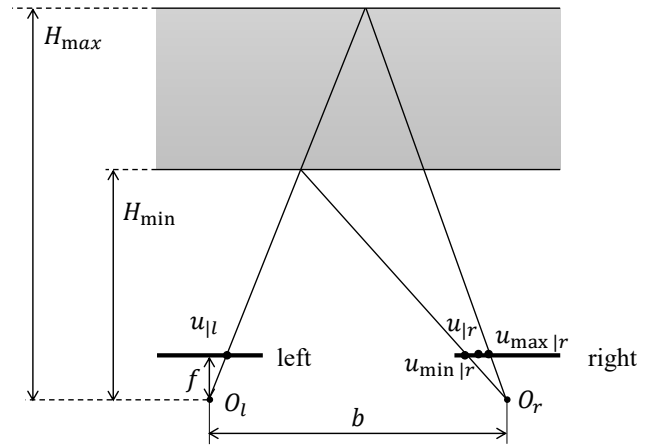


Fig. 4. Geometry constraint of the two-camera system. O_l and O_r are projection centers of the left and right cameras, respectively. The depth range is $[H_{min}, H_{max}]$. If the projection of a point on the left plane is u_l^p , its projection on the right image u_r^p is limited to $[u_{min|r}, u_{max|r}]$.

1) *Coarse Codeword*: The coarse codeword $\phi^c(u, v)$ is designed to encode the whole image. However, since an arctangent function is used in the Eq. (6), the obtained phases have the problem of 2π discontinuity, as shown in Fig.3 (h). The 2π discontinuity will lead to ambiguity in the correspondence establishment. To prevent the ambiguity, we introduce the geometry constraint, as illustrated in the Fig. 4. Suppose the measured range in the depth direction $[H_{min}, H_{max}]$ is known, then for any pixel on the left image whose u coordinate is $u_{|l}$, the u coordinate of its corresponding pixel on the right image will always locate in a closed interval $[u_{min|r}, u_{max|r}]$. The relationship is given by

$$u_{min|r} = u_{|l} - \frac{bf}{H_{min}}, \quad (8)$$

$$u_{max|r} = u_{|l} - \frac{bf}{H_{max}}, \quad (9)$$

where b, f are the baseline between the two cameras and the focal length, respectively. Let the period of the coarse codeword T^c equal to the range $u_{max|r} - u_{min|r} = bf/H_{min} - bf/H_{max}$ which is a fixed value, then the ambiguity can be removed.

2) *Precise Codeword*: The precise codeword is designed to encode the corresponding candidates decided by the coarse codeword. To assign each candidate with unique phase value, the period of the precise phase T^p should be larger than the maximum range of the u coordinates of the candidates, which is $\xi T^c/\pi$. Therefore, it is reasonable to choose the period of precise codeword as $\text{ceil}(\xi T^c/\pi)$.

With the proposed codification method, only five patterns are required, which is much less than the common codification methods [25] [13]. At the same time, the high accuracy is inherited from the high-frequency phase shifting fringes. Moreover, since the precise codeword encodes local pixels, the complex phase unwrapping is avoided. However, one limitation of the codification method is that the method can only reconstruct objects with limited depth change due to the disparity range constraint.

C. Correspondence Refinement

In the previous subsections, we assume the corresponding points always locate exactly on the pixels of the right image. However, since a pixel is a discrete sample of the continuous original image, most of the correspondence points locate

between two pixels. Therefore, we propose a refinement algorithm to estimate the true position of the correspondence point.

With the two-step matching method described in subsection II-A, the best correspondence $(u_{|r}, v_{|r})$ of a pixel $(u_{|l}, v_{|l})$ is obtained. Theoretically the true position of the correspondence point is considered to be located at the neighbourhood of the pixel $(u_{|r}, v_{|r})$, whose phase values can be estimated by the Taylor series shown in Eq. (10). The higher order items are ignored in this paper.

$$\phi_{|r}^p(u) = \phi_{|r}^p(u_{|r}) + \frac{\partial \phi_{|r}^p}{\partial u}(u - u_{|r}) + \frac{1}{2} \frac{\partial^2 \phi_{|r}^p}{\partial u^2}(u - u_{|r})^2, \quad (10)$$

where

$$\frac{\partial \phi_{|r}^p}{\partial u} = \phi_{|r}^p(u_{|r}, v_{|r}) - \phi_{|r}^p(u_{|r} - 1, v_{|r}), \quad (11)$$

$$\frac{\partial^2 \phi_{|r}^p}{\partial u^2} = \phi_{|r}^p(u_{|r} - 1, v_{|r}) - 2\phi_{|r}^p(u_{|r}, v_{|r}) + \phi_{|r}^p(u_{|r} + 1, v_{|r}), \quad (12)$$

Let Eq. (10) equal to the $\phi_{|l}^p(u_{|l}, v_{|l})$, the estimated true correspondence can be obtained by solving for the quadratic equation. Although there are two solutions for the quadratic equation, only one locates within the neighbours of $(u_{|r}, v_{|r})$. Therefore, unique correspondence can be found in the refinement.

III. EXPERIMENTS

To validate the effectiveness and the performance of the proposed 3D surface reconstruction method, various experiments are conducted on the system shown in Fig. 6. The system consists of a projector (DLP Lightcrafter 4500), two CMOS cameras (MER-502-79U3M, monochrome) and a computer. The cameras are synchronized with the projector by signal triggering. The resolutions of the projector and the cameras are 1024×768 and 2448×2048 , respectively. The two cameras are calibrated using the method reported in [27]. The length of baseline between the two cameras is 262.2773 mm. The objects are placed 500 – 1000 mm away from the system. The projector is not calibrated. The 3D surface reconstruction scheme is programmed by C++ and implemented on a desktop computer (CPU: Intel core i7-8700, 16 GB RAM, GPU: GeForce GTX 1080 Ti). The generation of point clouds and surface is implemented with the Point Cloud Library [28].

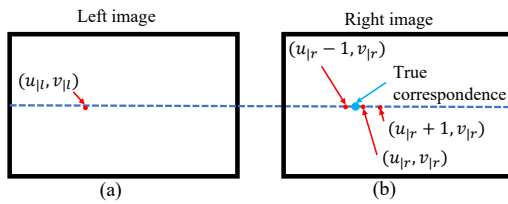


Fig. 5. Illustration of true position of the corresponding point. (a) A pixel $(u_{|l}, v_{|l})$ on the left image, (b) the correspondence pixel $(u_{|r}, v_{|r})$ on the right image obtained by the method introduced in subsection II-A and II-B, the true corresponding point can be estimated by correspondence refinement.

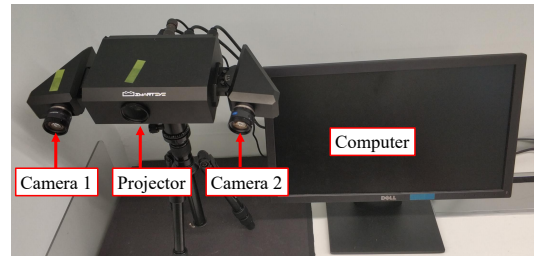


Fig. 6. The 3D vision system utilized in this paper.

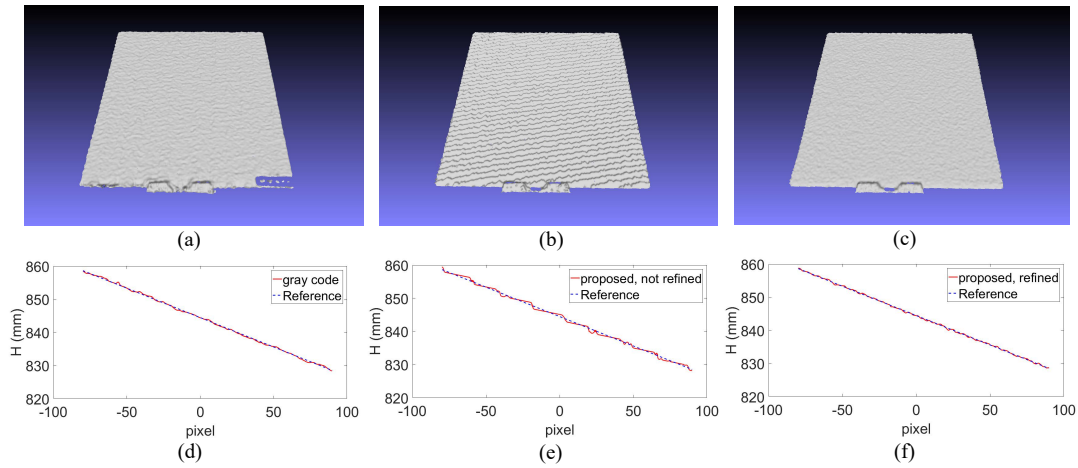


Fig. 7. Results of experiments conducted on a plate. (a)(b)(c) The reconstructed surface using gray code method, the proposed method without correspondence refinement, and the proposed method with correspondence refinement, respectively. (d)(e)(f) Cross-sections of the three surface and the fitted lines, respectively.

A. Precision and Accuracy tests

To validate the precision and accuracy, we conduct experiments on two standard objects, namely a plate and a cylinder.

1) *Plate*: First, we verify the precision and accuracy of the proposed scheme by conducting experiments on a plate made of glasses, which is considered as a standard plane surface. The size of the plate is 188.04 mm \times 188.04 mm, measured by a vernier caliper.

In the precision test, comparative experiments are conducted using different methods: the gray code [25] which is one of the most accurate 3D surface reconstruction methods [13], the proposed two-step matching scheme without correspondence refinement, and the proposed scheme with correspondence refinement. The reconstructed surfaces are illustrated in Fig. 7(a)-(c) and the supporting video. The numbers of reconstructed points are 159870, 163743, and 163730, respectively. For the method without refinement, the reconstructed result presents step-like surface. Because the correlation points are considered to locate exactly on the pixels of the image, the disparities are discrete integers, and the resultant depth are discrete values. Cross-sections of the three surface are shown in Fig. 7(d)-(f), whose root-mean-square errors (RMSEs) are 0.0128 mm, 0.0229 mm and 0.0078 mm, respectively. The results show that through the correspondence refinement, the reconstruction precision has been largely improved. Besides, the proposed scheme can achieve better performance in precision compared with the gray code method.

TABLE I
RESULTS OF THE ACCURACY TESTS

	Plate		Cylinder
	Height (mm)	Width (mm)	Diameter (mm)
True size	188.04	188.04	71.93
Measured size	187.2855	187.7280	71.8082
Relative Error	0.401%	0.166%	0.169%

In the accuracy test, the size of the 3D surface reconstructed by the proposed scheme is measured five times. The average of the five measurements are listed in the Table. I. The height and width of the plate surface are 187.02855 mm and 187.7280 mm, respectively. The relative errors are 0.401% and 0.166%, respectively. The results show that the accuracy of 3D surface reconstruction is sufficient for common robotic applications.

2) *Cylinder*: We also conduct experiments on a cylindrical bottle, as shown in Fig. 8. The diameter of the cylinder is 71.93mm. The reconstructed surface of the bottle is shown in Fig. 8(b). 94123 points are obtained. A cross-section of the reconstructed surface is optimally fitted to a circle, as shown in Fig. 8(c). The diameter of the fitted circle is 71.8082, whose relative error is 0.169%. The RMSE of the points is 0.1285 mm. These results also prove that the propose scheme can obtain precise and accurate 3D surface reconstruction results for objects with curved surface.

B. Reconstruction of Complex Objects

We further carry out more experiments to validate the performance of the proposed scheme in reconstructing complex objects with complicated curved surface, holes, thin lines and occlusions. The objects used in the experiments

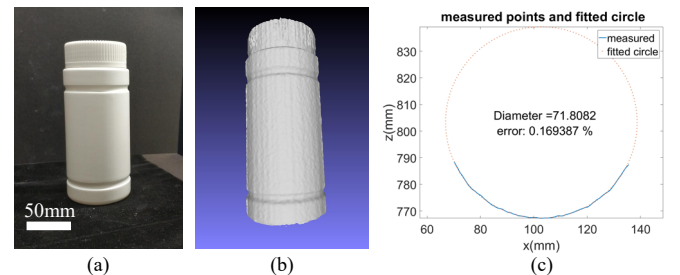


Fig. 8. Results of the experiment conducted on a bottle. (a) Photo of the bottle. (b) The reconstructed surface. (c) A cross-section of the surface and the fitted circle.

are a card, a bag, a head model and a liver model as shown in Fig. 9(a),(c),(e) and Fig. 1(a), respectively. The 3D surfaces obtained by the proposed scheme are demonstrated in Fig. 9(b),(d),(f) and Fig. 1(b), respectively. The numbers of obtained points of the four objects are 96387, 385334, 844172, and 164634, respectively. The liver model (Fig. 1(a)) and the bag (Fig. 9(c)) verify the capability of the proposed scheme in reconstructing the complicated curved surface. The letters on the card (Fig. 9(a)) has a height and minimum width of 1 mm. The resultant surface demonstrates that the proposed scheme can achieve high-resolution reconstruction. Besides, the holes on the card and liver model are well reconstructed. The head model shown in Fig. 9(e) are used to validate the ability of the scheme to avoid occlusion. The 3D surfaces reconstructed from left camera and right camera are shown in Fig. 9(g)(h) and Fig. 9(i)(j), respectively. Some areas such as the right temple and the nose cannot be reconstructed from both cameras. By combining the point clouds, the entire surface without occlusion can be obtained, as shown in Fig. 9 (f).

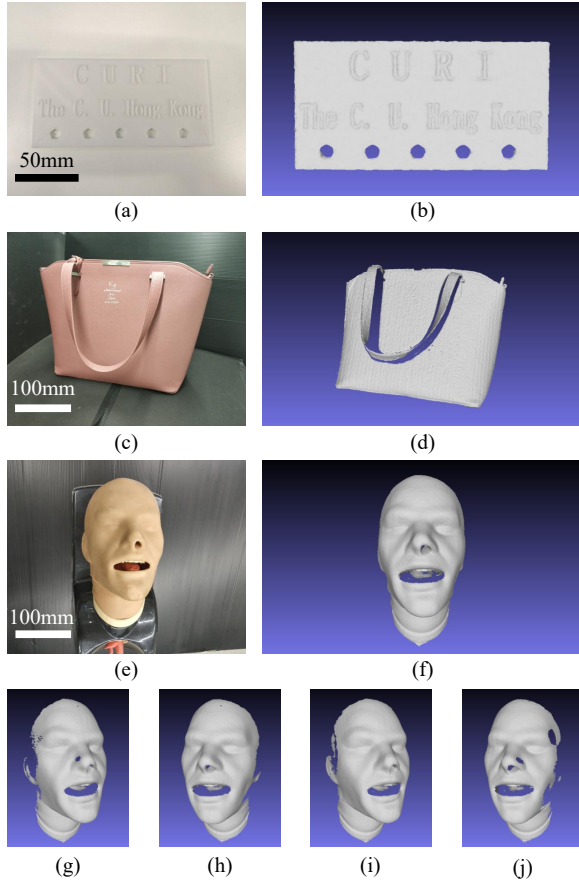


Fig. 9. Demonstration of 3D surface reconstruction of complex objects. (a) Photo of a card. (b) The reconstructed surface of the card-(a). (c) Photo of a bag. (d) The reconstructed surface of the bag-(c). (e) Photo of a head model. (f) The reconstructed surface of the head model-(e). (g)(h) The 3D surface of head model-(e) reconstructed from right camera. (i)(j) The 3D surface of head model-(e) reconstructed from left camera.

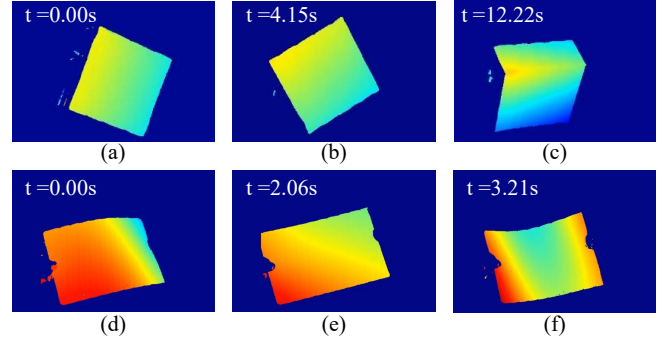


Fig. 10. Real-time experiment results. (a)-(c) Disparity maps of a moving box. (d)-(f) Disparity maps showing the deformation of paper.

C. Real-time Experiments

To demonstrate the real-time performance, two experiments are conducted using the proposed scheme. We first carry out experiments on a rotating box, whose sequences of disparity map are shown in Fig. 10(a)-(c). The colors on disparity map represent the depth information. When a point gets closer to the camera, its color varies from blue to red correspondingly. For the second real-time experiment shown in Fig. 10(d)-(f), a piece of paper is tested to verify the ability of the proposed scheme to reconstruct the deformation of soft objects. The average time for stereo matching is only 3.44 ms, which verifies the efficiency of the proposed matching method. The time for the entire reconstruction procedure is 140.56 ms by average, where the image acquisition time (average 137.12 ms) occupy the majority of the time. Please see the supporting video for more details.

IV. CONCLUSIONS

In this paper, we presented a 3D surface reconstruction scheme. A two-step matching method was proposed which has advantages of high reconstruction accuracy and high algorithm efficiency. Moreover, we proposed a pattern codification method which utilized five patterns to realize high encoding accuracy and robustness to noises. In addition, a correspondence refinement method was proposed to further enhance the matching precision. Various experiments were carried out to validate the effectiveness and performance of the proposed scheme. Experiments conducted on standard objects showed that the relative error of the reconstructed surface is less than 0.5%, and the RMSE is less than 0.2mm. Experiments on the complex objects validated the method has the capability to reconstruct complicated objects, and achieve high-resolution reconstruction and occlusion avoidance. Besides, real-time performance of the proposed method was also demonstrated.

ACKNOWLEDGMENT

The authors would like to thank Mr. Lidong Yang, Mr. Yong Peng, Dr. Guo Huiwen, Mr. Wang Daochuan and Dr. Suming Tang for their valuable advice.

REFERENCES

- [1] Z. Xiong, Y. Zhang, F. Wu, and W. Zeng, "Computational depth sensing : Toward high-performance commodity depth cameras," *IEEE Signal Processing Magazine*, vol. 34, no. 3, pp. 55–68, May 2017.
- [2] S. Chen, Y. Li, and N. M. Kwok, "Active vision in robotic systems: A survey of recent developments," *International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.
- [3] X. Li, X. Su, and Y. Liu, "Adaptive region control for robotic soldering of flexible pcbs," in *2017 18th International Conference on Advanced Robotics (ICAR)*, July 2017, pp. 216–221.
- [4] Y. Zhu, R. Mottaghi, E. Kolve, J. J. Lim, A. Gupta, L. Fei-Fei, and A. Farhadi, "Target-driven visual navigation in indoor scenes using deep reinforcement learning," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 3357–3364.
- [5] Y. Huang, S. Huang, H. Chen, Y. Chen, C. Liu, and T. S. Li, "A 3d vision based object grasping posture learning system for home service robots," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2017, pp. 2690–2695.
- [6] Z. Wang, X. Li, D. Navarro-Alarcon, and Y.-h. Liu, "A unified controller for region-reaching and deforming of soft objects," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 472–478.
- [7] D. Navarro-Alarcon, H. M. Yip, Z. Wang, Y. Liu, F. Zhong, T. Zhang, and P. Li, "Automatic 3-d manipulation of soft objects by robotic arms with an adaptive deformation model," *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 429–441, April 2016.
- [8] F. Alambeigi, Z. Wang, R. Hegeman, Y.-H. Liu, and M. Armand, "A robust data-driven approach for online learning and manipulation of unmodeled 3-d heterogeneous compliant objects," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4140–4147, 2018.
- [9] F. Zhong, D. Navarro-Alarcon, Z. Wang, Y. Liu, T. Zhang, H. M. Yip, and H. Wang, "Adaptive 3d pose computation of suturing needle using constraints from static monocular image feedback," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, 2016, pp. 5521–5526.
- [10] L. Qian, A. Deguet, and P. Kazanzides, "Arssist: augmented reality on a head-mounted display for the first assistant in robotic surgery," *Healthcare Technology Letters*, vol. 5, no. 5, pp. 194–200, 2018.
- [11] Z. Wang, Z. Liu, Q. Ma, A. Cheng, Y.-h. Liu, S. Kim, A. Deguet, A. Reiter, P. Kazanzides, and R. H. Taylor, "Vision-based calibration of dual rcm-based robot arms in human-robot collaborative minimally invasive surgery," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 672–679, 2018.
- [12] C. Sui, Z. Wang, and Y. Liu, "A 3d laparoscopic imaging system based on stereo-photogrammetry with random patterns," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018, pp. 1276–1282.
- [13] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern recognition*, vol. 43, no. 8, pp. 2666–2680, 2010.
- [14] B. Tippetts, D. J. Lee, K. Lillywhite, and J. Archibald, "Review of stereo vision algorithms and their suitability for resource-limited systems," *Journal of Real-Time Image Processing*, vol. 11, no. 1, pp. 5–25, 2016.
- [15] J. Smisek, M. Jancosek, and T. Pajdla, "3d with kinect," in *Consumer depth cameras for computer vision*. Springer, 2013, pp. 3–25.
- [16] A. Reiter, A. Sigaras, D. Fowler, and P. K. Allen, "Surgical structured light for 3d minimally invasive surgical imaging," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2014, pp. 1282–1287.
- [17] C. Albitar, P. Graebbling, and C. Doignon, "Design of a monochromatic pattern for a robust structured light coding," in *2007 IEEE International Conference on Image Processing*, vol. 6, Sept 2007, pp. VI – 529–VI – 532.
- [18] Z. Song, R. Chung, and X. Zhang, "An accurate and robust strip-edge-based structured light means for shiny surface micromasurement in 3-d," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 3, pp. 1023–1032, March 2013.
- [19] X. Chen, S. Chen, J. Luo, M. Ma, Y. Wang, Y. Wang, and L. Chen, "Modified gray-level coding method for absolute phase retrieval," *Sensors*, vol. 17, no. 10, p. 2383, 2017.
- [20] Z. Dai and X. Zha, "An accurate phase unwrapping algorithm based on reliability sorting and residue mask," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 2, pp. 219–223, March 2012.
- [21] D. Zheng, F. Da, Q. Kemao, and H. S. Seah, "Phase-shifting profilometry combined with gray-code patterns projection: unwrapping error removal by an adaptive median filter," *Optics express*, vol. 25, no. 5, pp. 4700–4713, 2017.
- [22] J.-S. Hyun and S. Zhang, "Superfast 3d absolute shape measurement using five binary patterns," *Optics and Lasers in Engineering*, vol. 90, pp. 217–224, 2017.
- [23] W. Lohry, V. Chen, and S. Zhang, "Absolute three-dimensional shape measurement using coded fringe patterns without phase unwrapping or projector calibration," *Opt. Express.*, vol. 22, no. 2, pp. 1287–1301, 2014.
- [24] K. Zhong, Z. Li, Y. Shi, C. Wang, and Y. Lei, "Fast phase measurement profilometry for arbitrary shape objects without phase unwrapping," *Opt. Laser. Eng.*, vol. 51, no. 11, pp. 1213–1222, 2013.
- [25] S. Inokuchi, "Range imaging system for 3-d object recognition," *ICPR*, 1984, pp. 806–808, 1984.
- [26] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc., 2008.
- [27] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, Nov 2000.
- [28] D. Holz, A. E. Ichim, F. Tombari, R. B. Rusu, and S. Behnke, "Registration with the point cloud library: A modular framework for aligning in 3-d," *IEEE Robotics & Automation Magazine*, vol. 22, no. 4, pp. 110–124, 2015.