

# Motion-oriented high speed 3-D measurements by binocular fringe projection using binary aperiodic patterns

SHIJIE FENG,<sup>1,2,3,4</sup> QIAN CHEN,<sup>2,5</sup> CHAO ZUO,<sup>1,2,\*</sup> TIANYANG TAO,<sup>1,2</sup> YAN HU,<sup>1,2</sup> AND ANAND ASUNDI<sup>3</sup>

<sup>1</sup>Smart Computational Imaging (SCI) Laboratory, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

<sup>2</sup>Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

<sup>3</sup>Centre for Optical and Laser Engineering, School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore

<sup>4</sup>geniushijie@163.com

<sup>5</sup>chenqian@njjust.edu.cn

\*surpasszuo@163.com

**Abstract:** Fringe projection is an extensively used technique for high speed three-dimensional (3-D) measurements of dynamic objects. To precisely retrieve a moving object at pixel level, researchers prefer to project a sequence of fringe images onto its surface. However, the motion often leads to artifacts in reconstructions due to the sequential recording of the set of patterns. In order to reduce the adverse impact of the movement, we present a novel high speed 3-D scanning technique combining the fringe projection and stereo. Firstly, promising measuring speed is achieved by modifying the traditional aperiodic sinusoidal patterns so that the fringe images can be cast at kilohertz with the widely used defocusing strategy. Next, a temporal intensity tracing algorithm is developed to further alleviate the influence of motion by accurately tracing the ideal intensity for stereo matching. Then, a combined cost measure is suggested to robustly estimate the cost for each pixel and lastly a three-step framework of refinement follows not only to eliminate outliers caused by the motion but also to obtain sub-pixel disparity results for 3-D reconstructions. In comparison with the traditional method where the effect of motion is not considered, experimental results show that the reconstruction accuracy for dynamic objects can be improved by an order of magnitude with the proposed method.

© 2017 Optical Society of America

**OCIS codes:** (120.0120) Instrumentation, measurement, and metrology; (150.6910) Three-dimensional sensing; (110.6880) Three-dimensional image acquisition.

## References and links

1. T. Tao, Q. Chen, J. Da, S. Feng, Y. Hu, and C. Zuo, "Real-time 3-d shape measurement with composite phase-shifting fringes and multi-view system," *Opt. Express* **24**, 20253–20269 (2016).
2. X. Yin, H. Zhao, J. Zeng, and Y. Qu, "Acoustic grating fringe projector for high-speed and high-precision three-dimensional shape measurements," *Appl. Opt.* **46**, 3046–3051 (2007).
3. S. Feng, Q. Chen, C. Zuo, and A. Asundi, "Fast three-dimensional measurements for dynamic scenes with shiny surfaces," *Opt. Commun.* **382**, 18–27 (2017).
4. E. B. Li, X. Peng, J. Xi, J. F. Chicharo, J. Q. Yao, and D. W. Zhang, "Multi-frequency and multiple phase-shift sinusoidal fringe projection for 3d profilometry," *Opt. Express* **13**, 1561–1569 (2005).
5. S. Van der Jeught and J. J. J. Dirckx, "Real-time structured light profilometry: a review," *Opt. Lasers Eng.* **87**, 18–31 (2016).
6. T. Monks and J. N. Carter, "Improved stripe matching for colour encoded structured light," in *Proceedings International Conference on Computer Analysis of Images and Patterns* (Springer), pp. 476–485.
7. Q. Zhang and X. Su, "High-speed optical measurement for the drumhead vibration," *Opt. Express* **13**, 3110–3116 (2005).
8. C. Guan, L. G. Hassebrook, and D. L. Lau, "Composite structured light pattern for three-dimensional video," *Opt. Express* **11**, 406–417 (2003).

9. J. Salvi, J. Pagès, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recogn.* **37**, 827–849 (2004).
10. C. Zuo, L. Huang, M. Zhang, Q. Chen, and A. Asundi, "Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review," *Opt. Lasers Eng.* **85**, 84–103 (2016).
11. Y. Wang and S. Zhang, "Superfast multifrequency phase-shifting technique with optimal pulse width modulation," *Opt. Express* **19**, 5149–5155 (2011).
12. C. Zuo, Q. Chen, G. Gu, S. Feng, and F. Feng, "High-speed three-dimensional profilometry for multiple objects with complex shapes," *Opt. Express* **20**, 19493–19510 (2012).
13. C. Zuo, Q. Chen, G. Gu, S. Feng, F. Feng, R. Li, and G. Shen, "High-speed three-dimensional shape measurement for dynamic scenes using bi-frequency tripolar pulse-width-modulation fringe projection," *Opt. Lasers Eng.* **51**, 953–960 (2013).
14. Z. Zhang, C. E. Towers, and D. P. Towers, "Time efficient color fringe projection system for 3d shape and color using optimum 3-frequency selection," *Opt. Express* **14**, 6444–6455 (2006).
15. M. Schaffer, M. Grosse, and R. Kowarschik, "High-speed pattern projection for three-dimensional shape measurement using laser speckles," *Appl. Opt.* **49**, 3622–3629 (2010).
16. T. Weise, B. Leibe, and L. V. Gool, "Fast 3d scanning with automatic motion compensation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2007), pp. 1–8.
17. P. Cong, Z. Xiong, Y. Zhang, S. Zhao, and F. Wu, "Accurate dynamic 3d sensing with fourier-assisted phase shifting," *IEEE J. Sel. Topics Signal Process.* **9**, 396–408 (2015).
18. O. Hall-Holt and S. Rusinkiewicz, "Stripe boundary codes for real-time structured-light range scanning of moving objects," in *Proceedings of IEEE International Conference on Computer Vision* (IEEE 2001), pp. 359–366 vol.2.
19. Z. Li, B. Curless, and S. M. Seitz, "Spacetime stereo: shape recovery for dynamic scenes," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2003), pp. 367–374.
20. B. Harendt, M. Große, M. Schaffer, and R. Kowarschik, "3d shape measurement of static and moving objects with adaptive spatiotemporal correlation," *Appl. Opt.* **53**, 7507–7515 (2014).
21. J. Davis, R. Ramamoorthi, and S. Rusinkiewicz, "Spacetime stereo: a unifying framework for depth from triangulation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2003), pp. 359–366.
22. M. Schaffer, M. Große, B. Harendt, and R. Kowarschik, "High-speed optical 3-d measurements for shape representation," *Opt. Photonics News* **22**, 49 (2011).
23. S. Heist, A. Mann, P. Kühmstedt, P. Schreiber, and G. Notni, "Array projection of aperiodic sinusoidal fringes for high-speed three-dimensional shape measurement," *Opt. Eng.* **53**, 112208 (2014).
24. S. Heist, P. Lutzke, I. Schmidt, P. Dietrich, P. Kühmstedt, A. Tünnermann, and G. Notni, "High-speed three-dimensional shape measurement using gobo projection," *Opt. Lasers Eng.* **87**, 90–96 (2016).
25. S. Zhang, D. Van Der Weide, and J. Oliver, "Superfast phase-shifting method for 3-d shape measurement," *Opt. Express* **18**, 9684–9689 (2010).
26. Y. Zhan, Y. Gu, K. Huang, C. Zhang, and K. Hu, "Accurate image-guided stereo matching with efficient matching cost and disparity refinement," *IEEE Trans. Circuits Syst. Video Technol.* **26**, 1632–1645 (2016).
27. S. Feng, Q. Chen, and C. Zuo, "Graphics processing unit-assisted real-time three-dimensional measurement using speckle-embedded fringe," *Appl. Opt.* **54**, 6865–6873 (2015).
28. X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and Z. Xiaopeng, "On building an accurate stereo matching system on graphics hardware," in *Proceedings of IEEE Computer Vision Workshops (ICCV Workshops)* (IEEE, 2011), pp. 467–474.
29. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision* **47**, 7–42 (2002).
30. Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 492–504 (2009).
31. A. N. Stein, A. Huertas, and L. Matthies, "Attenuating stereo pixel-locking via affine window adaptation," in *Proceedings of IEEE Conference on Robotics and Automation* (ICRA, 2006), pp. 914–921.
32. J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vision* **12**, 43–77 (1994).
33. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University, 2003).
34. J.-Y. Bouguet, "Camera calibration toolbox for matlab," (2004).
35. Q. Zhang, X. Su, L. Xiang, and X. Sun, "3-d shape measurement based on complementary gray-code light," *Opt. Lasers Eng.* **50**, 574–579 (2012).

## 1. Introduction

With rapid developments of the digital technology, high speed three-dimensional (3-D) measurements have received extensive attentions in recent years [1–4]. With the ability of measuring dynamic objects, they are showing great potentials for various fields including on-line inspec-

tions, fast reverse engineering, medical sciences and home entertainments. Benefited from merits of high measuring precision, full-field inspection and low requirements for surface features, the fringe projection technique becomes one of the most extensively used strategies for 3-D measurements [5].

To implement high speed 3-D measurements, intuitively, one should project as fewer patterns as possible, which can make the coding process of the moving surface less sensitive to the motion. Thus, various methods based on casting a unique pattern were developed. Monks et al. [6] used a De Bruijn sequence to generate a pattern with horizontal colored stripes where six colors were employed to paint the fringes. The coloring of the fringes was selected so that each sub-sequence of three colors appeared only once. Zhang and Su [7] measured a vibrating Chinese drum with the Fourier transform profilometry (FTP) by which a single sinusoidal fringe pattern was projected onto the drumhead and Fourier transform then followed to extract the profile from the captured pattern. Besides, Guan et al. [8] suggested to combine multiple phase shifting patterns into a single pattern by modulating each individual pattern along orthogonal direction with a distinct carrier frequency and then putting them together. Although these encoding techniques allow retrievals of dynamic contours with the minimum number of image projection, it is difficult for them to provide results with high spatial resolution since the codification must be condensed in a unique pattern [9].

Owing to the increasing rate of image projection and capture, researchers begin to resort to the coding strategy of time-multiplexing since a moving process can be assumed as a quasi-static one with the raised measuring speed. By the idea of time-multiplexing, a set of patterns are projected successively onto a measured object and the code word of a given pixel can be constructed from the sequence of the captured intensities. Therefore, with the strategy 3-D reconstructions can be obtained with higher resolution and accuracy compared with the methods using a unique pattern. Phase shifting profilometry (PSP) is one of the most widely used techniques based on this temporal coding strategy [10]. Wang and Zhang [11] presented a high speed multi-frequency phase-shifting technique where nine patterns were projected to recover a dynamic scene. Then, to project fewer patterns, Zuo et al. firstly proposed a four-frame pattern projection strategy to measure dynamic isolated objects in real time [12] and then developed a bi-frequency phase shifting algorithm based on tri-polar pulse width modulation with which one kilohertz measuring speed was reached [13]. In addition, Zhang et al. [14] reported a color fringe projection technique with which fringe images were encoded into different color channels of a same image for the increase of the projection efficiency. Except to use the phase information as a cue for surface reconstruction, some researchers utilize random statistical patterns to label a pixel where the codification is performed based on the uniqueness of captured intensities over time. Schaffer et al. [15] proposed to project randomly distributed laser speckles to encode a surface. Then a temporal correlation was used to determine the corresponding pixels.

Although for mentioned time-multiplexing methods the sensitivity towards movements has been reduced with the raising measuring speed, the impact of the motion cannot be simply ignored. The reason is that the 3-D reconstruction is extended over multiple frames when a pixel is temporally encoded. In order to reduce the motion artifacts, approaches based on motion compensation and intensity tracing have been presented respectively. To compensate the effect of motions, Weise et al. [16] reported a closed-form expression for the motion error of the phase obtained by the three-step phase shifting method. With the same idea, Cong et al. [17] applied the Fourier transform profilometry to estimate the phase distortion between successive phase shifting fringes for the compensation of the final phase result. Although these methods perform well in handling the motion artifacts, they are developed to correct erroneous phase values, therefore would be inapplicable for techniques that do not use phase information to mark a pixel. For these techniques, methods based on a tracing mechanism are more favorable. Hall-Holt and Rusinkiewicz [18] proposed a technique that projected boundary-coded stripe patterns over time.

By finding the nearest stripes over time, a unique code word can be detected for every stripes in spite of the movements. However, it has limitations that the measured object should move slowly to ensure correct temporal correlations, and only surfaces at the stripe boundaries can be perceived. Zhang et al. [19] reported a space-time stereo technique by which intensities captured by two cameras were matched with a temporal and spatial varying window. Since the sum of squared difference is the only measure used for the cost measurement, unreliable matches may be obtained due to the random noise or the variation of ambient illuminations. Similarly, Harendt et al. [20] developed an adaptive spatiotemporal correlation method where the inter-frame motions were compensated by the difference of disparity maps obtained over time. But, as a window-based spatial correlation is required to acquire initial disparity maps frame-wisely, the spatial resolution of reconstructed dynamic surfaces tends to be limited. Additionally, Davis et al. [21] also presented a spatial and temporal framework to trace the changing intensities. But their work was focusing in analyzing and presenting results for geometrically motionless objects, thus may find difficulties when apply to dynamic scenes.

In this work, we present a novel high speed 3-D shape measuring technique that combines the stereo vision and fringe projection for accurate and robust reconstructions. A few of aperiodic sinusoidal patterns are employed with the purpose of encoding each pixel with the strategy of time-multiplexing. To reduce the sensitivity to motions, the conventional aperiodic sinusoidal fringes have been changed into binary ones which can be projected at kilohertz by a digital light processing (DLP) projector. Since movements will cause artifacts in the reconstructions, we introduce an intensity-tracing algorithm by which the motion or deformation can be simulated adaptively within a motion-oriented temporal support window. For the right camera, this facilitates the search of desired intensities corresponding to the ones captured by the left camera, thus reducing the mismatches due to the motion. Since conventional cost measurement is conducted with a single cost measure, different measures may lead to various results due to their own strengths and weaknesses. To achieve better performance, we integrate the measures of absolute difference of image intensity, absolute difference of image gradient, normalized cross correlation and census transform into a synthetic measure to make use of the combined advantages. Further, we develop a three-step refinement framework to remove outliers among initial disparity results and to obtain precise sub-pixel disparity maps for 3-D reconstructions. Experimental results show that dynamic scenes can be accurately reconstructed by the proposed motion-oriented measuring technique.

## 2. Principles

### 2.1. Pattern design: binary aperiodic sinusoidal fringes

In PSP or FTP, sinusoidal fringes are usually projected onto measured objects to encode the surface with phase information. However, the gamma distortion in projectors will cause the deviation of captured fringe patterns from ideal sinusoidal waveforms, which may introduce errors in obtained phase maps. Additionally, sinusoidal patterns are often represented by pixels of 8-bit depth so that the projection rate is generally limited to hundreds of hertz (normally <255Hz) [22]. Compared with the capturing rate up to tens of thousands of Hz offered by cameras, we can find that the projection bottleneck imposes restrictions on the system's measuring speed.

To circumvent the issues, we project the aperiodic sinusoidal patterns and retrieve the surface by stereo matching [23]. Owing to the stereo imaging, the intensity deviation resulted from the gamma distortion will not negatively affect the reconstructions. Because as long as the same intensity can be captured by the two cameras, the stereo matching can be performed correctly no matter whether the captured intensity is deviated or not. For traditional aperiodic sinusoidal fringes, their intensities can be expressed as

$$I_i(x, y) = a_i(x, y) + b_i(x, y) \cdot \sin [c_i(x, y)x + e_i(x, y)] \geq 0 \quad (1)$$

where  $i$  is the frame index,  $a(x, y)$  the DC component,  $b(x, y)$  the amplitude of the sine wave,  $c(x, y)$  the angular frequency and  $e(x, y)$  the phase shift. Since these parameters can vary spatially and temporally in a random way, the obtained pattern will no longer be a standard periodic sinusoidal image. These aperiodic sinusoidal patterns can be projected at a high speed with an array projector [23] or a GOBO wheel [24]. As it may not be convenient to build these projection devices, we suggest to modify the traditional aperiodic sinusoidal patterns so that they can be projected at kilohertz but with a DLP projector. For the modification, the original gray fringes as shown by Eq. (1) are changed into a few binary stripes with different periods. For each pair of white and black stripes, the intensity can be represented as

$$S^j(x, y) = \begin{cases} 1 & 0 \leq x < h^j \\ 0 & h^j \leq x < g^j \end{cases} \quad (2)$$

where  $j$  is an index of the pair of 1-0 stripes,  $h$  the width of the contained white stripe and  $g$  the total period of the pair. The  $h$  and  $g$  can be randomly selected to generate basic binary stripes for the construction of binary  $I_i(x, y)$  by

$$I_i'(x, y) = U[S_i^1, S_i^2, S_i^3, \dots, S_i^j] \quad (3)$$

where  $U(\cdot)$  means to cascade the pairs of 1-0 stripes spatially. In the measurement, the generated aperiodic fringes are then projected onto the measured surface sequentially with the purpose of encoding the object with temporally varying artificial texture. When the binary patterns are projected by a defocused DLP projector, they will be captured as aperiodic sinusoidal fringes [25]. With the increased projection rate, the sensitivity to motions can be reduced accordingly.

## 2.2. Combined matching cost measurement

For a binocular fringe projection system, the matching cost of corresponding pixels is estimated by a single measure traditionally [15, 19, 24]. The unique cost measure may not be reliable enough when the captured intensity is affected by random noise, camera gain and camera bias [26]. Thus, to improve the robustness of the cost measurement, we employ the strategy of a combined cost measurement to leverage strengths of different cost measures. To this end, we develop a novel synthetic cost measure which consists of absolute difference of image intensity, absolute difference of image gradient, normalized cross correlation and census transform.

Before introducing the details, we firstly define some terminology and variables to facilitate the understanding of the procedure of our cost measurement. We have two primary camera views, *Left* and *Right*, between which correspondences are to be established. We define that pixel  $p = (x, y)$  is in the left camera view and pixel  $p_d = (x - d, y)$  the corresponding point in the right image with a disparity level  $d$ . It is noted that the images of both cameras need to be rectified so that the corresponding point  $p_d$  can be searched horizontally. The basic cost measurements between  $p$  and  $p_d$  are respectively denoted as  $C_{AD}(p, d)$ ,  $C_{gradAD}(p, d)$ ,  $C_{NCC}(p, d)$  and  $C_{census}(p, d)$  for the measures of absolute difference of image intensity, absolute difference of gradient, normalized cross correlation and census transform.

Firstly, the absolute difference of image intensity is a commonly adopted cost measure, as it indicates an intuitive difference between two corresponding pixels. With this measure, the matching ambiguity for regions with repetitive textures can be reduced. In addition, the calculation of the intensity difference is easy to implement, thus it can be obtained efficiently. Given the captured fringe images, the measure of the absolute difference between the image intensities is defined as

$$C_{AD}(p, d) = |I^{Left}(p) - I^{Right}(p_d)| \quad (4)$$

Then, the image gradient contains important structured information of captured patterns and shows a good tolerance to the ambient illumination. Therefore, it is introduced into the work to

calculate the matching cost by

$$C_{gradAD}(p, d) = \left| I_{grad}^{Left}(p) - I_{grad}^{Right}(pd) \right| \quad (5)$$

Thirdly, the normalized cross correlation has found various applications in computer vision. It is one of the simplest but effective measures for the estimation of similarity because of its invariability to linear variations of brightness and contrast. Moreover, it can be implemented by parallel computations, which shows favorable performance for real-time measurements [27]. To calculate the cost by normalized cross correlation, we have

$$C_{NCC}(p, d) = 1 - \left| \frac{\sum_t [Q_{Left}(t) - \bar{Q}_{Left}] [Q_{Right}(t) - \bar{Q}_{Right}]}{\sqrt{\sum_t [Q_{Left}(t) - \bar{Q}_{Left}]^2} \sqrt{\sum_t [Q_{Right}(t) - \bar{Q}_{Right}]^2}} \right| \quad (6)$$

where  $Q$  is the temporal sequence of captured intensities,  $\bar{Q}$  the average intensity of the sequence and  $t$  the index of a captured intensity.

Finally, the census transform is introduced as the last basic cost measure. The census is able to encode local image structures with relative orderings of the pixel intensities rather than the intensity values themselves, thus being robust to radiometric changes and image noise [28]. Further, since the aperiodic sinusoidal fringes are generated with randomly varying parameters, if a fringe pitch is too large, the pixels within the fringe may be encoded with insufficient artificial textures. Therefore, the census transform obtained over a small window is more robust than the single pixel-based intensity difference. For initialization, images from the two camera are processed with the census transform with a  $3 \times 3$  window. The cost can then be obtained by

$$C_{census}(p, d) = HamDist \left[ I_{cen}^{Left}(p), I_{cen}^{Right}(pd) \right] \quad (7)$$

where  $HamDist$  is to calculate the hamming distance of the images handled by the census transform.

With all of the cost measures mentioned above, we integrate them into a combined measure which can be expressed by

$$C(p, d) = \frac{w_{AD}C_{AD}(p, d) + w_{gradAD}C_{gradAD}(p, d) + w_{NCC}C_{NCC}(p, d) + w_{census}C_{census}(p, d)}{w_{AD} + w_{gradAD} + w_{NCC} + w_{census}} \quad (8)$$

where  $w_{AD}$ ,  $w_{gradAD}$ ,  $w_{NCC}$  and  $w_{census}$  are the weight parameters used to control the influence of each cost measurement.

### 2.3. Motion-oriented cost aggregation

This step is to aggregate each pixel's matching cost over a support region, which is able to reduce erroneous matches caused by matching ambiguities or image noise. The aggregation is based on a simple but useful assumption is that the central pixel and its surrounding pixels should have similar disparities. To achieve a high measuring resolution, the support window usually spreads on the temporal axis and spatially the window size is  $1 \times 1$ . Actually, for this strategy there is a hidden constraint that the disparity between two corresponding pixels should keep unchanged within the period of the temporal correlation. This constraint generally holds well for static objects since the same point is measured constantly by two fixed corresponding pixels from the cameras. For example, as shown in Fig. 1(a), the central pixel of the left camera will capture the same color sequence as that of the right camera does. However, when facing a moving object, the constraint is difficult to meet. This is illustrated by Fig. 1(b) in which the pixel of interest in the left camera captures a sequence of varying colors as the object is moving. The corresponding

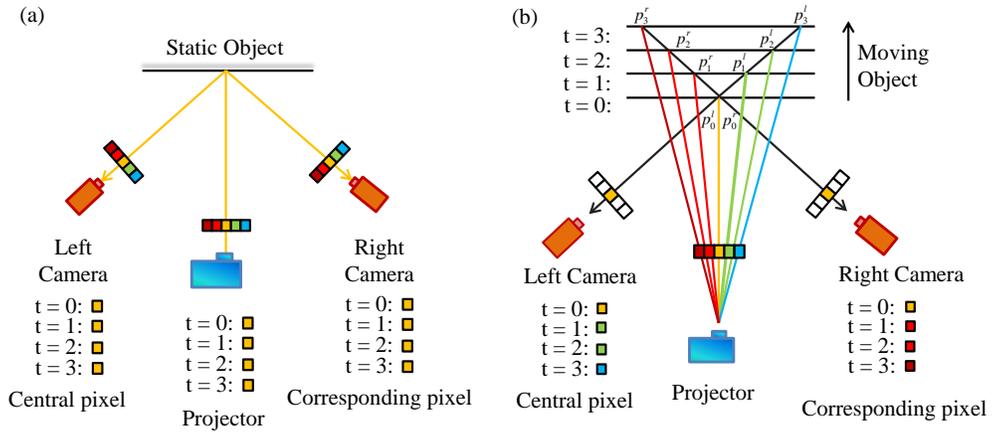


Fig. 1. (a) Temporal sequence of captured intensities from a static object; (b) Temporal sequence of captured intensities from a moving object.

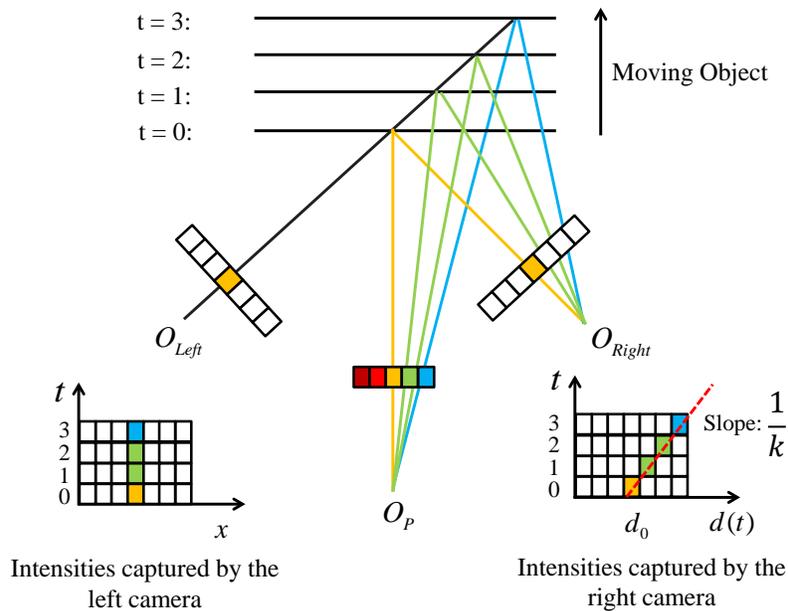


Fig. 2. Schematic of the intensity tracing algorithm.

pixel of the right camera also photographs a sequence but is distinct from the one obtained by the left view. Since the pair of pixels captures different sequence, it would be hard for the traditional temporal correlation method using a fixed temporal window to correctly estimate the disparity.

To handle this issue, we develop a motion-oriented cost aggregation method with the purpose of tracing the varying intensity in the right camera given the influence of the motion. Figure 2 shows the principle of the tracing algorithm. The optical centers of the cameras and the projector are represented by  $O_{Left}$ ,  $O_{Right}$  and  $O_p$  respectively. We assume that the speed of the moving object is constant over a short period of time. This is generally valid for high speed fringe projection measurements. From the temporal axis (vertical axis) of the right image, we can find that the corresponding pixel is actually shifting far away from the central pixel and moving towards the right edge of the image. Hence, to trace the desired intensity, we can establish a linear relation between the time and the disparity for each frame. It is noted that since the projected fringes used in our work are vertical, the tracing process for the corresponding light intensities is conducted horizontally. To perform the trace, the ideal disparity of the corresponding pixel in the right camera is to be determined with respect to the first frame ( $t = 0$ ). Given that the size of the temporal support window is  $T$  ( $t \in [0, 1, \dots, T - 1]$ ) and the initial disparity at time  $t = 0$  is  $d_0$ , we predict the offset of the matching point in consecutive images by

$$d(t) = \frac{t}{k} + d_0 \quad (9)$$

where  $\frac{1}{k}$  is the slope of the linear relation between the time and disparity. From Fig. 2 we can find that this parameter is closely related to the speed of the object motion. For example, a large value of  $k$  indicates a movement at a low speed, thus leading to a small disparity shift between successive frames. To simulate motions with different velocities, here  $k$  takes values ranging from  $[-K, K]$ ,  $[-\frac{1}{K}, -\frac{1}{K-1}, \dots, -1]$  and  $[1, \dots, \frac{1}{K-1}, \frac{1}{K}]$ , where  $K$  is selected as an integer to simplify the tracing procedure. It is noted that when the  $k$  takes the number of zero, we set  $d(t)$  equals to  $d_0$  which is the case of a fixed temporal window. Different from the conventional methods, our approach allows the collection of the costs flexibly with the consideration of motions. Since different  $k$  produces different tracing results, the ideal one should have the minimum cost since it fits the motion best. So the cost volume for the left camera is calculated as

$$C(p, d) = \min_k C(p, d(t)) \quad (10)$$

#### 2.4. Initial estimation of disparity and three-step framework of refinement

To obtain the disparity map from the aggregated cost volume, researchers usually adopt the winner takes all method by which the disparity with the lowest aggregated cost is selected [15, 19, 24]. However, according to practical experiments, we find that the disparity with the lowest cost may not always be the real disparity. For some cases, the second or the third lowest cost would correspond to the actual disparity. Therefore to cope with this problem, we set a searching range for the ideal disparity level. Suppose that for each pixel  $N$  disparity candidates with lowest costs are taken into account for the search. The  $N$  disparity candidates can be found by

$$d_1 = \arg \min_{\{d | d_{\min} \leq d \leq d_{\max}\}} C(p, d) \quad (11)$$

$$d_2 = \arg \min_{\{d | d_{\min} \leq d \leq d_{\max}, d \neq d_1\}} C(p, d) \quad (12)$$

$$d_3 = \arg \min_{\{d | d_{\min} \leq d \leq d_{\max}, d \neq d_1, d \neq d_2\}} C(p, d) \quad (13)$$

$$d_N = \arg \min_{\{d | d_{\min} \leq d \leq d_{\max}, d \neq d_1, d \neq d_2, \dots, d \neq d_{N-1}\}} C(p, d) \quad (14)$$

where  $d_{\min}$  and  $d_{\max}$  represent the minimum and maximum disparity level. With the calculated disparity candidates, we determine the optimal one among them with considerations of neighboring structures. Since the projected fringes are vertical, we choose a one dimensional horizontal window to compare the similarity. Within the support window, we also use the idea of the combined cost measure where the absolute difference of image intensity, absolute difference of image gradient and normalized cross correlation are included. So the measures can be written as

$$C_{Nb}(d_n) = \frac{C_{AD}^W(d_n) + C_{gradAD}^W(d_n) + C_{NCC}^W(d_n)}{3} \quad (15)$$

$$C_{AD}^W(d_n) = \frac{\sum_{x=-r_N}^{x=+r_N} |W^{Left}(x) - W^{Right}(x - d_n)|}{(2r_N + 1)\eta_1} \quad (16)$$

$$C_{gradAD}^W(d_n) = \frac{\sum_{x=-r_N}^{x=+r_N} |W_{grad}^{Left}(x) - W_{grad}^{Right}(x - d_n)|}{(2r_N + 1)\eta_2} \quad (17)$$

$$C_{NCC}^W(d_n) = 1 - \frac{\sum_{x=-r_N}^{x=+r_N} [W^{Left}(x) - \bar{W}^{Left}(x - d_n)] [W^{Right}(x) - \bar{W}^{Right}(x - d_n)]}{\sqrt{\sum_{x=-r_N}^{x=+r_N} [W^{Left}(x) - \bar{W}^{Left}(x - d_n)]^2} \sqrt{\sum_{x=-r_N}^{x=+r_N} [W^{Right}(x) - \bar{W}^{Right}(x - d_n)]^2}} \quad (18)$$

where  $W^{Left}$  and  $W^{Right}$  are the light intensities in the support window,  $W_{grad}^{Left}$  and  $W_{grad}^{Right}$  the image gradients,  $r_N$  the radius of the window,  $C_{Nb}$  the combined cost measure,  $\eta_1$  and  $\eta_2$  the scalars, and  $n$  the index of the disparity candidate. Thus for each pixel, the real disparity  $d'$  can be found by

$$d' = \arg \min_{d_n \in \{d_1, d_2, \dots, d_N\}} C_{Nb}(d_n) \quad (19)$$

Since we only choose the best disparity from the candidates, the usage of the horizontal window will not affect the measuring resolution. By Eq. (19), we can calculate an initial disparity map. To further refine the disparity results, a framework of refinement is developed. From practical experience, we find that outliers are difficult to be removed effectively by a single step, e.g. the left-right constancy check. Therefore, our framework consists of three steps to obtain accurate disparity maps.

**Step 1: Elimination of Occluded Regions.** Pixels in the occluded areas normally show incorrect disparities which will adversely influence the 3-D reconstruction result. For the occluded regions discussed here, they are classified into two kinds. The first class is caused by the fringe projection setup since no fringes can be captured in the shadowed areas. The second type of occlusion is induced by the binocular system because the two cameras are measuring an object from different sides. To detect the pixels for the first kind of occlusion, we have

$$(x, y) \in \{I_0(x, y) < \alpha \cup I_2(x, y) < \alpha \dots \cup I_{T-1}(x, y) < \alpha\} \quad (20)$$

where  $\alpha$  is a threshold for the captured intensity. If a pixel is constantly capturing very low intensities, it is likely for the pixel to be in the shadow. To identify the second class of occluded regions, we utilize the left-right consistency check [29]. This check is effective because of the fact that the disparity value of two corresponding pixels from the conjugate views should be the same. Suppose that the disparity results obtained from both cameras are denoted as  $D_{Left}$  and  $D_{Right}$ , then we have

$$(x, y) \in \{|D_{Left}(x, y) - D_{Right}(x - D_{Left}(x, y), y)| > \beta\} \quad (21)$$

where  $\beta$  is a threshold of the absolute difference between left and right disparities. By Eqs. (20) and (21), pixels satisfying the relations are eliminated as occluded pixels. Additionally, some pixels corresponding to right image points with  $(x - D_{Left}(x, y)) < 0$  are also treated as occluded pixels and need to be removed.

**Step 2: Left-right Consistency Check Based Outlier Correction.** As for the correct disparity results, they should be constant for the left and right camera views, which can be written as

$$D_{Left}(x, y) = D_{Right}(x - D_{Left}(x, y), y) \quad (22)$$

Thus, for the outliers surviving previous step, we can use Eq. (22) to locate their positions. For the correction, we utilize the reliable pixels surrounding an outlier to predict a suitable disparity for it. To find reliable points near the outlier, firstly the selected pixels should meet the requirement of Eq. (22). Then, we consider that pixels with shorter Euclidean distance and higher similarity of gray intensity with respect to the central pixel (the outlier) would be more reliable than other pixels. So the weight function can be expressed as

$$w_r(p', p'') = \exp \left[ - \left( \frac{\sqrt{(p'_x - p''_x)^2 + (p'_y - p''_y)^2}}{\gamma_d} + \frac{|I(p') - I(p'')|}{\gamma_i} \right) \right] \quad (23)$$

where  $p'$  is the image coordinate of the neighboring pixel and  $p''$  that of the outlier.  $\gamma_d$  and  $\gamma_i$  are used to adjust the spatial and intensity effect in a  $M \times M$  window. By setting a threshold  $\delta$  for the weight, the reliable pixels within the support window can be found with a weight value above the threshold. Since for these selected pixels, they may have slightly different disparities, a histogram is used to choose the disparity with the highest frequency. Finally, we replace the erroneous disparity of the outlier with the selected one indicated by its reliable neighbors.

**Step 3: Sub-pixel Disparity Enhancement.** For the last step of the refinement framework, a sub-pixel disparity enhancement is conducted to reduce the errors discontinuities caused by the discrete quantization in the depth. Conventionally, the methods of parabola fitting [30] and Lucas–Kanade [19] are used to obtain sub-pixel disparity results. Since artificial peaks may emerge due to the effect of pixel-locking when using the method of parabola fitting [31], we adopt the latter in our work. For pixel  $p$ , the final sub-pixel disparity is expressed as

$$d'' = d_{ref} + d_{off} \quad (24)$$

where  $d_{ref}$  is the integer disparity obtained after the implementations of the refining steps and  $d_{off}$  a small fractional offset. By the Lucas–Kanade method, the sub-pixel offset can be calculated by

$$d_{off} = flow_{vx} \quad (25)$$

$$flow = LK [I^{Left}(x, y), I^{Right}(x - d_r, y)] \quad (26)$$

where  $flow$  represents the optical flow solved by Lucas–Kanade algorithm  $LK(\cdot)$  [32] and the sub-script  $vx$  the velocity component along the horizontal direction.

## 2.5. Surface reconstruction

As the correspondence is obtained, this subsection describes the method to compute the position of a point  $P$  in 3-D space given its image in two views ( $p$  and  $p_d$ ) and the projection matrices ( $H^{Left}$  and  $H^{Right}$ ) that can be obtained once the cameras are calibrated. Since in each image we have a measurement  $p = H^{Left} P$  and  $p_d = H^{Right} P$ , the point  $P$  can be simply computed by combining these equations

$$\begin{cases} p = H^{Left} P \\ p_d = H^{Right} P \end{cases} \quad (27)$$

Eq. (27) shows an ideal triangulation by back-projecting rays from the measured image points. In general, however, there are errors in the measured points  $p$  and  $p_d$ , which means the rays back-projected from the points may be skew. Thus there may not be a point  $P$  which exactly satisfies the Eq. (27). In Fig. 3, this can be illustrated by the dashed lines which are skew in 3-D space.

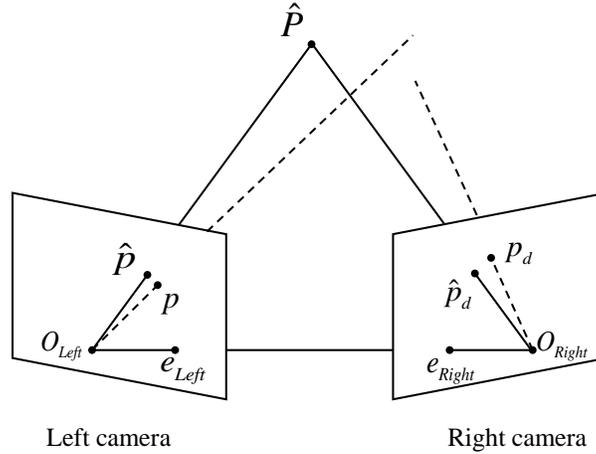


Fig. 3. Refinement of the measured correspondence ( $e_{Left}$  and  $e_{Right}$  are the epipoles).

To cope with this issue, we introduce the optimal triangulation method to retrieve the 3-D reconstruction [33]. A typical observation consists of a noisy point correspondence  $p \leftrightarrow p_d$  which does not generally meet the requirement of the epipolar constraint. In reality, the correct corresponding image points should be points  $\hat{p} \leftrightarrow \hat{p}_d$  lying close to the measured points  $p \leftrightarrow p_d$  and satisfying the epipolar constraint exactly. Therefore the optimal triangulation method aims at calculating a 3-D point  $\hat{P}$  from refined image points  $\hat{p} \leftrightarrow \hat{p}_d$  instead of from raw points  $p \leftrightarrow p_d$ , which is shown by the triangle of solid lines in Fig. 3. To seek the desired pair of points, we can minimize the function

$$C_{dist}(p, p_d) = dist(p, \hat{p})^2 + dist(p_d, \hat{p}_d)^2 \text{ subject to } \hat{p}_d^T F \hat{p} = 0 \quad (28)$$

where  $dist(\cdot)$  is the Euclidean distance between the points and  $F$  is the fundamental matrix which can be computed by the normalized 8-point algorithm [33]. Since any pair of points satisfying the epipolar constraint must lie on a pair of corresponding epipolar lines in the two images, the optimum point  $\hat{p}$  will lie on an epipolar line  $l$  and  $\hat{p}_d$  on the corresponding epipolar line  $l'$  as shown by Fig. 4. From the figure we can see that of all pairs of points on the lines  $l$  and  $l'$ , the points  $p_{\perp}$  and  $p_{d\perp}$  lie closest to the measured points  $p$  and  $p_d$ , which means they can minimize the Eq. (28). So it follows that  $\hat{p} = p_{\perp}$  and  $\hat{p}_d = p_{d\perp}$ , where  $p_{\perp}$  and  $p_{d\perp}$  are defined according to a pair of matching epipolar lines  $l$  and  $l'$ . Consequently, we can write  $dist(p, \hat{p}) = dist(p, l)$  and  $dist(p_d, \hat{p}_d) = dist(p_d, l')$ , and Eq. (28) can be rewritten as

$$C_{dist}(p, p_d) = dist(p, l)^2 + dist(p_d, l')^2 \quad (29)$$

where  $dist(\cdot)$  now indicates the perpendicular distance from a point to a line. As  $l$  and  $l'$  range over all possible corresponding epipolar lines, we need to find the optimum ones that minimize Eq. (29). The approach is detailed as follows.

For the left view, we parametrize the pencil of epipolar lines by a parameter  $v$ . To simplify the process, we apply a rigid transformation to each view in order to place both points  $p$  and  $p_d$  at

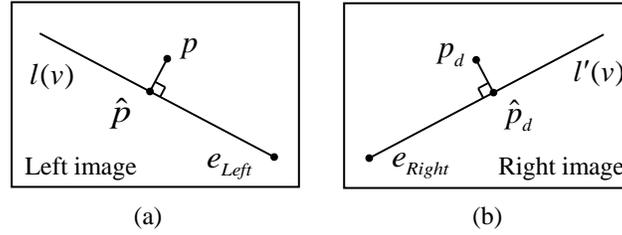


Fig. 4. The search for optimum points  $\hat{p}$  and  $\hat{p}_d$ . (a) Left view; (b) Right view.

the origin  $(0, 0, 1)^T$  in homogeneous coordinates. For the epipoles, they will be located on the x-axis at points  $(1, 0, f)^T$  and  $(1, 0, f')^T$  after the transformation. The value of  $f$  and  $f'$  equal to 0 indicates that the epipoles are at infinity. Since these two rigid transforms do not influence the sum of squares in Eq. (29), they will not change the minimization result. According to the two-view geometry, since  $F(1, 0, f)^T = (1, 0, f')^T F = 0$ , with the simplification the fundamental matrix can be represented by a special form

$$F = \begin{pmatrix} F_4 f f' & -F_3 f' & -F_4 f' \\ -F_2 f & F_1 & F_2 \\ -F_4 f & F_3 & F_4 \end{pmatrix} \quad (30)$$

Assume that an epipolar line in the left image passes through an image point  $(0, v, 1)^T$  and epipole  $(1, 0, f)^T$ , the vector representing the epipolar line can then be obtained by the cross product  $(0, v, 1) \times (1, 0, f) = (vf, 1, -v)$ . We denote this epipolar line by  $l(v)$ , so the squared distance from the line to the origin  $p$  can be expressed as

$$\text{dist}(p, l)^2 = \frac{v^2}{1 + (vf)^2} \quad (31)$$

With the fundamental matrix, we can find the corresponding epipolar line  $l'(v)$  in the right image by

$$l'(v) = F(0, v, 1)^T = [-f'(F_3 v + F_4), F_1 v + F_2, F_3 v + F_4]^T \quad (32)$$

So the squared distance of this line to the origin  $p_d$  can be written as

$$\text{dist}(p_d, l')^2 = \frac{(F_3 v + F_4)^2}{(F_1 v + F_2)^2 + f'^2 (F_3 v + F_4)^2} \quad (33)$$

The Eq. (29) can then be given by

$$C_{\text{dist}}(v) = \frac{v^2}{1 + (vf)^2} + \frac{(F_3 v + F_4)^2}{(F_1 v + F_2)^2 + f'^2 (F_3 v + F_4)^2} \quad (34)$$

As the fundamental matrix  $F$  is known, the problem of minimizing Eq. (29) is reduced to that of finding the minimum of the Eq. (34) of a single variable  $v$ . Once the optimum  $v$  is obtained, the epipolar lines  $l(v)$  and  $l'(v)$  can be computed readily. The points  $\hat{p}$  and  $\hat{p}_d$  are then determined by  $\hat{p} = p_{\perp}$  and  $\hat{p}_d = p_{d\perp}$ . Hence for the pair of corresponding points  $p$  and  $p_d$ , their 3-D point  $\hat{P}$  can be calculated from the refined points  $\hat{p}$  and  $\hat{p}_d$  by

$$\begin{cases} \hat{p} = H^{\text{Left}} \hat{P} \\ \hat{p}_d = H^{\text{Right}} \hat{P} \end{cases} \quad (35)$$

### 3. Experiments

To test the proposed method, we develop a dual-camera high speed fringe projection system. It consists of two monochrome cameras (Basler acA640-750) with the resolution of  $640 \times 480$ , a DLP light crafter with the resolution of  $608 \times 684$  and a laptop (Lenovo Y430P). The distance between the cameras is  $325\text{mm}$ . The projector is arranged between the cameras as illustrated by Fig. 1. The measured object can be placed arbitrarily in front of the system as long as the regions of interest can be illuminated and captured by both cameras. With the projector, the binary aperiodic sinusoidal fringes can be projected at  $4000\text{Hz}$ . However, due to the need of synchronization with the cameras with a top speed of  $490\text{Hz}$ , the measuring speed of our system is limited to 490 frames per second (fps).

Table 1. Parameter Settings

$w_{AD}$	$w_{gradAD}$	$w_{NCC}$	$w_{census}$	$K$	$T$	$d_{\min}$	$d_{\max}$	$N$
8	3	18	3	4	5	-100	280	6
$r_N$	$\eta_1$	$\eta_2$	$\alpha$	$\beta$	$\gamma_d$	$\gamma_i$	$\delta$	$M$
2	100	100	10	10	4.5	30	0.3	11

Firstly, we performed the stereo camera calibration with 20 images of a randomly placed calibration board using Bouguet's calibration toolbox [34]. Then, for captured fringe images from both cameras, they were rectified according to the calibration result. Table 1 shows the parameters of our technique. Since the employed support window  $T$  only spreads on the temporal axis, the spatial resolution of obtained results is equal to the full resolution of the used camera. With a sliding temporal window, high speed 3-D measurements can be carried out at 490 fps with the setup.

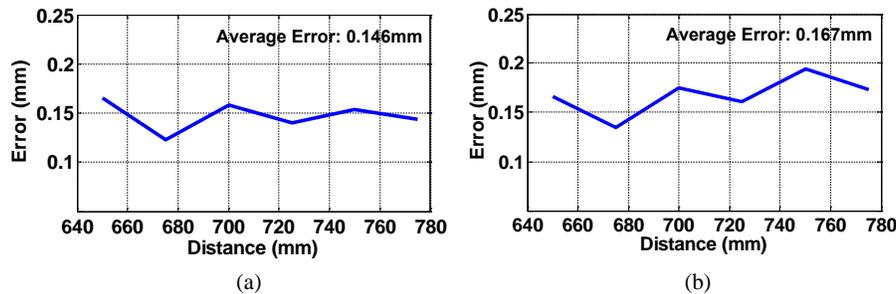


Fig. 5. Reconstruction error of a static plane. (a) TNCC; (b) Our method.

#### 3.1. Quantitative evaluation

For quantitative evaluation, we firstly measured a static planar surface and a couple of gage balls to see the method's performance of handling static objects. Then, a moving plane and a pendular ball were scanned respectively to analyze reconstructions for dynamic planar and sphere objects.

For the first static scene, a plane was placed at different discrete depths. The distance between the plane and the midpoint of the base line of cameras ranged from  $650\text{mm}$  to  $775\text{mm}$  with a step size of  $25\text{mm}$ . For comparison, the traditional temporal normalized cross correlation method (TNCC) [24] was also used to measure the plane with a temporal window size of 9 pixels. At each position, the 3-D reconstruction was calculated by averaging over ten measurements and the measuring accuracy was obtained by the average error of the 240th row of the retrieved surface.

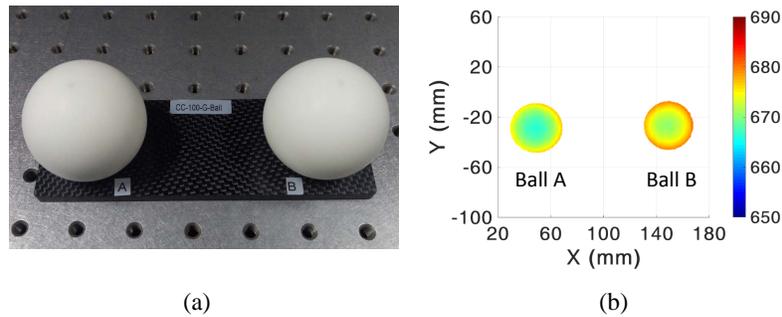


Fig. 6. Measurement of a couple of gage balls. (a) The gage balls; (b) 3-D reconstruction.

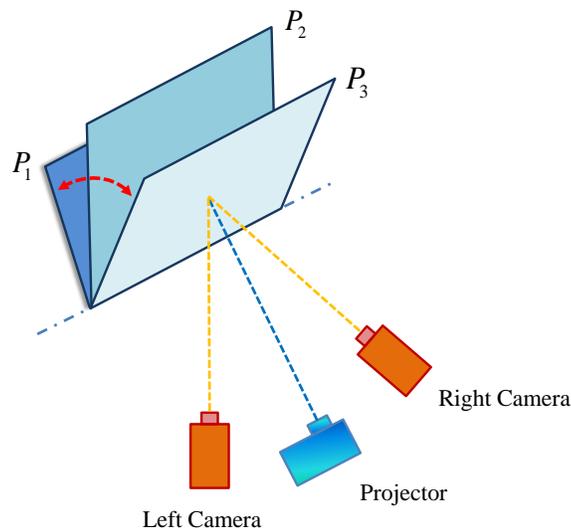


Fig. 7. Illustration of the moving process of the plane.

Figures 5(a) and 5(b) show the results obtained by the TNCC and our method respectively. From the figures, we can see that in general our technique has similar accuracy compared to traditional TNCC except that our method shows slightly larger reconstruction error when the plane was at  $750\text{mm}$ . For the quantitative results, the average error of our method over the selected depths is  $0.167\text{mm}$  and that of TNCC is  $0.146\text{mm}$ . Since our result was obtained by a temporal window size of 5 pixels and the observed largest error of our method is less than  $0.2\text{mm}$ , the proposed method would still meet the accuracy requirements of high speed measurements. Then, we measured a couple of ceramic gage balls as shown by Fig. 6(a). Firstly, the dimension of gage balls obtained with a coordinate measuring machine was referred as the ground truth. The radii of the balls and the distance between centers of the spheres were then computed based on the 3-D point cloud reconstructed with our method, as shown in Fig. 6(b). For the ground truth, the radius of Ball A is  $25.398\text{mm}$ , that of Ball B is  $25.403\text{mm}$ , and their distance is  $100.069\text{mm}$ . By the proposed technique, the obtained radii are  $25.533\text{mm}$  and  $25.568\text{mm}$  respectively for Ball A and Ball B, and the distance of their centers is  $100.198\text{mm}$ . It can be seen that for the tested gage balls the maximum error is  $0.165\text{mm}$ , the minimum error is  $0.129\text{mm}$  and the average error is  $0.143\text{mm}$ .

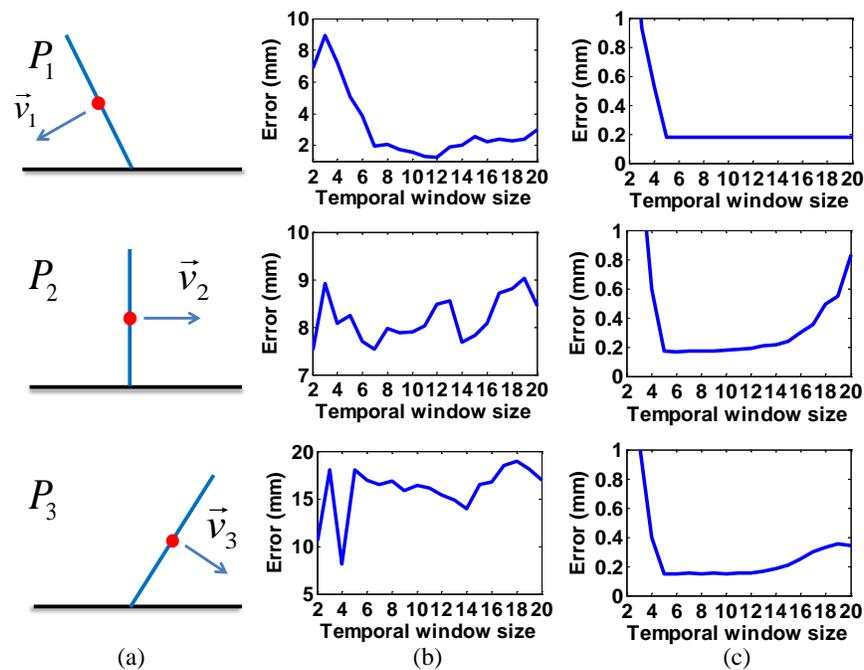


Fig. 8. (a) Selected three positions of the plane; (b) Reconstruction error of the TNCC for each position; (c) Reconstruction error of our method for each position.

To see the method's accuracy of handling dynamic planar surfaces, we firstly inspected a plane that was turned up and down by a hand. Figure 7 illustrates the moving process of the plane with which we selected three different moments with positions  $P_1$ ,  $P_2$  and  $P_3$  to test the instantaneous accuracy. Since an appropriate size of the temporal support window is beneficial to achieving favorable results, we also compared the calculated 3D reconstructions with different temporal windows and the reconstruction error was still computed by averaging the ones over the 240th row. Figure 8(a) displays the measured three positions, Figs. 8(b) and 8(c) the reconstruction error of the TNCC and our method for the positions respectively. In general, it can be seen that our approach outperforms the traditional method of TNCC. To be specific, for TNCC although it shows a relatively small error when the plane was at  $P_1$ , the smallest error is  $0.92\text{mm}$  (window size takes 12 pixels) and the average error is  $1.97\text{mm}$  when the window size is larger than 7. However, for the same position, with our method the smallest error is  $0.178\text{mm}$  and the average error is  $0.179\text{mm}$  calculated over the same range of windows. For the positions  $P_2$  and  $P_3$ , the traditional technique begins to show much larger measuring errors which are generally above  $7\text{mm}$  as can be observed from Fig. 8(b). But, for our method, it still performs well with the average errors of  $0.176\text{mm}$  and  $0.152\text{mm}$  respectively for the window size ranging from 5 to 12. From the comparison, we can find that the accuracy can be increased by an order of magnitude with the presented method. For further analysis, according to results obtained by our method, we can find that the plane moved at a relatively low speed when it was at  $P_1$  because the measuring error shows a steady distribution with the increase of the size of the temporal window. When the plane was turned to the position  $P_2$ , the speed was supposed to reach its highest value among the three moments as can be seen the measuring error is going to soar once the window size is more than 12. Based on the results as shown in Fig. 8(c), we can see that the error has a steady distribution with small variations when the window size starts from 5. Therefore, we set the size of the temporal window as 5 pixels in our experiments.

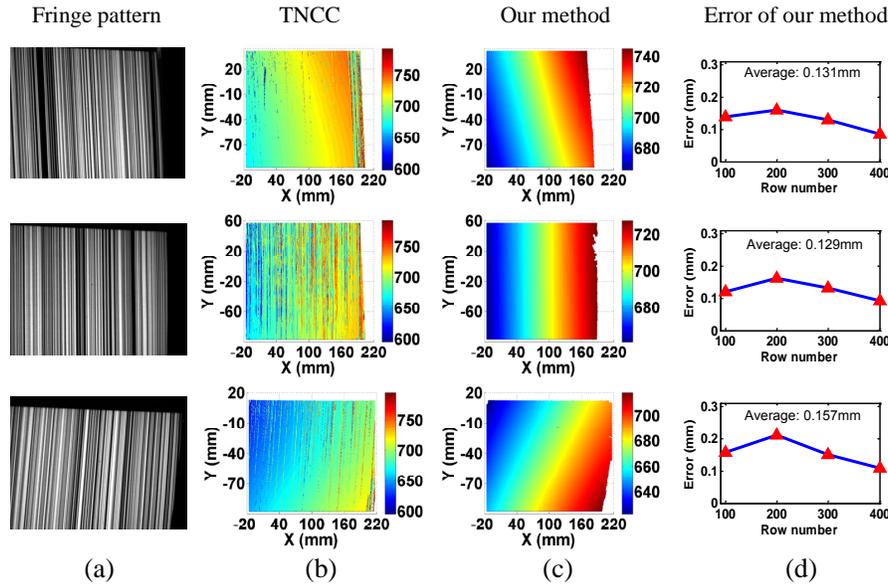


Fig. 9. Measurement results obtained with the window size of 5. (a) Captured fringe images for positions  $P_1$  to  $P_3$ ; (b) 3-D reconstructions of the three positions obtained by TNCC; (c) 3-D reconstructions of the three positions obtained by our method; (d) Reconstruction error of our method for the 320th column of the plane.

With the selected window, we then show the whole 3-D reconstructions of the plane in Fig. 9. From the global view, our results indicate smooth and flat reconstructed shapes while lot of errors can be observed as abrupt prickles when TNCC was used. As the plane was rotating, it is not hard to find that the tangential velocity will be different due to the varying distance to the axis of rotation. To see the accuracy obtained at a certain moment with different rotation velocities, we display the error of a vertical column (the 320th column) of the plane, as shown in Fig. 9(d). Although the speed is changing, the error of the retrieved surface is generally small. Lastly, to visualize the intensity tracing process, by Fig. 10 we display the tracing procedure for the plane at the position  $P_2$ . We arbitrarily chose two pixels from the left camera and to observe whether the matching points calculated by our method and TNCC could capture the same intensity sequence as the left pixel did or not. From the results, we can see that by the proposed method the captured intensity sequences have been accurately matched to the ones in the left view while with TNCC the maximum temporal correlation coefficient failed to correctly indicate the intensities.

Next, we changed the plane into a table tennis ball to show the reconstruction accuracy for dynamic spherical objects. During the experiment, the ball was suspended from a static pivot and swung freely. We calculated the radius of the ball at five different moments selected from a same period of oscillation. The captured fringe patterns and the corresponding retrieved results are shown in Fig. 11. From the reconstructions, we can see that the ball has been successfully measured for each moment. For detailed analysis, the obtained 3-D point cloud was fitted to a spherical function for the computation of sphere's center and radius. The real radius of the used table tennis ball is  $20\text{mm}$  and the measured results are shown in the second column of Table 2 from which the average error of the calculated radii is  $0.14\text{mm}$ . So in spite of the moving process, the spherical surface can be correctly recovered with the proposed method. Furthermore, the centers of the sphere were used to calculate the instantaneous velocity of the ball. By computing

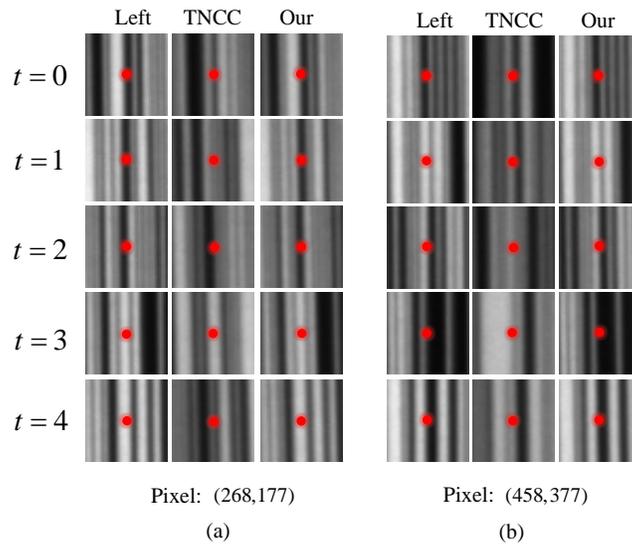


Fig. 10. Demonstration of the practical intensity tracing procedure. (a) The result of pixel (268,177); (b) The result of pixel (458,377).

the spatial distance between successive 3-D locations and dividing by the temporal interval of  $1/490$  second, the velocity was obtained as shown in the third column of Table 2. According to the results, we can see that the ball moved at a relatively low speed of  $263.38\text{mm/s}$  at the first moment and then began to accelerate as indicated by the velocity for the second moment. It reached the highest speed of  $601.32\text{mm/s}$  at the third moment and after that started to slow down. The speed of the ball was then reduced to the minimum  $229.83\text{mm/s}$  at the fifth moment.

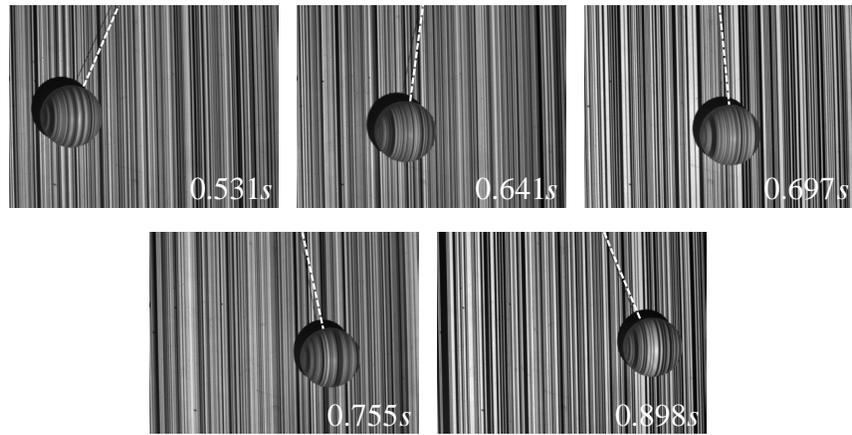
Table 2. Measurement Results of the Swinging Ball

Time	Radius (mm)	Speed (mm/s)
0.531s	19.88	263.38
0.641s	20.17	529.28
0.697s	19.82	601.32
0.755s	19.85	490.11
0.898s	20.08	229.83

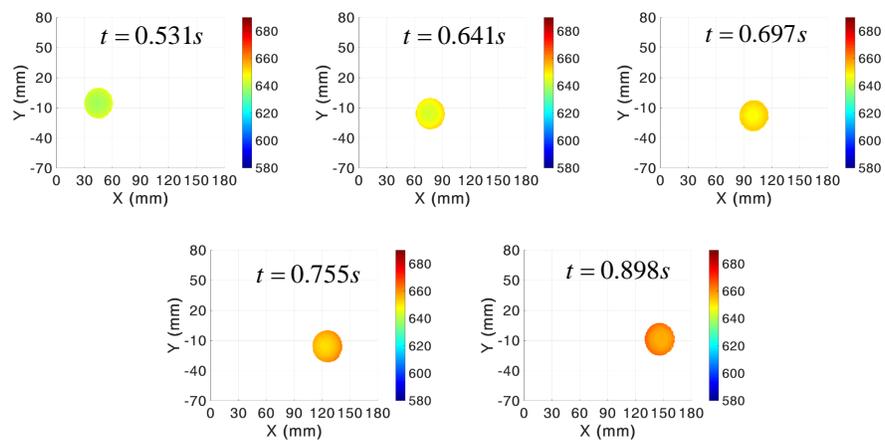
### 3.2. Qualitative results

To evaluate 3-D reconstructions qualitatively, we also conducted measurements for both static and dynamic scenes. Figure 12(a) displays a captured fringe pattern of a motionless plaster model. With the presented method, the obtained disparity map and the depth map are shown in Figs. 12(b) and 12(c) respectively. The depth was represented with negative values with the purpose of better visualization of the 3-D modal. From the results, we can see that the surface of the Chinese arch model has been reconstructed faithfully. Then, we scanned a more complex static scene that includes a table tennis ball, a white paper box (below the ball), a scroll of tape and a calibration board with white dots. Figures 12(d)-12(f) show the fringe image, calculated disparity map and depth image respectively. From the results, although these objects are spatially isolated, we are able to successfully retrieve the discontinuous surfaces with the developed approach.

For reconstructions of dynamic processes, we carried out two measurements for different



(a)



(b)

Fig. 11. (a) Fringe patterns of a swinging table tennis ball captured at different times; (b) Corresponding 3-D reconstructions.

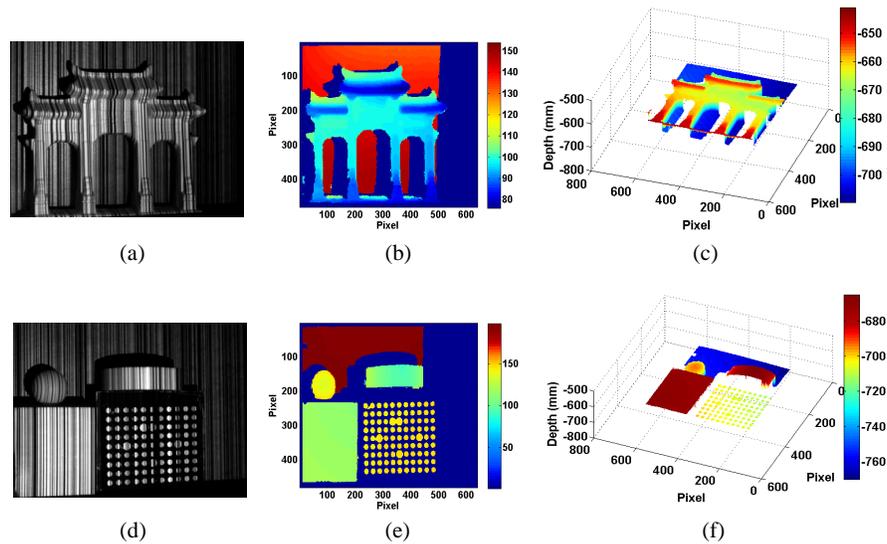


Fig. 12. (a) A captured fringe pattern of a motionless plaster model; (b) Calculated disparity map of the model; (c) The reconstructed depth; (d) A captured fringe image of a complex static scene; (e) Obtained disparity map of the scene; (f) The reconstructed depth.

objects. In the first one, we projected our patterns onto a moving hand. Its 3-D reconstruction is shown by [Visualization 1](#) and a screen shot of the video is displayed by Fig. 13. The left window shows the captured fringe image by the left camera, the central and right windows display the reconstructed 3-D model with a front view and a side view respectively. From [Visualization 1](#), the hand was steadily moving away from the system firstly and changed its moving direction to get close to the system afterwards. For the whole dynamic process, we can observe that the moving hand has been measured correctly and its movement retrieved smoothly. Then, we tested



Fig. 13. High speed 3-D measurement of a moving hand (see [Visualization 1](#)).

a deflating toy balloon for the second measurement. The result is shown in [Visualization 2](#) and a corresponding screen shot in Fig. 14. From the front view of the video, we can see that the global size of the balloon was becoming smaller and smaller with the process of the deflation. And from the side view, the height of the reconstructed object can be observed decreased gradually at the same time. According to the results, we can see that the entire process of the balloon from being swollen to shrunken has been reconstructed faithfully by our method. It is noted that since the high light intensity on the balloon affected the capture of projected patterns for both cameras, the corresponding regions were removed from the 3-D reconstruction (observed as two black dots). At the end of the deflating process, we also simulated its inverse—the inflating process by

playing the video in reverse.

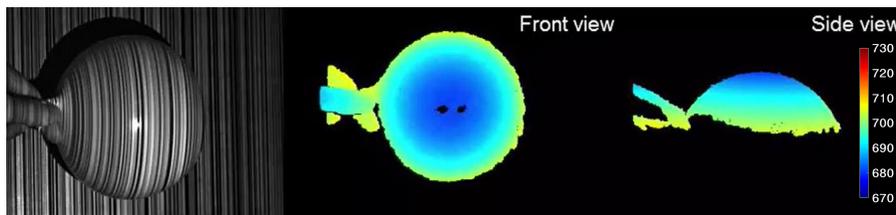


Fig. 14. High speed 3-D measurement of a deflating toy balloon (see [Visualization 2](#)).

#### 4. Conclusion

We have introduced a motion-oriented high speed 3-D measuring technique that combines the stereo and the fringe projection for robust and accurate 3-D measurements of fast moving objects. With the developed binary aperiodic sinusoidal fringes, the aperiodic sinusoidal patterns can be projected at kHz by a DLP projector with the aim of reducing the sensitivity to movements. To handle the problem that the motion violates the basic assumption that corresponding pixels from two cameras should capture intensities from the same object point, we suggest a motion-oriented temporal support window by which we can trace the desired light intensity temporally for the right camera according to the one captured from the left camera, thus reducing the number of mismatches. To robustly calculate the cost for pixels within the support window, an integrated cost measure has been presented to combine strengths of several single measures. In addition, a three-step optimization framework is further developed for refinement of obtained initial disparity results. With refining processes of the removal of outliers in occluded areas, corrections of residual outliers based on left-right consistency check and the sub-pixel disparity enhancement, accurate disparity maps at sub-pixel level are obtained for precise 3-D reconstructions. In experiments, high speed 3-D reconstructions of fast moving objects have been accurately obtained at 490 fps. Since the proposed method is able to trace the desired intensity, it may be further extended to applications where objects are encoded by gray code patterns [35] or statistical patterns [15].

Due to the employment of two cameras, only surfaces simultaneously viewed by the cameras and the projector can be retrieved correctly. Therefore, for complex objects, adjustments to object position or orientation or to the system configuration are preferred. Additionally, as the motion within the temporal support window is assumed to be linear, the method tends to compromise for the scenes vibrating or shaking too rapidly. We will carry on further researches on this issue in our future work.

#### Funding

National Natural Science Fund of China (NSFC) (11574152, 61505081, 61377003); 'Six Talent Peaks' project (2015-DZXX-009); '333 Engineering' research project (BRA2015294); Fundamental Research Funds for the Central Universities (30915011318); Final Assembly "13th Five-Year Plan" Advanced Research Project of China (30102070102); Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (3092014012200417).

#### Acknowledgments

C. Zuo thanks the support of the 'Zijin Star' program of Nanjing University of Science and Technology.