



Hybrid profilometry using a single monochromatic multi-frequency pattern

SEN XIANG,^{1,2} HUIPING DENG,¹ LI YU,^{1,*} JIN WU,¹ YOU YANG,²
QIONG LIU,² AND ZHENWEI YUAN²

¹School of Inform. Sci. and Engn., Wuhan Univ. of Sci. and Tech., Wuhan 430081, China

²School of Electron. Inform. and Commun., Huazhong Univ. of Sci. and Tech., Wuhan 430074, China

*huslyu@hust.edu.cn

Abstract: In this paper, we propose a novel profilometry scheme to acquire high quality depth, where only a single shot of a monochromatic pattern is utilized. We design a band-wise pattern consisting of fringe bands spatially modulated with coprime periods. After that, with the designed pattern, depth is obtained in a hybrid manner, where both phase-based profilometry and active stereo are incorporated. To be specific, pixels in smooth regions obtain their depth values through phases analysis. Especially, based on depth smooth property, we propose a novel phase unwrapping algorithm, which avoids the problem of error propagation and yields accurate unwrapping phases. On the other hand, for boundary regions, spatial stereo, which is more robust to depth discontinuities, is utilized to modify incorrect depth values. Both theoretical verification and experimental results demonstrate that the proposed scheme can generate high quality depth maps, even for complex scenes and isolated objects.

© 2017 Optical Society of America

OCIS codes: (110.6880) Three-dimensional image acquisition; (120.0120) Instrumentation, measurement, and metrology; (120.5050) Phase measurement.

References and links

1. A. Smolic, "3d video and free viewpoint video from capture to display," *Pattern Recogn.* **44**, 1958–1968 (2011).
2. Y. Yang, X. Wang, Q. Liu, M. Xu, and L. Yu, "A bundled-optimization model of multiview dense depth map synthesis for dynamic scene reconstruction," *Inform. Sci.* **320**, 306–319 (2015).
3. Y. Yang, H. Deng, J. Wu, and L. Yu, "Depth map reconstruction and rectification through coding parameters for mobile 3d video system," *Neurocomputing* **151**, 663–673 (2015).
4. H. Deng, J. Wu, L. Zhu, Z. Yan, and L. Yu, "Texture edge-guided depth recovery for structured light-based depth sensor," *Multimed. Tools Appl.* **76**(3), 4211–4226 (2017).
5. S. Naveen and R. S. Moni, "A robust novel method for face recognition from 2d depth images using DWT and DFT score fusion," in Proceedings of International Conference on Computational Systems and Communications (IEEE, 2015), pp. 1–6.
6. J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recogn.* **43**(8), 2666–2680 (2010).
7. S. S. Gorthi and P. Rastogi, "Fringe projection techniques: Whither we are?" *Opt. Laser Eng.* **48**(2), 133–140 (2010).
8. P. S. Huang and S. Zhang, "Fast three-step phase-shifting algorithm," *Appl. Opt.* **45**(21), 5086–5091 (2006).
9. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," *Appl. Opt.* **22**(24), 3977–3982 (1983).
10. Z. Zhang and J. Zhong, "Applicability analysis of wavelet-transform profilometry," *Opt. Express* **21**(16), 18777–18796 (2013).
11. S. Zhang and P. S. Huang, "High-resolution, real-time three-dimensional shape measurement," *Opt. Eng.* **45**(12), 123601 (2006).
12. J. L. Flores, J. A. Ferrari, G. G. Torales, R. Legarda-Saenz, and A. Silva, "Color-fringe pattern profilometry using a generalized phase-shifting algorithm," *Appl. Optics* **54**(30), 8827–8834 (2015).
13. C. Guan, L. Hassebrook, and D. Lau, "Composite structured light pattern for three-dimensional video," *Opt. Express* **11**(5), 406–417 (2003).
14. Y. Gong and S. Zhang, "Ultrafast 3-d shape measurement with an off-the-shelf dlp projector," *Opt. Express* **18**(19), 19743–19754 (2010).
15. K. Liu, Y. Wang, D. L. Lau, Q. Hao, and L. G. Hassebrook, "Dual-frequency pattern scheme for high-speed 3-D shape measurement," *Opt. Express* **18**(5), 5229–5244 (2010).
16. 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**(17), 19493–19510 (2012).

17. 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. Laser Eng.* **51**(8), 953–960 (2013).
18. E. Li, X. Peng, J. Xi, J. Chicharo, J. Yao, and D. Zhang, "Multi-frequency and multiple phase-shift sinusoidal fringe projection for 3d profilometry," *Opt. Express* **13**(5), 1561–1569 (2005).
19. W.-H. Su and H. Liu, "Calibration-based two-frequency projected fringe profilometry: a robust, accurate, and single-shot measurement for objects with large depth discontinuities," *Opt. Express* **14**(20), 9178–9187 (2006).
20. Y. Ding, J. Xi, Y. Yu, and J. Chicharo, "Recovering the absolute phase maps of two fringe patterns with selected frequencies," *Opt. Lett.* **36**(13), 2518–2520 (2011).
21. Y. Ding, J. Xi, Y. Yu, W. Cheng, S. Wang, and J. F. Chicharo, "Frequency selection in absolute phase maps recovery with two frequency projection fringes," *Opt. Express* **20**(12), 13238–13251 (2012).
22. X. Zhu, S. Cohen, S. Schiller, and P. Milanfar, "Estimating spatially varying defocus blur from a single image," *IEEE Trans. Image Process.* **22**(12), 4879–4891 (2013).
23. M. Zhao, L. Huang, Q. Zhang, X. Su, A. Asundi, and Q. Kemao, "Quality-guided phase unwrapping technique: comparison of quality maps and guiding strategies," *Appl. Optics* **50**(33), 6214–6224 (2011).
24. C. Zuo, L. Huang, M. Zhang, Q. Chen, and A. Asundi, "Temporal phase unwrapping algorithms for fringe projection profilometry: a comparative review," *Opt. Laser Eng.* **85**, 84–103 (2016).
25. Y. Xu and C. Ai, "Simple and effective phase unwrapping technique," *Proc. SPIE* **2003**, 254–263 (1993).
26. S. Liu and L. Yang, "Regional phase unwrapping method based on fringe estimation and phase map segmentation," *Opt. Eng.* **46**(5), 051012 (2007).
27. S. Zhang, X. Li, and S.-T. Yau, "Multilevel quality-guided phase unwrapping algorithm for real-time three-dimensional shape reconstruction," *Appl. Opt.* **46**(1), 50–57 (2007).
28. K. Chen and L. Song, "A composite quality-guided phase unwrapping algorithm for fast 3d profile measurement," *Proc. SPIE* **8563**, 856305 (2012).
29. K. Chen, J. Xi, and Y. Yu, "Fast quality-guided phase unwrapping algorithm for 3d profilometry based on object image edge detection," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (IEEE, 2012), pp. 64–69.
30. Z. Dai and X. Zha, "An accurate phase unwrapping algorithm based on reliability sorting and residue mask," *IEEE Geosci. Remote Sens. Lett.* **9**(2), 219–223 (2012).
31. H. Zhong, J. Tang, S. Zhang, and M. Chen, "An improved quality-guided phase-unwrapping algorithm based on priority queue," *IEEE Geosci. Remote Sens. Lett.* **8**(2), 364–368 (2011).
32. X. Su, W. Chen, Q. Zhang, and Y. Chao, "Dynamic 3-d shape measurement method based on ftp," *Opt. Laser Eng.* **36**(1), 49–64 (2001).
33. A. Asundi and Z. Wensen, "Fast phase-unwrapping algorithm based on a gray-scale mask and flood fill," *Appl. Optics* **37**(23), 5416–5420 (1998).
34. M. A. Herráez, D. R. Burton, M. J. Lalor, and M. A. Gdeisat, "Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path," *Appl. Opt.* **41**(35), 7437–7444 (2002).
35. A. Baldi, "Phase unwrapping by region growing," *Appl. Opt.* **42**(14), 2498–2505 (2003).
36. T. Pribanić, S. Mrvođa, and J. Salvi, "Efficient multiple phase shift patterns for dense 3d acquisition in structured light scanning," *Image Vision Comput.* **28**(8), 1255–1266 (2010).
37. Y. Wang and S. Zhang, "Superfast multifrequency phase-shifting technique with optimal pulse width modulation," *Opt. Express* **19**(6), 5149–5155 (2011).
38. Z. Lei, C. Wang, and C. Zhou, "Multi-frequency inverse-phase fringe projection profilometry for nonlinear phase error compensation," *Opt. Laser Eng.* **66**, 249–257 (2015).
39. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision* **47**(1-3), 7–42 (2002).
40. A. F. Bobick and S. S. Intille, "Large occlusion stereo," *Int. J. Comput. Vision* **33**(3), 181–200 (1999).

1. Introduction

Depth is a fundamental element in various applications, such as free-viewpoint video (FVV) [1], scene reconstruction [2–4] and face recognition [5]. Therefore, high quality depth maps are very much desired. Among various depth generation approaches, structured light (SL) [6] techniques have attracted attentions from both the academic and the industry fields.

As shown in Fig. 1, a typical SL system consists of a projector and a camera. Projecting a designed pattern I_{prj} to a reference plane H_{ref} and the actual measured object will report a reference pattern I_{ref} and a captured pattern I_{cap} , respectively. After that, by detecting the deformation between I_{cap} and I_{ref} , depth of the measured object can be obtained. Based on this principle, phase-based profilometry [7] embeds phase information to sinusoidal patterns and thus provides depth values in an accurate and convenient manner.

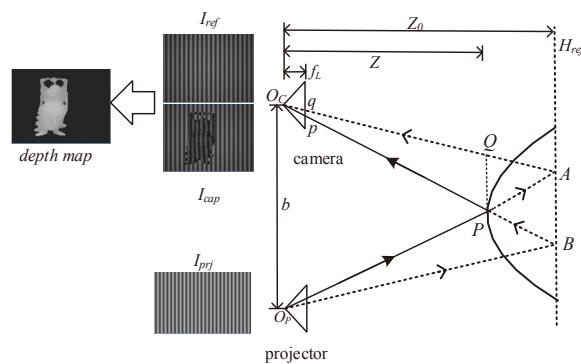


Fig. 1. Sketch of a SL system.

1.1. Phase-based profilometry

Phase-based profilometry can be further categorized into phase shifting profilometry (PSP) [8], Fourier transform profilometry (FTP) [9] and wavelet transform profilometry [10]. PSP utilizes multiple fringes with specified initial-phases to calculate the phase map, which in general needs multiple shots. Most improved versions focus on reducing the shots and use multiplexing techniques [11–13] or ultrafast imaging device [14]. FTP [9] is based on signal transform and processing between spatial domain and frequency domain. Briefly speaking, in FTP, phase map is obtained via successive steps including Fourier transform, band-pass filtering, inverse Fourier transform and phase computation with ‘arctan’ function. Compared with PSP, FTP strictly needs only a single shot and therefore is applicable to dynamic and colorful scenes. WTP [10] is similar with FTP, but it replaces Fourier transform with wavelet transform. Since wavelet transform is more sensitive to frequency changes, WTP improves phase accuracy over FTP, but it faces much higher complexity due to wavelet analysis.

However, the calculated phases are wrapped to the range of $(-\pi, \pi]$, i.e., the real phase ϕ_{uw} and the calculated phase ϕ_w satisfy $\phi_{uw} = \phi_w + 2m\pi$, where $m \in \mathbb{N}$. This causes phase ambiguity, and phase unwrapping is necessary to recover ϕ_{uw} from ϕ_w . To achieve this goal, multi-frequency phase profilometry methods are introduced. In these methods, multiple patterns modulated with several frequencies are applied. Correspondingly, several phase maps are obtained, with which the unwrapped phase map ϕ_{uw} can be retrieved. On the other hand, to acquire the phase maps, multiple patterns are necessary, no matter for PSP or FTP, which degrades the real-time performance. Some researchers tried to design composite patterns that can reduce the number of projections [15–17]. Especially, in [18, 19], multi-frequency components are integrated into only a single pattern. Other researchers focus on the selection of the modulation frequencies. For example in [20, 21], the authors investigated the relationship between the real phase and the two wrapped phases with given co-prime modulation periods, and derived a look-up-table to achieve fast phase unwrapping.

In modern applications, high quality depth maps are of great importance. To be specific, depth maps should be accurate, dense and attainable from a single-shot. Unfortunately, it is still challenging to meet these demands. The first concern is the accuracy of depth maps. The second challenge is the efficiency, especially that dynamic scenes need single-shot solutions. The last one is solving the important phase unwrapping problem. More importantly, these demands are not easy to meet simultaneously. For example, FTP strictly needs only a single shot, but with larger error at depth transitions. Another example is that multi-frequency profilometry is a good choice for phase unwrapping, but its multiple shots reduce the applicability in dynamic scenes. Therefore, novel structured light schemes are still to be developed.

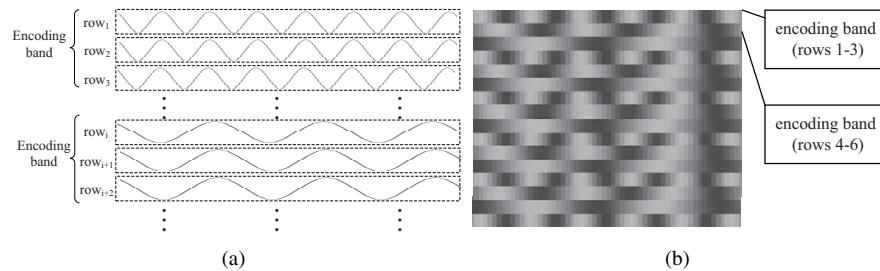


Fig. 2. The proposed encoding band and the pattern. (a): A sketch map. (b): Details of the pattern.

1.2. Overview of the proposed scheme

We propose, in this paper, a novel hybrid scheme, which only needs a single-shot of a monochromatic fringe, to generate high quality depth maps. The proposed scheme is based on phase-analysis and active stereo, thus inheriting the advantages of both approaches. The contributions of this paper are three-fold. First, the pattern is compatible with classic FTP, so wrapped phases can be calculated conveniently. Second, based on the designed pattern, we propose a novel phase unwrapping algorithm: local-smoothness-based phase unwrapping, which can handle scenes with large discontinuities and isolated objects. Last but not least, the pattern is locally unique, and active stereo can be applied to modify and improve the depth at object boundaries.

The remainder of this paper is organized as follows. In Section 2, we present the proposed scheme in detail. After that, simulation and experimental results are shown in Section 3. Finally, Section 4 summarizes this paper with a conclusion.

2. The proposed hybrid depth measuring scheme

2.1. Pattern design

In most monochromatic fringes, the whole pattern is modulated with a single frequency. In contrast, the modulation frequency of the proposed pattern is location-dependent, i.e., the intensity of pixel (i, x) is

$$\begin{aligned} I_{prj}(i, x) &= \text{Acos}(2\pi f_i^m x) + B \\ &= \text{Acos}(2\pi x/T_i^m) + B \end{aligned} \quad (1)$$

where f_i^m and T_i^m are the frequency and the period, respectively.

In the simplest case, the frequency varies across lines. However, in practice, the depth-of-field of the projector lens is relatively short and defocus blur occurs. To be specific, the pattern on the measured object can be modeled as the convolution of the in-focus image and a Gaussian kernel [22], which is related to depth. As the measured object goes further from the in-focus plane, the blurring effect becomes severer. Therefore, if the pattern is modulated by lines, it will be totally blurred.

Instead, we define ‘encoding band’, which is a union of neighboring lines, as the basic modulation unit as shown in Fig. 2(a). Its size should be large enough to deal with blurring. More specifically, as the measured object goes further from the in-focus plane, the encoding band should become wider. In our experiment, we find three-pixel-wide is enough for 3dsMax simulation and five-pixel-wide performs well in the real structured light system. All lines in an encoding band share an identical frequency f_i^m and period T_i^m . On the other hand, f_i^m and T_i^m are not determined by the row number i , but by the encoding band. Fig. 2(b) presents the details of the designed pattern.

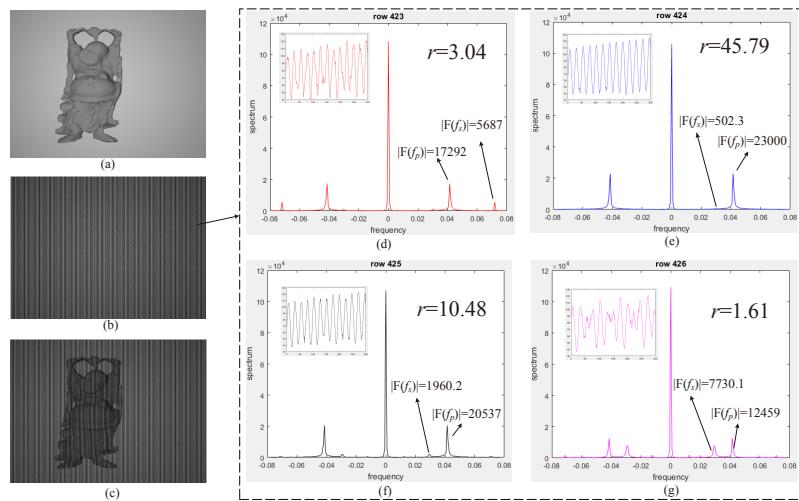


Fig. 3. An example of determining the typical line. (a)-(c): are the 3D model, the reference pattern and the captured pattern, respectively. (d)-(g): The spectrum diagram of line 423-426 respectively. The energy $|F(f_p)|$, $|F(f_s)|$ and the ratio r are shown in the figures. The waveform of each line is partly presented in the upper-left corner.

The modulation period T_i^m is determined due to two considerations. On the one hand, the periods of neighboring bands must be coprime, which brings two advantages. First, based on the idea of coprime periods and temporal phase unwrapping, we propose a novel local-smoothness based phase unwrapping method as detailedly addressed in Section 2.2.2. Second, coprime periods make the projected pattern unique in terms of local intensity distribution, with which we can further modify incorrect depth values as presented in Section 2.2.4. On the other hand, once meeting the ‘coprime’ condition, relative higher frequencies are preferred. This is because fringes with higher frequencies are more robust to the impairment of the DC component in the pattern, and yields more accurate phases.

2.2. Depth calculation

Depth values are obtained with a hybrid scheme, which makes full use of the designed pattern. More detailedly, it is further divided into several sub-steps: (1) wrapped phase calculation, (2) phase unwrapping, (3) phase-depth conversion and (4) depth modification with stereo matching.

2.2.1. Wrapped phase calculation

In implementation, the projected pattern I_{prj} and the captured pattern I_{cap} usually do not form a pixel-to-pixel mapping. In I_{prj} , the modulation frequency is constant within an encoding band, but varies across encoding bands. Therefore, several frequency components may exist in a line of the captured pattern I_{cap} , and this impairs the phase calculated with FTP.

With the reference pattern I_{ref} , we can solve this problem efficiently. Among the frequency components in I_{ref} , the two major ones with the largest and the second largest energy are named primary frequency $f_p(i)$ and secondary frequency $f_s(i)$, respectively. We take the pattern of ‘Buddha’ as an example, and the waveforms and spectra of rows 423-426 are presented in Fig. 3. In Fig. 3(d), $f_p(423) = 0.0414$ has the largest energy and $f_s(423) = 0.0719$ has the second largest energy. Moreover, I_{prj} and I_{ref} are related, and thus neighboring rows should have the same $f_p(i)$. We combine these rows with the same $f_p(i)$ as a ‘decoding band’. In Fig. 3(d)-(g), the four rows share the same $f_p = 0.414$, and they form a ‘decoding band’.

Within a decoding band, the rows have different spectrum distributions, and the energy ratio between $f_p(i)$ and $f_s(i)$ of row i is defined as

$$r_i = \frac{|F(f_p(i))|}{|F(f_s(i))|} \quad (2)$$

where F means Fourier transform, and $|F(f)|$ is the energy at frequency f . The border rows of a decoding band receive more interference, and thus have more frequency components and smaller r values. On the contrary, the center row is often dominated by a single frequency with the largest r . In fact, I_{ref} is dependent on I_{prj} , and the major frequencies in I_{ref} correspond to the modulation frequencies in I_{prj} . So the frequencies of different rows in a decoding band are similar but with different contributions. For every decoding band, we can always find the line with the largest ratio r , and it is named ‘typical line’. In Fig. 3, row 424 has the largest ratio and it is selected as the typical line of the decoding band.

For every decoding band, the typical line with the largest r is dominated by a single frequency, and therefore it generates the most accurate phase. By following the procedure of FTP [9], the phase map can be obtained by performing Fourier transform, band pass filtering, inverse Fourier transform and phase calculation with ‘arctan’ function. On the other hand, considering that most object regions are smooth and neighboring lines have similar phases, we assign the phase of the typical line to the entire decoding band. Note that at object boundaries, phases are not smooth, and the proposed hybrid scheme will further modify the depth with active spatial stereo as will be described in Section 2.2.4.

2.2.2. Local-smoothness-based phase unwrapping (LSPU)

Analyzing the reference pattern I_{ref} and the captured pattern I_{cap} in Fig. 1 reports the wrapped reference phase ϕ_w^{ref} and the wrapped calculated phase ϕ_w^{cap} , respectively. The parameters of the band pass filter are determined based on the primary frequencies of each decoding band. Based on the two phases, the deformed phase $\Delta\phi_w$ is

$$\Delta\phi_w = \phi_w^{cap} - \phi_w^{ref} \quad (3)$$

$\Delta\phi_w$ is determined by the depth of the measured object. Since ϕ_w^{cap} , ϕ_w^{ref} and $\Delta\phi_w$ are wrapped, phase unwrapping must be applied before converting phase to depth.

In general, there are two types of phase unwrapping methods: spatial approaches [23] and temporal ones [24]. Spatial phase unwrapping is based on the smooth assumption of objects, and thus compensate abrupt phase changes. In implementation, it follows specified paths, such as the scan lines [25, 26] in the simplest cases. In more advanced quality guided phase unwrapping (QGPU) [27–29], the path is determined according ‘phase quality’ such as phase derivative variance (PDV) [30], the maximum phase gradient (MPG) [31], and the amplitude of the reconstructed fringe (AMP) [32]. Based on phase quality, guiding strategies such as flooding fill [33] and group growing [34, 35] determine the exact unwrapping order of pixels by following the principle ‘reliable pixels first’. However, errors will also propagate and accumulate along the unwrapping path. Temporal or multi-frequency phase unwrapping is based on mathematical theory and is coupled with multi-frequency profilometry. In these methods, multiple phase maps are first obtained through phase analysis with varying modulation frequencies. After that, for each pixel, the unwrapped phase can be directly derived from its multiple wrapped phases by following a proper mathematical constraint [36–38]. This procedure is pixel-wise independent, so it avoids phase error propagation and enjoys better robustness. However, the use of multiple phase maps also requires multiple shots of patterns, which impairs real-time performance and innovates composite multi-frequency patterns [15–17].

We propose, in this part, a novel method named local-smoothness-based phase unwrapping (LSPU) based on the local-smoothness of the measured object. In LSPU, we first define phase

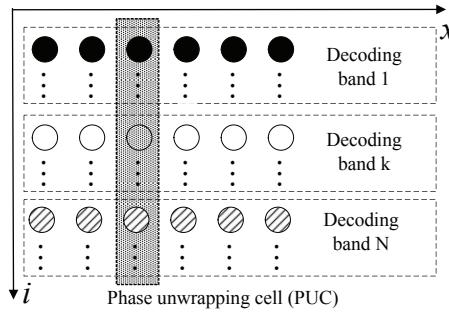


Fig. 4. Sketch of phase unwrapping cell (PUC).

unwrapping cell (PUC). As shown in Fig. 4, N neighboring decoding bands are combined together, and a single column is defined as a PUC. For the k^{th} decoding band, the primary frequency, the wrapped and the unwrapped phases are denoted as $f_{DB}(k)$, $\Delta\phi_w(k)$ and $\Delta\phi_{uw}(k)$, respectively.

For the k^{th} decoding band, phase unwrapping can be described as

$$\Delta\phi_{uw}(k) = \Delta\phi_w(k) + m(k)2\pi, \quad m(k) \in \mathbb{Z} \quad (4)$$

According to the local smoothness assumption, pixels in a PUC often share an identical depth value, so the unwrapped phases should satisfy

$$\frac{\Delta\phi_{uw}(1)}{2\pi f_{DB}(1)} = \dots = \frac{\Delta\phi_{uw}(k)}{2\pi f_{DB}(k)} = \dots = \frac{\Delta\phi_{uw}(N)}{2\pi f_{DB}(N)}, \quad (5)$$

In practice, the terms in Eq. (5) may not strictly equal to each other, and we rewrite it as an equivalent energy term of the PUC

$$E(\mathbf{m}, PUC) = \frac{2}{N(N-1)} \sum_{\tau=1}^{N-1} \sum_{\delta=\tau+1}^N \left[\frac{\Delta\phi_w(\tau) + m(\tau)2\pi}{2\pi f_{DB}(\tau)} - \frac{\Delta\phi_w(\delta) + m(\delta)2\pi}{2\pi f_{DB}(\delta)} \right]^2 \quad (6)$$

where \mathbf{m} is the vector of unwrapping coefficients $\mathbf{m} = [m(1), \dots, m(N)]$ of the PUC. The optimal vector \mathbf{m}^* minimizes the energy in Eq. (6), i.e.

$$\mathbf{m}^* = \arg \min_{\mathbf{m}} (E(\mathbf{m}, PUC)) \quad (7)$$

To solve this equation seems difficult, but the elements in \mathbf{m} are dependent. Equations (4) and (5) indicate that $m(i)$ and $m(j)$ in a PUC satisfy

$$m(j) = \text{round} \left[\frac{f_{DB}(j)}{f_{DB}(i)} m(i) + \frac{f_{DB}(j)\Delta\phi_w(i) - f_{DB}(i)\Delta\phi_w(j)}{2\pi f_{DB}(i)} \right] \quad (8)$$

When $f_{DB}(i) < f_{DB}(j)$, $m(j)$ can be uniquely determined with every given $m(i)$. Moreover, the unwrapped phase is in a limited range with only several valid candidates of \mathbf{m} . Supposing $f_{DB}(i)$ is the minimal frequency in PUC, we assign several candidates to $m(i)$, i.e., $m(i) \in [-M, M]$, and obtain the corresponding vectors of \mathbf{m} . Afterward, the vector \mathbf{m}^* minimizing the energy $E(\mathbf{m}, PUC)$ is obtained, with which phase unwrapping will be further accomplished with Eq. (4). The selection of M is based on two considerations. On the one hand, structured light is mostly for in-door use, and thus depth values, as well as unwrapped phases, are in a limited range. On the other hand, depth is reversely-related to the unwrapped phase, and small changes in m will lead to great differences in depth. Therefore, the valid candidates of m are actually in a small range, and M can be initialized.

2.2.3. Phase-depth conversion

With the unwrapped phase, depth can be computed. In a SL system as shown in Fig. 1, the deformed phase $\Delta\phi_{uw}$ uniquely corresponds to disparity and depth. To be specific, the unwrapped phase of (i, x) can be first converted to disparity and depth as follows

$$d_{FTP}(i, x) = \frac{\Delta\phi_{uw}(i, x)}{2\pi f_i^P} \quad (9)$$

$$Z_{FTP}(i, x) = \frac{bf_L Z_0}{f_L b + Z_0 d_{FTP}(i, x)} \quad (10)$$

where $\Delta\phi_{uw}$ is the unwrapped phase, and f_i^P is the primary frequency in of the i^{th} row. b is the base-line between the projector and the camera. f_L is the focal length. Z_0 is the depth of the reference plane.

2.2.4. Local-uniqueness-based spatial stereo

For a pixel (i, x) in I_{cap} , its correspondence in I_{ref} is $(i, x + d_{FTP}(i, x))$, where the disparity $d_{FTP}(i, x)$ is given in Eq. (9). If $d_{FTP}(i, x)$ is accurate, the intensities of $I_{cap}(i, x)$ and $I_{ref}(i, x + d_{FTP}(i, x))$ should be similar. In practice, it is more robust to evaluate the similarity between local patches, such as the Pearson correlation:

$$S(\mathbf{B}_{cap}, \mathbf{B}_{ref}) = \frac{\langle \mathbf{B}_{cap} - \bar{\mathbf{B}}_{cap}, \mathbf{B}_{ref} - \bar{\mathbf{B}}_{ref} \rangle}{|\mathbf{B}_{cap} - \bar{\mathbf{B}}_{cap}| |\mathbf{B}_{ref} - \bar{\mathbf{B}}_{ref}|} \quad (11)$$

\mathbf{B}_{cap} is a block in I_{cap} and its center is (i, x) . Similarly, \mathbf{B}_{ref} is a block in I_{ref} and its center is $(i, x + d_{FTP}(i, x))$. Pixels inside objects satisfy local-smoothness assumption well, so $d_{FTP}(i, x)$ is accurate. As a result, \mathbf{B}_{cap} and \mathbf{B}_{ref} are similar, and $S(\mathbf{B}_{cap}, \mathbf{B}_{ref})$ is large. In contrast, at depth discontinuities, $d_{FTP}(i, x)$ is incorrect and $S(\mathbf{B}_{cap}, \mathbf{B}_{ref})$ is small. Therefore, we use a threshold T_s to identify incorrect disparities, i.e., d_{FTP} is believed to be correct if its similarity $S(\mathbf{B}_{cap}, \mathbf{B}_{ref})$ exceeds T_s . T_s is a threshold of Pearson correlation, so its value is within the range of $[-1, 1]$. In general, as T_s increases, more incorrect disparities will be detected. However, a too large T_s will treat correct disparities as incorrect ones and will introduce unnecessary computation load. In our implementation, T_s is an empirical value that obtained by several trials. For the detected incorrect disparities, stereo matching is applied for modification. To be specific, two blocks $\mathbf{B}_{cap}(i, x)$ and $\mathbf{B}_{ref}(i, x + d_{stereo})$ are extracted from I_{cap} and I_{ref} , respectively. By testing a series of candidates $d_{stereo} \in [d_{min}, d_{max}]$, stereo matching finds the optimal disparity

$$d_{stereo}^*(i, x) = \arg \max_{d_{stereo}} \{S(\mathbf{B}_{cap}(i, x), \mathbf{B}_{ref}(i, x + d_{stereo}))\} \quad (12)$$

where the similarity measure S is the same with Eq. (11). With the detected optimal disparity, depth can be calculated as

$$Z_{stereo}(i, x) = \frac{f_L b Z_0}{f_L b + Z_0 d_{stereo}^*(i, x)} \quad (13)$$

In such an approach, the incorrect depth values are modified, which improves the depth accuracy.

Finally, with Z_{FTP} and Z_{stereo} , the final depth value is determined, i.e.,

$$Z_{hybrid}(i, x) = \begin{cases} Z_{FTP}(i, x) & d_{FTP}(i, x) \text{ is correct} \\ Z_{stereo}(i, x) & \text{otherwise} \end{cases} \quad (14)$$

Here, the correctness of Z_{FTP} has been determined with the similarity of $S(\mathbf{B}_{cap}, \mathbf{B}_{ref})$ as aforementioned.

We would also like to discuss the selection of the matching block **B**. A general concept in stereo matching is that, with a proper size, the matching block should be locally unique so as to report accurate disparity maps. For the proposed multi-frequency pattern, we set the block size adaptively. To be specific, the block covers three decoding bands, whose periods are T_{DB1} , T_{DB2} and T_{DB3} respectively. Since the periods are coprime, the minimal distance between two identical blocks is their product, i.e., $T_{DB1}T_{DB2}T_{DB3}$. This value is very large, and therefore the block is locally unique. If the block gets smaller and covers less decoding bands, the blocks uniqueness is weakened, which impairs disparity accuracy. In contrast, as the block size increases, more decoding bands are involved to achieve accurate matching, but the computation load also goes up and the boundary fatten effect becomes stronger [39]. Based on the above considerations, we find that it balances the accuracy and complexity when the block covers three decoding bands. In addition, we also use shiftable window [40] to improve the matching accuracy in practice.

3. Experimental results

In this section, we conduct simulations and experiments to demonstrate the performance of the proposed scheme. First, we theoretically verify the local-smoothness-based phase unwrapping (LSPU) with additive noise. After that, the proposed hybrid scheme is tested with 3dsMax, where practical factors such as defocus are considered. Finally, we build a real structured light system with practical distortions and noise to test the proposed scheme.

3.1. Theoretical verification

In simulation, I_{ref} and I_{cap} are 800×800 images defined as

$$\begin{cases} I_{ref}(i, x) = 120 + 60 \cos(2\pi f_i x) \\ I_{cap}(i, x) = 120 + 60 \cos\left[2\pi f_i x + \frac{f_0}{f_i} \Delta\phi(i, x)\right] \end{cases} \quad (15)$$

where $\Delta\phi$ is the additional deformed phase. $\Delta\phi$ is given in Eq. (16) and illustrated in Figs. 5(a) and 5(f). It jumps from 0 to 2.2π in the top, but varies smoothly in the bottom.

$$\Delta\phi(i, x) = \begin{cases} 2.2\pi u(x - 400) & i < 400 \\ 2.2\pi x / 800 & i \geq 400 \end{cases} \quad (16)$$

where $u(x)$ is the step function. In addition, the patterns are corrupted with uniformly distributed noise with expectation $\mu=0$ and variance $\sigma^2=33.33$.

We also reproduce the results of classic FTP [9], two-frequency composite pattern (TFCP) [19], and two-frequency phase-shifting-profilometry (TFPSP) [21] for comparison. Classic FTP, TFCP and the proposed method belong to FTP that only use a single pattern. TFPSP [21] uses two frequencies to modulate patterns, and uses three-step phase-shifting [8] to calculate wrapped phases for each frequency, which requires six patterns in total. The standard frequency $f_0=1/25$, and the three frequencies used in the proposed hybrid pattern are $1/11$, $1/19$ and $1/27$. For TFCP [19], the low and high frequencies are set to $1/125$ and $1/25$, respectively. In TFPSP [21], the two frequencies are $1/100$ and $1/53$, respectively. For classic FTP, the patterns are obtained by replacing all frequencies f_i in Eq. (15) with f_0 .

Figure 5 presents the recovered phases of $\Delta\phi$. Subfigures (b)-(e) show the recovered phases within the range of $[0, 2.2\pi]$. The corresponding phase errors are presented in subfigures (g)-(j). It is clear that classic FTP fails at the phase discontinuity, and the upper-right quarter of the phase map is incorrect. Among the single-shot methods, the proposed LSPU performs better than classic FTP and TFCP [19] with higher accuracy. As to TFPSP [21], it yields more accurate phases than the proposed one, but at the cost of six shots. In general, the proposed LSPU has quite good performance with a single shot.

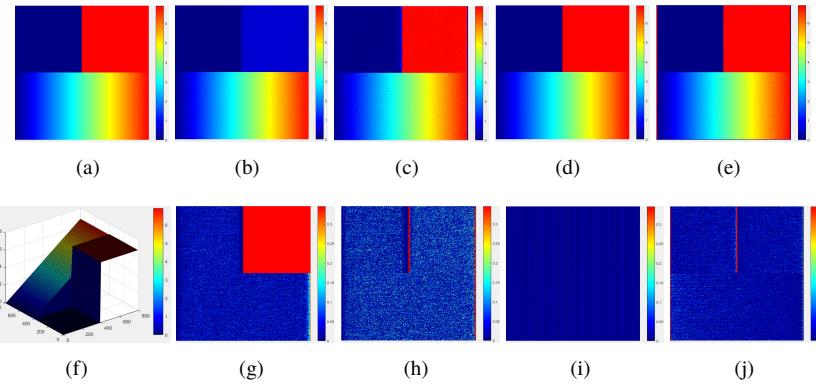


Fig. 5. Simulated phases obtained with different algorithms. (a)-(f): Color-view and mesh of the ground truth of $\Delta\phi$, respectively. (b): Classic FTP with scanline-based phase unwrapping. (c): TFCP [19]. (d): TFPSP [21]. (e): The proposed LSPU. (g)-(j): Error maps corresponding to (b)-(e), respectively.

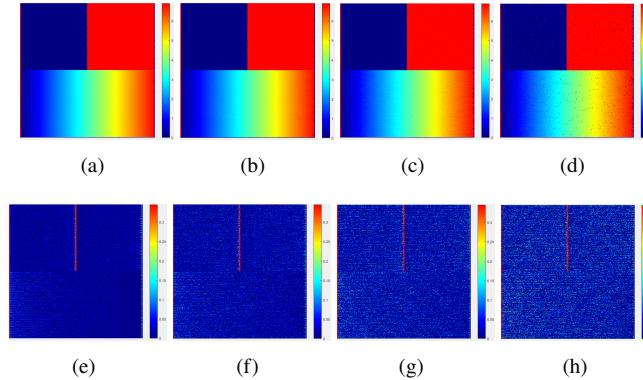


Fig. 6. Performance of the proposed LSPU with different noise. The upper row shows the unwrapped phases, and the lower row illustrates the phase errors. (a)-(e): $\mu=0$, $\sigma^2 = 8.33$. (b)-(f): $\mu=0$, $\sigma^2 = 33.33$. (c)-(g): $\mu=0$, $\sigma^2 = 75$. (d)-(h): $\mu=0$, $\sigma^2 = 133.33$.

We also test the proposed LSPU with different levels of noise, and the results are shown in Fig. 6. In general, errors increase and phase quality degrades with increasing noise. Fortunately, the proposed LSPU is independent among phase unwrapping cells (PUCs), and thus the errors are in isolated pixels or small blocks, which provides acceptable overall quality. Furthermore, with the successive spatial stereo, these errors can be further modified.

3.2. Experiment with 3dsMax

Using 3dsMax, we simulate a structured light system to test the proposed method. Important parameters are as follows: baseline $b=70\text{mm}$, focal length $f_L=35.572\text{mm}$. The resolutions of the projector and the camera are 800×1280 and 960×1280 , respectively. The camera focuses at the plane of 600mm . The field-of-view of the camera is smaller than that of the projector, and practical factors such as light attenuation, glisten rate and defocus blur are considered. Three 3D models: head, dragon and Buddha are utilized in our test. The models are located away from the reference plane with large depth discontinuities.

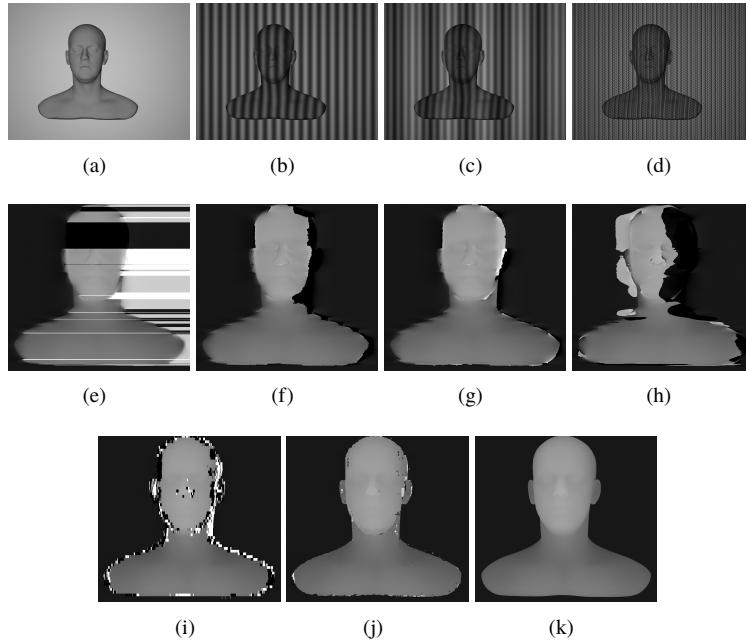


Fig. 7. Results of head. (a): 3D model. (b)-(d): Captured patterns of conventional fringe, TFCP [19] and the proposed method, respectively. (e)-(g): Results of SPU, A+F and P+F, respectively. (h): Result of TFCP [19]. (i)(j): Results of the proposed LSPU and the hybrid scheme, respectively. (k): Ground truth.

Several one-shot FTP-based approaches are used for comparison, including both single-frequency ones and dual-frequency ones. In the single-frequency pattern, the modulation frequency is 1/50. We achieve phase unwrapping via three approaches, i.e scanline based phase unwrapping, A+F (QGPU, reliability metric: amplitude [32], guiding strategy: flooding) and P+F (QGPU, reliability metric: PDV [30], guiding strategy: flooding). For the dual-frequency method, two-frequency composite pattern (TFCP) [19] is applied, and the modulation frequencies of the pattern are 1/15 and 1/75. For the proposed LSPU and the hybrid scheme, the three frequencies in the projected pattern are 1/11, 1/19 and 1/35, respectively. In addition, in simulation, the influence of noise and distortions is weak, and the patches are quite similar. As a result, we use a large threshold $T_s=0.9$ to detect incorrect disparities.

3.2.1. Qualitative results

The acquired depth maps of the tested scenes are shown in Figs. 7, 8 and 9. For a better illustration, the background regions are ignored and only the objects are presented.

SPU cannot well handle sharp depth changes, and errors propagate along scanlines. Therefore, in subfigures (a), we see ‘white bands’ starting from object boundaries to the end of the depth maps. As shown in subfigures (b) and (c), compared with SPU, QGPU, including A+F and P+F, can better handle phase error propagation, and therefore the incorrect depth bands are gone. However, instead of scanlines, phase errors propagate along the unwrapping path determined by QGPU. For example, the irregular blocks in Figs. 7(f) and 8(f) are filled with invalid depth values. In TFCP [19], the pattern is composed of two fringes modulated with a low-frequency and a high-frequency, based on which pixel-wise phase unwrapping is achieved without error propagation. However, for composed patterns, the spectrum, especially the low-frequency one, is

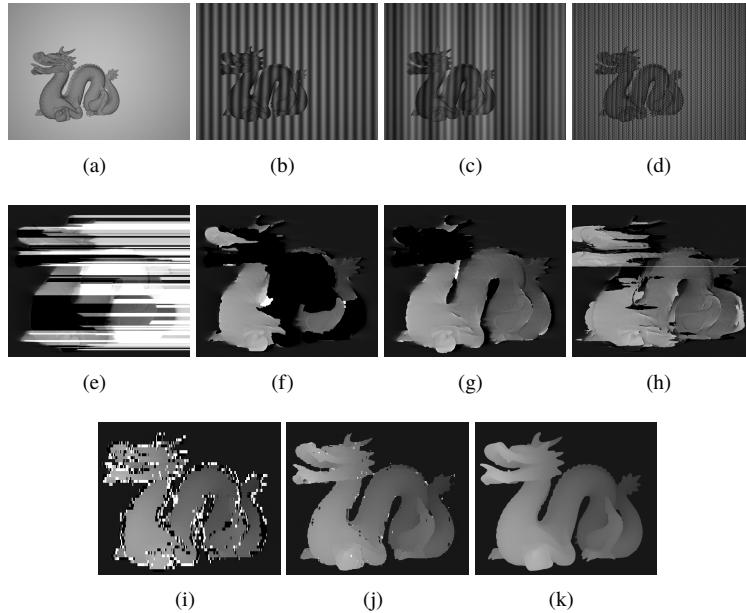


Fig. 8. Results of dragon. (a): 3D model. (b)-(d): Captured patterns of conventional fringe, TFCP [19] and the proposed method, respectively. (e)-(g): Results of SPU, A+F and P+F, respectively. (h): Result of TFCP [19]. (i)(j): Results of the proposed LSPU and the hybrid scheme, respectively. (k): Ground truth.

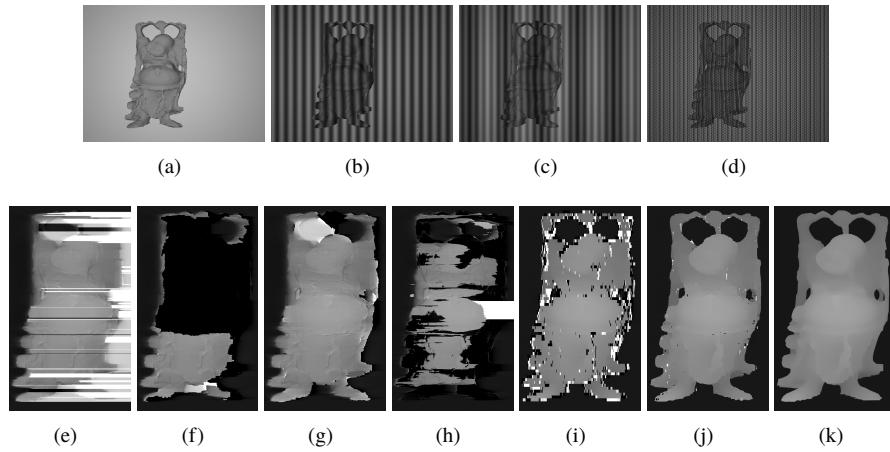


Fig. 9. Results of Buddha. (a): 3D model. (b)-(d): Captured patterns of conventional fringe, TFCP [19] and the proposed method, respectively. (e)-(g): Results of SPU, A+F and P+F, respectively. (h): Result of TFCP [19]. (i)(j): Results of the proposed LSPU and the hybrid scheme, respectively. (k): Ground truth.

limited to a narrow range and is often mixed with the DC component. As a result, the wrapped phases are not that accurate and the errors are amplified in phase unwrapping. Figure 7(h) is an example of such a failure. In the right part of the face, the foreground-background transition corresponds to the low frequency component. These frequency components are not accurately

extracted in the filtering, and the reported wrapped low-frequency phases are incorrect, which finally causes depth errors in Fig. 7(h).

As to the proposed scheme, LSPU achieves reliable phase unwrapping if the local smoothness assumption is well satisfied. Therefore, the results presented in subfigures (i) have valid depth values inside objects. In contrast, at object boundaries, LSPU fails and the reported depth values are incorrect. Fortunately, the errors can be further modified with the hybrid scheme, and the results are shown in subfigures (j), where both the inner parts and the boundaries have correct results.

3.2.2. Quantitative results

In addition to illustrating the depth maps, we also quantify the depth quality with mean absolute difference (MAD) and root mean squared error (RMSE) as presented in Tables. 1 and 2, respectively, where smaller values indicate higher accuracy.

In the experiment, the depth of the tested objects ranges from 500mm to 850mm. In the proposed method, the MAD is between 1.8 mm and 3.9 mm, and the relative error is about 0.5%, which demonstrates the accuracy of our method. Moreover, compared with other methods, the proposed hybrid scheme yields the smallest MAD/RMSE.

Table 1. MAD of the obtained depth values (mm)

	SPU	A+F	P+F	TFCP	LSPU	Hybrid
Buddha	46.06	64.90	14.80	47.30	18.85	3.04
Dragon	88.80	39.16	21.11	20.88	21.04	2.87
Head	50.96	8.56	5.70	22.25	11.13	1.88

Table 2. RMSE of the obtained depth values (mm)

	SPU	A+F	P+F	TFCP	LSPU	Hybrid
Buddha	85.86	96.94	38.10	80.10	49.57	14.31
Dragon	122.29	68.95	51.04	48.11	53.66	14.75
Head	94.81	27.16	16.61	53.80	38.94	10.98

3.3. Experiment with a real structured light system

We further test the proposed method with a real structured light system. The system consists of an Acer-H6510BD projector and a Crevis camera, which are vertically settled. Correspondingly, the fringe is modulated along the vertical axes. Important parameters of the system are as the following: baseline $b=200\text{mm}$, focal length $f_L=12.23\text{mm}$, and depth of the reference plane $Z_0=1330\text{mm}$. The resolutions of the projector and the camera are 1920×1080 and 1296×966 , respectively. Due to noise and distortions, pattern patches are not similar as they did in 3dsMax simulation, and a threshold of $T_s=0$ performs well (note that the similarity is in the range of -1 and 1). Two plaster statues, ‘fist’ and ‘portrait’ are used for depth measuring. They are placed on a low shelf in the experiment and thus are at the bottom of the captured images. In addition to the proposed method, classic scanline-based phase unwrapping (SPU) and two-frequency phase-shifting-profilometry (TFPSP) [20,21] are used for comparison. In [20,21], wrapped phases are calculated with phase shifting for each modulation frequency, and thus six patterns are needed to retrieve the unwrapped absolute phase. The modulation frequencies are determined based on two considerations. On the one hand, using a higher frequency can benefit band pass filtering in FFP and reports more accurate phases. On the other hand, both TFPSP and the proposed LSPU need multi-frequency phase unwrapping, and each of them needs several coprime and

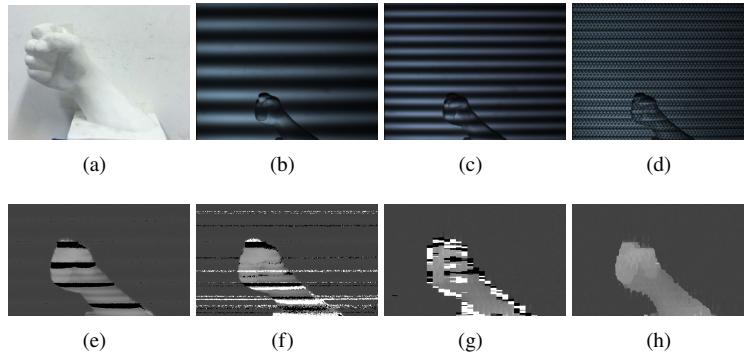


Fig. 10. Results of ‘fist’. (a): Color map. (b)(c): The second patterns of the low frequency and the high frequency fringes, respectively. (d): Captured pattern of the proposed scheme. (e) to (h): Depth maps of SPU, TFPSP [20, 21], the proposed LSPU and the proposed hybrid scheme, respectively.

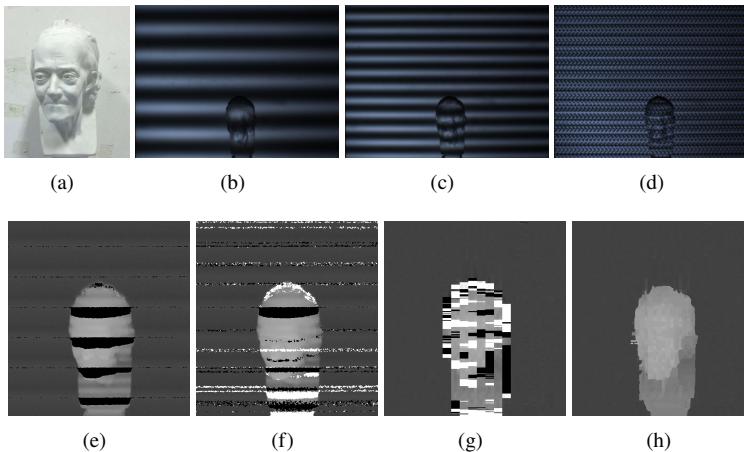


Fig. 11. Results of ‘portrait’. (a): Color map. (b)(c): The second patterns of the low frequency and the high frequency fringes, respectively. (d): Captured pattern of the proposed scheme. (e) to (h): Depth maps of SPU, TFPSP [20, 21], the proposed LSPU and the proposed hybrid scheme, respectively.

quite different modulation periods. Therefore, the frequencies are set as follows. The frequencies in the proposed hybrid pattern are 1/13, 1/8 and 1/5. As to TFPSP, the two frequencies are 1/15 and 1/8. Last but not least, the frequency of SPU is 1/15. As can be seen, the lowest frequencies are similar, but the proposed method also has higher frequency fringes because the pattern needs three frequencies.

The depth maps of ‘fist’ and ‘portrait’ are shown in Figs. 10 and 11. The objects are emphasized and the background plane is ignored. For scan-line based phase unwrapping, errors appear at the boundaries of fringe periods. In the depth maps of TFPSP [20, 21], there are many black/white points, which are with extremely small/large depth values. This is because the errors in wrapped phases are further amplified by referring to a look-up-table, and it leads to extremes in the unwrapped phases and the depth values. Figures 10(g) and 11(g) illustrate the results of the proposed LSPU. The depth-homogeneous regions generate correct depth results, while depth

Table 3. RMSE of the obtained depth values (cm)

	SPU	TFPSP	LSPU	Hybrid
fist	23.20	66.16	45.30	14.62
portrait	40.12	62.51	61.71	20.46
average	31.66	64.34	53.51	17.54

transitions suffer from mistakes. Fortunately, with the further steps of error-checking and stereo matching, the quality of boundaries are improved as shown in Figs. 10(h) and 11(h). The root mean squared error (RMSE) values are presented in Table 3, and the reported error indicates that the proposed scheme generates depth maps with higher accuracy.

3.4. Limitation of the proposed scheme

Although the proposed scheme can generate high quality depth maps, it also faces the problem of block effect. To be specific, trace of stripes can be observed in the resultant depth maps. The reason lies in the operation that we assign the wrapped phase of the typical line to the entire decoding band, including the non-typical lines. This in fact quantifies the phases, where minor phase changes of the non-typical lines are lost. As a result, depth cannot vary smoothly across decoding bands and block artifacts appear. In the future, we will remove these artifacts in two approaches: modifying the measured phase of non-typical lines based on their original wrapped phase, and smooth the phase between typical lines. We believe that both methods can remove the block effects and facilitate more accurate depth results.

4. Conclusion

In this paper, we proposed a novel profilometry scheme to acquire high quality depth maps with a single-shot of a multi-frequency monochromatic pattern. Based on this pattern, depth values are obtained in a hybrid manner. First, for the inner smooth regions, depth values are obtained through FTP-based phase analysis. Especially, motivated by local-smoothness assumption of objects, we proposed the LSPU algorithm to achieve phase unwrapping accurately and conveniently. Afterward, for object boundaries, where accurate phases are unavailable, the more robust spatial stereo matching was applied to obtain correct depth values.

The proposed scheme combined conventional phase-based SL and stereo matching, and therefore it utilizes the advantages while overcome the shortcomings of the conventional methods. Theoretical verification and experimental results demonstrated that, the proposed hybrid scheme performs robustly in generating depth maps for various scenes, including those with large depth discontinuities. Therefore, it can facilitate various depth-based applications, especially those applied in dynamic scenes.

Funding

National Natural Science Foundation of China (Nos. 61231010, 61702384, 61502357); Natural Science Foundation of Hubei Province (2017CFB348); Research Foundation of Hubei Educational Committee (Q20171106); Research Foundation for Young Scholars of WUST (2017xz008).