Published in final edited form as:

Opt Eng. 2019 June ; 58(6): . doi:10.1117/1.OE.58.6.060501.

Robust structured-light depth mapping via recursive decomposition of binary codes

Xiaohua Feng^a and Liang Gao^{a,b,*}

^aUniversity of Illinois at Urbana-Champaign, Department of Electrical and Computer Engineering, Urbana, Illinois, United States

^bUniversity of Illinois at Urbana-Champaign, Beckman Institute for Advanced Science and Technology, Urbana, Illinois, United States

Abstract

Structured-light depth cameras rely on projecting and resolving coded patterns on a three-dimensional scene with high contrast. The front-end optics of such depth cameras impose a fundamental restriction on the depth-sensing range and accuracy: the patterns only remain sharp within the depth of field jointly determined by the camera and projector. We present here a robust method to improve the depth-sensing range and accuracy for a structured-light depth camera without changing the underlying optical design. Moreover, it shows the unique advantage in macrophotography of highly light-scattering objects. We analyze the proposed method theoretically and validate it in experiments.

Keywords

optical metrology; three-dimensional modeling; camera; depth camera; structured light; range finding

1 Introduction

Structured light has been widely used in various imaging modalities, such as hyperspectral imaging,¹ optical diffraction tomography,² and three-dimensional (3-D) photography. In structured-light depth cameras,^{3–10} coded patterns are projected onto a scene to encode the correspondence between the projector and camera. Decoding the patterns correctly in the acquired images is of utmost importance to retrieve the correspondence, which is the key for accurate depth measurement. To obtain fine-grained correspondence, most codes rely on projecting high-frequency patterns. This is the case for phase-shifting profilometry and structured-light using temporal binary coding.¹¹ However, the finite band-width of the front-end optics and defocus generally attenuate high-frequency patterns. Moreover, when imaging highly light-scattering objects such as biological tissue in the macrophotography setting, light diffusion becomes evident and attenuates high-frequency patterns further.

*Address all correspondence to Liang Gao, gaol@illinois.edu.

These factors lead to a deteriorated contrast-to-noise ratio (CNR) of the coded images and cause erroneous decoding, limiting the depth-sensing range and accuracy.

Substantial efforts have been made to alleviate defocus effects while still maintaining excellent measurement accuracy. Microphase shifting¹² uses fringe patterns of high frequencies within a narrow band to improve its robustness against defocus and global illuminations. This method is based on the fact that fringe patterns of very close frequencies are almost defocus-invariant relative to each other. However, those high-frequency patterns are not viable under large defocus. The coded aperture projector¹³ uses a mask to modulate the projector's point spread function and performs deconvolution to recover high-frequency patterns. Yet, as defocus gets larger, deconvolution becomes increasingly ill posed because the high-frequency patterns gradually get washed out, which is a common issue for high-frequency structured-light methods.^{14–16} Other methods, such as using telecentric lens,¹⁷ works by increasing the depth of field of the camera/projector. None of these methods, however, deal well with large defocus and light diffusion.

To achieve robust depth mapping against defocus and light diffusion, it is therefore necessary to use only low-frequency patterns. The phase-shifting method with low-frequency gray-scale fringe patterns is appealing in this regard because it provides correspondence with subpixel accuracy and therefore superior accuracy than the structured light with temporal binary coding, where each column/row of the projector is encoded with a unique binary code and the correspondence is determined at an accuracy of 1 pixel. However, the gray-scale patterns are more prone to intensity noises and system nonlinearity than the binary codes.^{6,15} For instance, when intensity errors are severe, phase-shifting with high-frequency binary code (strip-edge structured light¹⁸) is preferred. Also, under defocus and light diffusion, the phase unwrapping errors¹⁵ are aggravated, producing depth measurement errors. Alternative methods using binary codes that are robust against defocus and light diffusion while still offering good measurement accuracy are therefore highly desirable.

We present here a robust method to achieve this goal for a binary-code based structured-light camera. We show that any binary code can be decomposed into and reconstructed by a pair of twice lower frequency codes and this process can be applied recursively. Therefore, the high-frequency binary code can be preserved by first decomposing it into multiple lower-frequency codes to enable its passage through front-end optics and light diffusion and finally numerically reconstructed from its decompositions. This allows fine-grained correspondence to be preserved despite high-frequency codes themselves being not viable. This simple yet highly effective method substantially extends the depth-sensing range and is particularly robust against light diffusion. It offers an accuracy close to the popular gray-code phase-shifting^{14,15} (GCPS) method, but free from the phase unwrapping error, which is hard to completely eliminate for highly light-diffusive objects as demonstrated in the experiments. Compared with binary coding without decomposing high-frequency codes, the proposed method improves the depth measurement accuracy substantially for both defocused and light-diffusive objects.

2 Method

We define, for arbitrary binary codes, the period of the code as the interval between two successive rising or falling edge of the code. The code's period could therefore be spatially varying, as shown in Fig. 1(a). The proposed method generates two lower-frequency codes by alternative flipping: using the rising edge as the beginning of a period and keeping one period intact while the next period flipped. The second code is similarly generated by using the falling edge as the start of a period and then applying alternative flipping. The two lower-frequency codes reconstruct the original signal perfectly via an efficient exclusive OR (XOR) operation. By applying this method recursively, one can decompose any high-frequency codes into a tree of much lower frequencies. We refer here the decomposition level as the number of recursions in the decomposition of one code. For instance, a level 2 decomposition corresponds to decomposing the highest frequency codes twice that results in four codes with 4 times lower frequency. The decomposition level is one less for the second-highest frequency in the original code set and this goes on until the level decreases to 0.

To analyze the proposed approach quantitatively, we model the projection and imaging of the codes in Fourier domain

$$P_{\text{img}} = [O(k) * C(k)P_j(k)]P_c(k) + U, \quad (1)$$

where $*$ denotes convolution operation, U is the measurement noise, and $O(k)$ and $C(k)$ are the Fourier transform of the object and coded pattern, respectively. $P_j(k)$ and $P_c(k)$ are the depth-dependent incoherent optical transfer function (iOTF) of the projector and camera, respectively. The iOTF of both the projector and camera shows a low-pass filtering characteristic. As indicated in Eq. (1), their effects on the coded pattern are multiplicative. It thus shows enhanced attenuation of high frequencies and induces intersymbol interferences (ISI) (smoothing of the edges between the binary codes) in the coded patterns. During the decoding process, the pattern images are locally thresholded against $(Ap + An)/2$ to recover the binary pattern, where Ap and An are the magnitudes of the direct and inverse images, respectively. Such binary thresholding is sensitive to noises and ISI. Borrowing from communication theory, the performance of recovering the codes from noisy measurement could be characterized by bit error rate (BER): the ratio of correctly recovered binary codes over the total number of projected codes. Assuming the noises U is Gaussian with zero mean and standard deviation of σ , the relationship between the BER and CNR is¹⁹

$$\text{BER} = \mathcal{O}(-\text{CNR}), \quad (2)$$

where $\mathcal{O}$ is the cumulative distribution function of Gaussian process, and CNR is the contrast-to-noise ratio of the coded image

$$\text{CNR} = (Ap - An)/\sigma. \quad (3)$$

The BER is inversely related to CNR and gets larger with ISI. By decomposing the high-frequency codes into lower frequencies, we can largely bypass the low-pass filtering effect on the codes. As a result, the BER will be significantly reduced by reconstructing the high-frequency code from these low-frequency decompositions, which shows better CNR. This improves correspondence accuracy and therefore enables more accurate depth measurement. As defocus and light diffusion are major contributors to the low-pass filtering effects, this effectively extends the depth measurement range and makes it robust against light diffusion.

To validate the analysis, a one-dimensional (1-D) numerical simulation was conducted. Given a uniform object, the combined iOTF of the projector and camera was modeled as a second-order Bessel low-pass filter, which corresponds to an optimistic case of imaging an in-focus object. The cut-off frequency of the filter was set at 10 lp/mm. A binary code with a spatial frequency of 20 lp/mm and 50% duty cycle was generated by gray code scheme. The binary code was decomposed into two lower-frequency codes using the proposed method, as shown in Fig. 1(b). After passing through the filter, a zero mean Gaussian noise with a standard deviation (std.) of 0.05 was added to simulate the received signal by the camera. The signals were then thresholded against the mean of the signal to recover the binary code. The received signals and recovered high-frequency codes, with and without decomposition, are compared in Figs. 1(c) and 1(d), respectively. Decomposed into lower frequencies, the received signals exhibit a higher CNR [top black trace in Fig. 1(c)] than the high-frequency code itself (bottom red). By reconstructing the high-frequency code with these lower-frequency codes, the proposed method shows substantially better recovery [top black trace in Fig. 1(d)] than the case without applying decomposition (bottom red): the BER is about 7% and 20%, respectively. For defocus that shows a strong low-pass filtering effect, simulated by fourth-order Bessel filter with the same cut-off frequency, the BER is close to random guess at 50% without decomposition. By contrast, the proposed decomposition-reconstruction method achieves a BER about 10%, which is an improvement over fourfolds. Recursive decomposition is necessary when the codes show frequency way beyond the cut-off frequency. This happens when the object is heavily defocused.

3 Results

We demonstrated the proposed decomposition-reconstruction method in structured-light with temporal encoding using gray codes and compared it with the same method but without decomposition (referred to as conventional method hereafter). To make the decoding more robust for inhomogeneous surfaces, the inverse pattern of each code was also projected, as proposed in Ref. 11. With the inverse-coded images, their differential image improves the CNR of each individual code by a factor of 1.4, and the threshold value for each code recovery was calculated as the average of complementarily coded images. The experimental setup for the depth camera is shown in Fig. 2. An LCD projector (NEC PA550W, resolution 1280×800) illuminates the object through a relay lens pair at an oblique angle ($\sim 30^\circ$). The image is captured by a monochrome CMOS camera (1920×1200 resolution, Blackfly 2.0MP, Point Grey Inc.) with a 35-mm focal length CCTV camera lens. The system working distance is about 100 mm with a field of view of $\sim 22 \times 14 \text{ mm}^2$ at the focal plane. The projector and camera show a joint depth of field around 4 mm for a circle of confusion of 0.05 mm. They are coordinated by a computer and calibrated using a checker-board pattern

to extract both the intrinsic and extrinsic parameters (see Ref. 11 for more details). The depth was calculated using the triangulation method based on ray-plane intersection.

The accuracy of the proposed method over an extended depth range was characterized using a flat plane as the object and calculated as the std. from the fitted plane (via moving least squares²⁰). Figure 3(a) shows the measurement accuracy using different decomposition levels for both in-focus and defocused (10 mm) plane. The accuracy improvement is not substantial for the in-focused plane. However, for the defocused plane, the accuracy improvement becomes evident: at level 4 decomposition, the measurement accuracy is more than 5 times better. As the decomposition is most effective when the decomposed frequency falls into the pass-band of the iOTF, it is not surprising to see such sharp transition in accuracy improvement. The depth accuracy using level 4 decomposition and conventional method are compared at four different depths: one in focus, and the other three being defocused by about 6, 10, and 14 mm, respectively. The results are shown in Fig. 3(b). Unlike the conventional method that shows quick degradation of measurement accuracy with defocus, the proposed method consistently achieved a better depth accuracy across all the depth range and offered an accuracy around 100 μm even when the scene was defocused by 14 mm.

The improved accuracy is crucial to map objects that show small depth variations. To illustrate, a 10-cent coin was imaged by the depth camera. Figures 4(a)–4(c) shows, from left to right, the bright-field photography of the coin at the focus, the depth maps of the proposed and conventional methods, respectively. The proposed method rendered well the contour of the head, whereas the conventional method suffered from some strip artifacts on the right, where the obliquely incident-coded pattern had undergone slight defocus. Even when the coin was defocused by 12 mm [photograph in Fig. 4(d)], the proposed method still resolved the depth variation [Fig. 4(e)] well over the coin, whereas the conventional method, shown in Fig. 4(f), was overwhelmed by artifacts. The blurring of the depth map caused by the defocus is noticeable when comparing Figs. 4(b) and 4(e). As the proposed method only converted the projected codes to lower-frequency range but not the object's high-frequency components, it cannot improve spatial resolution of the depth map when the object is defocused. Nonetheless, it is feasible to use coded aperture camera²¹ to overcome this issue, which is beyond the scope of current work.

Lastly, we demonstrated the proposed method in measuring depths of highly light-diffusive objects and compared it with the GCPS method as well, where care was taken to align the phase-wrapping edge with that of the gray code.¹⁴ To reduce the phase-unwrapping errors, the unwrapped correspondence was corrected by a median filter (9 pixels in size) before depth calculation. The fringe periods of GCPS were chosen to be 128 and 256 pixels owing to their robustness against light diffusion. Smaller fringe periods such as 64 produce noticeable stripping artifacts that are absent for the nondiffusive object. The object is a 3-D printed rabbit [Fig. 5(a) inset] that shows similar light-scattering properties as biological tissues and was placed inside the focal region so that only light-diffusion is present. The bright-field image of the rabbit belly region is shown in Fig. 5(a) and a representative image of a binary pattern is given in Fig. 5(b), which showed significantly reduced contrast than that of the paper object [inset of Fig. 5(b)] due to light-diffusion. Such low-pass filtering on

the pattern is independent of the front-end optics of the depth camera and cannot be physically eliminated. The depth maps produced by the conventional and proposed methods with level 5 decomposition are shown in Figs. 5(c) and 5(d), respectively. The proposed method rendered a smooth depth map, whereas the conventional method, as in previous cases, suffered from larger errors as evidenced by the strip artifacts. The depth maps produced by GCPS with fringe periods of 128 and 256 pixels are shown in Figs. 5(e) and 5(f), respectively. It is noteworthy that GCPS suffered from depth errors caused by the phase-unwrapping error in both cases, though GCPS with fringe period of 256 pixels showed fewer unwrapping errors because the lower frequencies were inherently more robust against diffusion. To eliminate these phase-unwrapping errors, a much larger filter is needed,^{15,22} leading to a compromised spatial resolution. In contrast, the proposed method is free from the phase-unwrapping errors and shows an accuracy close to the GCPS in other regions.

The proposed method is hence a simple and highly effective improvement for structured-light depth camera using temporal binary coding. It enables more robust and accurate depth measurement over a larger depth range. The decomposition and reconstruction of the codes is computationally light and runs in real time. One unique advantage of the proposed method is its effectiveness for depth mapping of highly light-scattering objects such as biological tissues in macrophotography mode, where light diffusion severely attenuates high-frequency patterns. Light diffusion is independent of the front-end optics, and as a result, most existing methods that extend the depth-sensing range are not applicable. A drawback of the proposed approach is that it generates more coded patterns that need to be projected, reducing the frame rate of depth map acquisition and thus hampering its application in imaging dynamics. With more images, it is also tempting to use temporal averaging to increase the CNR. However, temporal averaging is much less efficient once the coded patterns are highly blurred by defocus or light scattering. The lost high-frequency pattern is hard to recover by simple filtering process, either in the temporal or spatial domain. In the proposed method, the high-frequency coded pattern is represented by lower-frequency codes and then reconstructed algorithmically. It thus enables efficient projection and capturing through the front-end optics and highly light-scattering objects. Since the accuracy of the depth camera depends heavily on the high-frequency codes, the capability in preserving them in the proposed method is the trick for its better depth-sensing accuracy.

4 Conclusion

We proposed a recursive decomposition method for a structured-light depth camera using binary codes for highly accurate depth measurement over an extended depth range. It improves the measurement accuracy over 5 times for significantly defocused objects, and it shows the unique advantage in robust depth sensing of highly light-scattering object. It could find applications in fields such as medical imaging and robotics.

Acknowledgments

This work was supported in part by NSF CBET 1652150 and NIH R35GM128761, and we thank Ricoh Innovations for the donation of the projector.

References

1. Ma C et al., "Digital micro-mirror device based multispectral imaging using compressed Fourier spectrum," *Opt. Commun* 426, 348–358 (2018).
2. Lee K et al., "Time-multiplexed structured illumination using a DMD for optical diffraction tomography," *Opt. Lett* 42(5), 999–1002 (2017). [PubMed: 28248352]
3. Tang S et al., "Three-dimensional surface reconstruction via a robust binary shape-coded structured light method," *Opt. Eng* 56(1), 014102 (2017).
4. Zhang S and Yau S-T, "Three-dimensional shape measurement using a structured light system with dual cameras," *Opt. Eng* 47(1), 013604 (2008).
5. Pan B et al., "Phase error analysis and compensation for nonsinusoidal waveforms in phase-shifting digital fringe projection profilometry," *Opt. Lett* 34(4), 416–418 (2009). [PubMed: 19373326]
6. Zuo C et al., "Phase shifting algorithms for fringe projection profilometry: a review," *Opt. Lasers Eng* 109, 23–59 (2018).
7. Kim K et al., "Real-time visualization of 3-D dynamic microscopic objects using optical diffraction tomography," *Opt. Express* 21(26), 32269–32278 (2013). [PubMed: 24514820]
8. Feng S et al., "Motion-oriented high speed 3-D measurements by binocular fringe projection using binary aperiodic patterns," *Opt. Express* 25(2), 540–559 (2017). [PubMed: 28157945]
9. Yin W et al., "High-speed 3D shape measurement using the optimized composite fringe patterns and stereo-assisted structured light system," *Opt. Express* 27(3), 2411–2431 (2019). [PubMed: 30732279]
10. Geng J, "Structured-light 3D surface imaging: a tutorial," *Adv. Opt. Photonics* 3(2), 128–160 (2011).
11. Moreno D and Taubin G, "Simple, accurate, and robust projector-camera calibration," in *Second Int. Conf. 3D Imaging, Model., Process., Visualization and Transm.*, pp. 464–471 (2012).
12. Gupta M and Nayar SK, "Micro phase shifting," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 813–820 (2012).
13. Kawasaki H. Active one-shot scan for wide depth range using a light field projector based on coded aperture. *IEEE Int. Conf. Comput. Vision (ICCV)*; 2015. 3568–3576.
14. Sansoni G, Carocci M, and Rodella R, "Three-dimensional vision based on a combination of gray-code and phase-shift light projection: analysis and compensation of the systematic errors," *Appl. Opt* 38(31), 6565–6573 (1999). [PubMed: 18324191]
15. Zheng D et al., "Phase-shifting profilometry combined with gray-code patterns projection: unwrapping error removal by an adaptive median filter," *Opt. Express* 25(5), 4700–4713 (2017). [PubMed: 28380741]
16. Zhou P, Zhu J, and Jing H, "Optical 3-D surface reconstruction with color binary speckle pattern encoding," *Opt. Express* 26(3), 3452–3465 (2018). [PubMed: 29401873]
17. Li D, Liu C, and Tian J, "Telecentric 3D profilometry based on phase-shifting fringe projection," *Opt. Express* 22(26), 31826–31835 (2014). [PubMed: 25607150]
18. Song Z, Chung R, and Zhang X-T, "an accurate and robust strip-edge-based structured light means for Shiny surface micromasurement in 3-D," *IEEE Trans. Ind. Electron* 60(3), 1023–1032 (2013).
19. Barry JR, Lee EA, and Messerschmitt DG, *Digital Communication*, 3rd edn., Springer, New York (2004).
20. Fleishman S, Cohen-Or D, and Silva CT, "Robust moving least-squares fitting with sharp features," in *ACM SIGGRAPH*, pp. 544–552 (2005).
21. Levin A et al., "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graphics* 26(3), 70 (2007).
22. An Y, Hyun J-S, and Zhang S, "Pixel-wise absolute phase unwrapping using geometric constraints of structured light system," *Opt. Express* 24(16), 18445–18459 (2016). [PubMed: 27505808]

Fig. 1.

(a) Decomposition and reconstruction of high-frequency binary codes. HF, high frequency; LF, low frequency. (b)–(d) One-dimensional simulation of recursive decomposition of binary codes. (b) The high-frequency (top) code is decomposed into two lower-frequency codes. (c) Simulated signals received by the camera. The top black is associated with the decomposed lower-frequency code, and lower red is associated with the high-frequency code. (d) Recovery of the high-frequency code. The top black is the result of the proposed method, the center blue is the ground-truth code, and the bottom red is the recovery using the conventional method (no decomposition).

Fig. 2.
Experimental setup for the depth camera.

Fig. 3. Quantitative comparison of depth accuracy. (a) Depth mapping accuracy versus different decomposition levels for in-focus (dashed line) and defocused scene (solid line). (b) Depth mapping accuracy versus different defocus distance for the conventional (dashed line) and level-4 decomposition method (solid line).

Fig. 4.

(a) and (d) Photography of the coin placed in focus and defocused. (b) and (e) Depth map of proposed method. (c) and (f) Depth map of conventional method.

Fig. 5.

- (a) Bright-field photography of the belly region of the 3-D printed rabbit (shown in inset).
- (b) The image of one representative binary pattern; inset is the pattern image on a white paper.
- (c) and (d) Depth map of conventional and proposed methods, respectively.
- (d) and (f) Depth map produced by GCPS with fringe period of 128 and 256, respectively.