Optics Letters

Single-shot absolute 3D shape measurement with deep-learning-based color fringe projection profilometry

Jiaming Qian,[1,2] Shijie Feng,[1,2] Yixuan Li,[1,2] Tianyang Tao,[1,2] Jing Han,[2] Qian Chen,[2,3] and Chao Zuo[1,2,*]

[1]Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China
[2]Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China
[3]e-mail: chenqian@njust.edu.cn
*Corresponding author: zuochao@njust.edu.cn

Received 23 January 2020; revised 23 February 2020; accepted 27 February 2020; posted 28 February 2020 (Doc. ID 388994); published 20 March 2020

Recovering the high-resolution three-dimensional (3D) surface of an object from a single frame image has been the ultimate goal long pursued in fringe projection profilometry (FPP). The color fringe projection method is one of the technologies with the most potential towards such a goal due to its three-channel multiplexing properties. However, the associated color imbalance, crosstalk problems, and compromised coding strategy remain major obstacles to overcome. Inspired by recent successes of deep learning for FPP, we propose a single-shot absolute 3D shape measurement with deep-learning-based color FPP. Through "learning" on extensive data sets, the properly trained neural network can "predict" the high-resolution, motion-artifact-free, crosstalk-free absolute phase directly from one single color fringe image. Compared with the traditional approach, our method allows for more accurate phase retrieval and more robust phase unwrapping. Experimental results demonstrate that the proposed approach can provide high-accuracy single-frame absolute 3D shape measurement for complicated objects. © 2020 Optical Society of America

https://doi.org/10.1364/OL.388994

Fringe projection profilometry (FPP) [1] is one of the most widely used three-dimensional (3D) measurement techniques because of its simple hardware facilities, flexible implementation, and high measurement precision. Recently, with the increasing demands of 3D information acquisition in various applications, such as online quality inspection and rapid reverse engineering, high-speed 3D shape measurement technologies based on FPP are becoming more and more popular and essential [2–5].

To achieve high-speed 3D imaging, it is necessary to improve the measurement efficiency, i.e., to reduce the number of patterns required per reconstruction [6,7]. Ideally, the absolute 3D surface of an object should be recovered from only a single image. The color-coded projection methods [8–13] have great advantages in dynamic scene measurement, since it can encode three independent patterns in its red, blue, and green channels. However, few of them can be used for high-accuracy measurements of complex objects. On one hand, to obtain high-accuracy phase information, the phase-shifting (PS) method [14] with high measurement resolution is preferred. However, PS requires at least three fringe images, which occupy all channels of an RGB image, so that only the spatial phase unwrapping method [15] (which will become vulnerable when encountering isolated or unjoined phase) can remove the phase ambiguity [9]. On the other hand, to achieve robust phase unwrapping, the strategies of combining fringe patterns with the fringe-order-coded information (gray-code patterns or stair intensities) or combining multi-frequency fringe images are adopted [10–13]. The former still cannot unwrap the phase stably due to the difficulty in identifying the edge of gray-code patterns or the sensitivity of intensity to ambient light noise and large surface reflectivity variations of the objects [10,11]. The latter can recover the absolute phase by the three-fringe number selection method [16] but compromises the accuracy due to the use of the Fourier transform (FT) method [15,17] (well known for its single-shot nature but with limited quality around discontinuities and isolated areas in the phase map) [12,13]. In addition, there are some inherent defects in color-coded projection methods, such as chromatic aberration and crosstalk between color channels, which will affect the quality of phase calculations. Some researchers propose some pre-processing methods [9,12,18] to compensate for these factors, but only to some extent reduce their impact.

Recently, the successes of deep learning in FPP [19–23] have provided new opportunities for color-coded projection technologies. In this Letter, we propose a deep-learning-based color fringe projection profilometry. With the help of a properly trained neural network, a high-accuracy absolute

phase can be "learned" from a color-coded image without any pre/post-processing or phase error compensation.

The flowchart of our method is shown in Fig. 1. Notably, instead of adopting an end-to-end network structure directly linking fringe images to the absolute phase, our network predicts a high-quality wrapped phase map along with a "coarse" absolute phase from the three-channel fringe patterns of different frequencies. There are three basic considerations for this strategy: (1) due to the inherent depth ambiguities in FPP, one single fringe pattern is generally insufficient to uniquely determine the fringe order in the presence of isolated surfaces or surface discontinuities [5]. Thus, to guarantee absolute 3D shape measurement independent of any assumptions and prior knowledge, multi-frequency fringe patterns and temporal phase unwrapping (TPU) [5] are necessary. (2) Without additional auxiliary information, a simple input–output network structure (linking fringe images to the absolute phase directly) usually produces an estimate with compromised accuracy; especially, the measured surface contains sharp edges, discontinuities, or large surface reflectivity variations [23]. Based on this consideration, we incorporate the idea of TPU into our network to determine the fringe order of the wrapped phase with the predicted "coarse" absolute phase on a pixel-by-pixel basis. (3) For the unwrapped phase map, our deep neural network is trained to predict the intermediate numerator and denominator of the arctangent function, to bypass the difficulties associated with reproducing abrupt 2π phase wraps and thus obtain a high-quality phase estimate [19].

Our pattern design is similar to that of [13], which encodes three fringe patterns of different wavelengths λ_R, λ_G, and λ_B into the red, green, and blue channels of one color image. The color-coded image is projected by a projector, modulated by the object, and finally captured by a color camera. The captured image can be represented as I^c_{RGB}, and the gray images of its three channels can be expressed as I^c_R, I^c_G, and I^c_B, respectively. Since three fringe images with different wavelengths are used, the phase unwrapping can be achieved by the projection distance minimization (PDM) approach [24], which obtains the pixel-wise qualified fringe order according to wrapped phase distribution in three fringe images. Because the wrapped phase is encoded in the fringe pattern and the deep learning can organically synthesize the spatial and temporal information [23], we can expect that the properly trained neural network can directly obtain the absolute phase from the fringe patterns of three frequencies. However, only the low-accuracy absolute phase can be predicted. To obtain high-quality phase information, the physical model of the traditional PS method is considered. For the N-step PS algorithm, the N fringe patterns can be expressed as

$$I_n = A + B\cos(\Phi + 2\pi n/N), \tag{1}$$

where I_n represents the $(n+1)$th fringe image, $n \in [0,\ N-1]$, A is the average intensity, B is the amplitude, Φ is the absolute phase, and $2\pi/N$ is the phase shift. The wrapped phase ϕ can be obtained by [14]

$$\phi = \arctan\frac{M}{D} = \arctan\frac{\sum_{n=0}^{N-1} I_n \sin(2\pi n/N)}{\sum_{n=0}^{N-1} I_n \cos(2\pi n/N)}, \tag{2}$$

where M and D represent the numerator and denominator of arctangent function, respectively. If the network predicts M and D, the phase information can be obtained by Eq. (2). Such an operation provides higher phase accuracy than the network structure of directly linking fringe pattern to phase

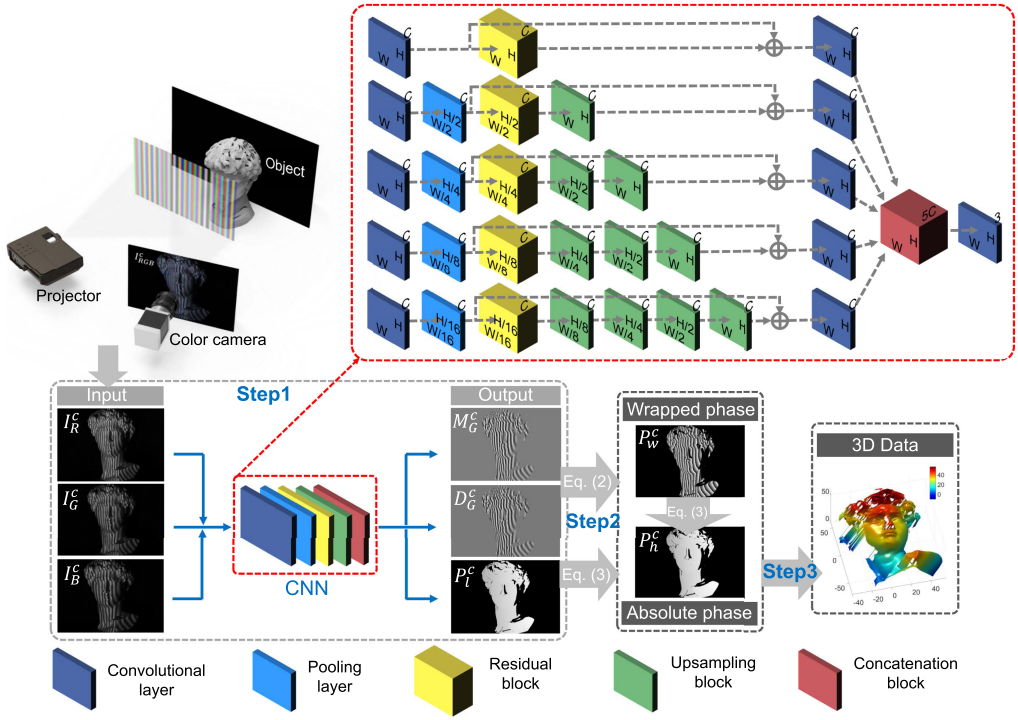

Fig. 1. Flowchart of our method. Step 1: input three gray fringe images I^c_R, I^c_G, and I^c_B of the color image channels, and output the numerator and denominator terms M^c_G and D^c_G and the low-accuracy absolute phase P^c_l in the green channel. Step 2: obtain the high-accuracy absolute phase P^c_h by Eqs. (2) and (3). Step 3: reconstruct the 3D information by the calibration parameters.

[19]. Therefore, I_R^c, I_G^c, and I_B^c are fed into the constructed neural network, and the outputs are the numerator M_G^c and denominator D_G^c, and a rough absolute phase map P_l^c (whose error is between $-\pi$ and π) in the green channel. In addition, to enable the network to resist crosstalk and chromatic aberration problems, the data without these factors are used as labels to train our network. After the network predicts M_G^c, D_G^c, and P_l^c, the wrapped phase P_w^c of wavelength λ_G can be obtained by Eq. (2). Then the high-quality absolute phase P_h^c can be acquired by

$$P_h^c = P_w^c + 2\pi \,\mathrm{Round}[(P_l^c - P_w^c)/2/\pi], \tag{3}$$

where Round represents the rounding function. Finally, the 3D reconstruction can be performed by utilizing the pre-calibrated parameters [25] of the system.

The upper right part of Fig. 1 reveals the internal structure of our network, which is a five-path convolutional neural network (CNN) and is more powerful than our previous CNNs [19,21–23]. For each convolutional layer, the kernel size is 3×3 with convolution stride one, and the output is a 3D tensor of shape (H, W, C), where (H, W) is the size of the input pattern, and C represents the number of filters used in each convolutional layer ($C = 64$). In the first path of the CNN, the inputs are processed by a convolutional layer, followed by four residual blocks and another convolutional layer. Also, implementing shortcuts between residual blocks contributes to the convolution stability [22]. In the other four paths, the data are down-sampled by the max-pooling layers by two times, four times, eight times, and 16 times, respectively, for better feature extraction, and then up-sampled by the up-sampling blocks to match the original size. The outputs of all paths are concatenated into a tensor with five channels. Finally, three channels are generated in the last convolution layer.

We construct an FPP system, which includes a LightCrafter 4500 (912×1140 resolution) and a Basler acA640-750uc camera (640×480 resolution), to test the effectiveness of our method. The wavelengths of three color channels are selected as $\lambda_R = 9$, $\lambda_G = 11$, and $\lambda_B = 13$, which provide unambiguous 3D reconstructions for the whole projection range. During the training session, 600 groups of images are captured. Each group contains a color-coded fringe pattern [Fig. 2(a)], the three channels of which are used as inputs of the network, as well as the 12-step PS fringe patterns of three frequencies [Figs. 2(b)–2(d)], which are used for the calculation of the ground-truth data (the three frequencies of the latter are consistent with those of the three channels of the former). To avoid crosstalk and chromatic aberration problems at the source, when collecting the PS images, the green non-composite fringe patterns are projected, and only the green channels of the captured images are utilized for the label. The numerator and denominator terms are calculated by Eq. (2). The absolute phase is obtained based on PDM. When training, 450 groups of data are used for training, and 150 groups are used for verification.

To verify the superiority of our method over traditional methods, we measure two scenarios with our method and the method of [13]. The measured results are shown in Fig. 3. The shadow-noised regions, with B smaller than a pre-defined threshold, are excluded in the subsequent processing [14]. Figures 3(b) and 3(f) are results of the method of [13], from which we can see that despite the crosstalk compensation in advance, the resolution of the details is limited due to the use of FT method. Also, the phase errors lead to the phase unwrapping errors at the edges. In contrast to the traditional method, our method produces more accurate absolute reconstruction, whose quality is even comparable to that obtained by PS and PDM methods.

Next, we apply our approach to the dynamic 360-deg 3D model reconstruction. A metallic workpiece is continually rotated for one cycle on a mechanical stage and measured with our approach (note that there are no objects of such type in our training and verification set). The measured results from two

Fig. 2. Group of data sets. (a) Color image. (b)–(d) 12-step PS images of λ_R, λ_G, and λ_B.

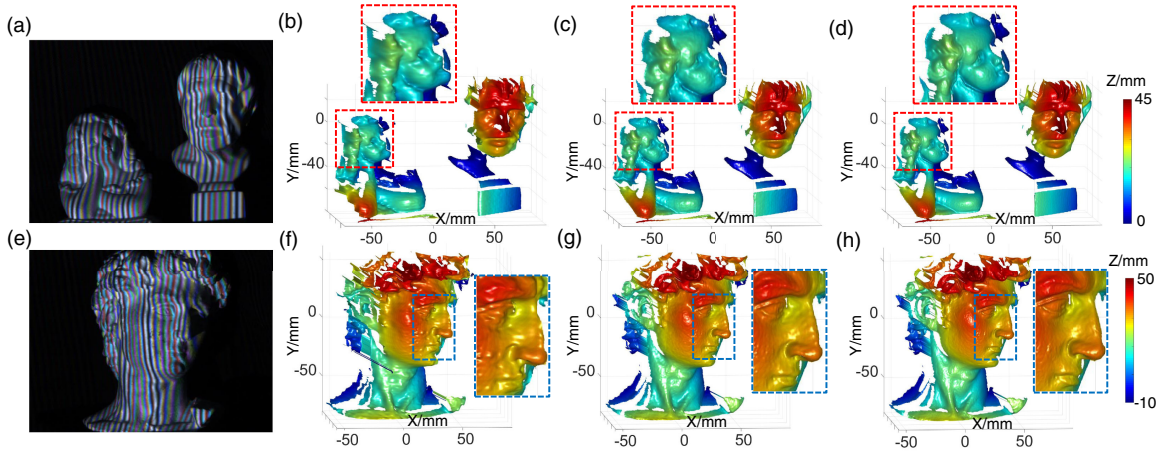

Fig. 3. Measurement results of two scenes. (a), (e) Captured color images. (b), (f) Results measured by the method of [13]. (c), (g) Results of our method. (d), (h) Ground truth obtained by 12-step PS and the projection distance minimization.

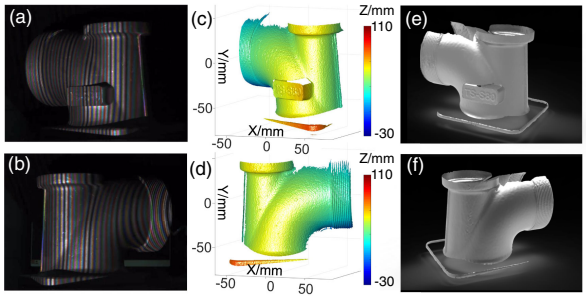

Fig. 4. Measured results of a 360-deg rotated workpiece with our method. (a), (b) Color images of different times. (c), (d) Results corresponding to (a), (b). (e), (f) Registration results (see Visualization 1, Visualization 2, and Visualization 3 for the whole process).

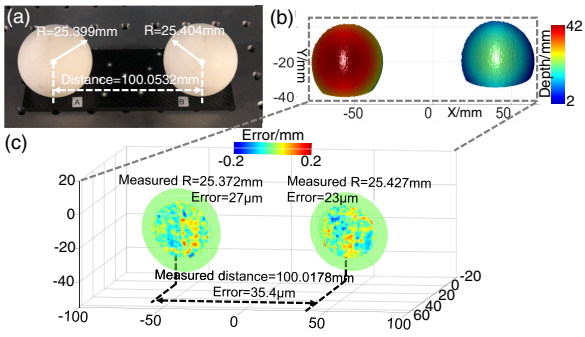

Fig. 5. Quantitative analysis of the accuracy of our method. (a) Measured standard spheres. (b) Measured results of our method. (c) Error distribution of the measured result.

different perspectives are shown in Figs. 4(a)–4(d). Due to the single-shot nature of our approach, the measurement can be performed uninterruptedly without being affected by motion artifacts. We register all the independent measurements into an integral 3D model, which is shown in Figs. 4(e) and 4(f). This experiment demonstrates the potential of our method for rapid reverse engineering applications.

Finally, to quantify the accuracy of our method, two standard spheres as shown in Fig. 5(a) are measured. The measured result is shown in Fig. 5(b). We perform sphere fitting to the results, and the error distribution is shown in Fig. 5(c). The radii of reconstructed spheres are 25.372 mm and 25.427 mm, with errors of 27 μm and 23 μm. The measured center distance is 100.0178 mm, with an error of 35.4 μm. This experiment validates the high measurement accuracy of our method.

In this Letter, we have presented a deep-learning-based color FPP. With deep learning, color-coded projection technologies are rejuvenated in single-frame, high-precision 3D imaging. Deep learning breaks the dependence of traditional methods on prior knowledge and can more efficiently utilize the raw information "hidden" in the original fringe pattern. However, it should be mentioned that for some other disadvantages of color-coded projection methods, such as being unsuited for measuring color objects, the proposed depth learning approach is still inadequate because the information cannot "come out of thin air." The fringe in a certain channel will be blended onto the object surface with the same color, leading to extremely poor fringe contrast and information lost. One possible solution is to create "invisible fringes" with infrared or ultraviolet light sources, which is an interesting direction for future work.

Funding. National Natural Science Foundation of China (61705105, 61722506); National Key Research and Development Program of China (2017YFF0106403); Outstanding Youth Foundation of Jiangsu Province of China (BK20170034); Fundamental Research Funds for the Central Universities (30917011204, 30919011222).

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. S. S. Gorthi and P. Rastogi, Opt. Lasers Eng. **48**, 133 (2010).
2. S. Feng, C. Zuo, T. Tao, Y. Hu, M. Zhang, Q. Chen, and G. Gu, Opt. Lasers Eng. **103**, 127 (2018).
3. J. Qian, S. Feng, T. Tao, Y. Hu, K. Liu, S. Wu, Q. Chen, and C. Zuo, Opt. Lett. **44**, 5751 (2019).
4. T. Tao, Q. Chen, S. Feng, J. Qian, Y. Hu, L. Huang, and C. Zuo, Opt. Express **26**, 22440 (2018).
5. C. Zuo, L. Huang, M. Zhang, Q. Chen, and A. Asundi, Opt. Lasers Eng. **85**, 84 (2016).
6. C. Zuo, Q. Chen, G. Gu, S. Feng, F. Feng, R. Li, and G. Shen, Opt. Lasers Eng. **51**, 953 (2013).
7. J. Qian, T. Tao, S. Feng, Q. Chen, and C. Zuo, Opt. Express **27**, 2713 (2019).
8. Z. Zhang, Opt. Lasers Eng. **50**, 1097 (2012).
9. P. S. Huang, Q. Hu, F. Jin, and F. P. Chiang, Opt. Eng. **38**, 1065 (1999).
10. W.-H. Su, Opt. Express **16**, 2590 (2008).
11. N. Karpinsky and S. Zhang, Opt. Eng. **49**, 063604 (2010).
12. Z. Zhang, C. E. Towers, and D. P. Towers, Opt. Express **14**, 6444 (2006).
13. Z. Zhang, D. P. Towers, and C. E. Towers, Appl. Opt. **49**, 5947 (2010).
14. C. Zuo, S. Feng, L. Huang, T. Tao, W. Yin, and Q. Chen, Opt. Lasers Eng. **109**, 23 (2018).
15. X. Su and Q. Zhang, Opt. Lasers Eng. **48**, 191 (2010).
16. D. Towers, C. Towers, and Z. Zhang, "Optical imaging of physical objects," U.S patent App. 12/377,180 (7 July 2010).
17. L. Huang, Q. Kemao, B. Pan, and A. K. Asundi, Opt. Lasers Eng. **48**, 141 (2010).
18. Z. Zhang, C. Towers, and D. Towers, Opt. Lasers Eng. **48**, 159 (2010).
19. S. Feng, Q. Chen, G. Gu, T. Tao, L. Zhang, Y. Hu, W. Yin, and C. Zuo, Adv. Photon. **1**, 025001 (2019).
20. S. Van der Jeught and J. J. Dirckx, Opt. Express **27**, 17091 (2019).
21. S. Feng, C. Zuo, W. Yin, G. Gu, and Q. Chen, Opt. Lasers Eng. **121**, 416 (2019).
22. W. Yin, Q. Chen, S. Feng, T. Tao, L. Huang, M. Trusiak, A. Asundi, and C. Zuo, arXiv preprint arXiv:1903.09836 (2019).
23. J. Qian, S. Feng, T. Tao, Y. Hu, Y. Li, Q. Chen, and C. Zuo, arXiv preprint arXiv:2001.01439 (2020).
24. C. Zuo, T. Tao, S. Feng, L. Huang, A. Asundi, and Q. Chen, Opt. Lasers Eng. **102**, 70 (2018).
25. Y. Yin, X. Peng, A. Li, X. Liu, and B. Z. Gao, Opt. Lett. **37**, 542 (2012).