Deep-learning-enabled temporally super-resolved multiplexed fringe projection profilometry: high-speed kHz 3D imaging with low-speed camera

Wenwu Chen[1,2,3], Shijie Feng[1,2,3*], Wei Yin[1,2,3], Yixuan Li[1,2,3], Jiaming Qian[1,2,3], Qian Chen[3*] and Chao Zuo[1,2,3*]

*Correspondence:
shijiefeng@njust.edu.cn;
chenqian@njust.edu.cn;
zuochao@njust.edu.cn

[1] Smart Computational Imaging
Laboratory (SCILab), School
of Electronic and Optical
Engineering, Nanjing University
of Science and Technology,
Nanjing 210094, Jiangsu
Province, China
[2] Smart Computational Imaging
Research Institute (SCIRI)
of Nanjing University of Science
and Technology, Nanjing 210019,
Jiangsu Province, China
[3] Jiangsu Key Laboratory
of Spectral Imaging & Intelligent
Sense, Nanjing University
of Science and Technology,
Nanjing 210094, Jiangsu
Province, China

Abstract

Recent advances in imaging sensors and digital light projection technology have facilitated rapid progress in 3D optical sensing, enabling 3D surfaces of complex-shaped objects to be captured with high resolution and accuracy. Nevertheless, due to the inherent synchronous pattern projection and image acquisition mechanism, the temporal resolution of conventional structured light or fringe projection profilometry (FPP) based 3D imaging methods is still limited to the native detector frame rates. In this work, we demonstrate a new 3D imaging method, termed deep-learning-enabled multiplexed FPP (DLMFPP), that allows to achieve high-resolution and high-speed 3D imaging at near-one-order of magnitude-higher 3D frame rate with conventional low-speed cameras. By encoding temporal information in one multiplexed fringe pattern, DLMFPP harnesses deep neural networks embedded with Fourier transform, phase-shifting and ensemble learning to decompose the pattern and analyze separate fringes, furnishing a high signal-to-noise ratio and a ready-to-implement solution over conventional computational imaging techniques. We demonstrate this method by measuring different types of transient scenes, including rotating fan blades and bullet fired from a toy gun, at kHz using cameras of around 100 Hz. Experiential results establish that DLMFPP allows slow-scan cameras with their known advantages in terms of cost and spatial resolution to be used for high-speed 3D imaging tasks.

Keywords: 3D imaging, Fringe projection profilometry, Multiplex, Deep learning, Temporal super-resolution

Introduction

Over recent decades, significant advancements in optoelectronics have ignited interests in capturing and documenting instantaneous phenomena. The ability to capture immediate three-dimensional (3D) geometric changes in objects provides invaluable insights into fast events, crucial for diverse fields such as industrial inspection [1], biomedicine [2], and solid mechanics [3]. Among the array of 3D imaging techniques, fringe projection profilometry (FPP) [4] is one of the most promising modalities due to its capacity for high-accuracy and full-field 3D measurements.

To enhance the speed of FPP, efforts have been made to improve the speed of measurement system. Binary defocusing techniques, for instance, have emerged to increase the projection speed of digital light processing (DLP) systems [5, 6]. By projecting binary fringes (1-bit) instead of grayscale patterns (8-bit) in a defocused manner, these techniques have demonstrated the capability to increase projection speeds from a hundred frames per second (fps) to thousands or even tens of thousands fps. Additionally, custom projectors utilizing rotating wheels [7] or LED arrays [8, 9] have also been developed to achieve high-speed pattern projection.

Although system speed has improved, motion can still compromise 3D measurements if numerous patterns are required for dynamic 3D reconstruction [10]. Therefore, researchers have presented methods using a small number of patterns, such as dual-frequency phase-shifting (PS) [11], bi-frequency PS [12], 2+2 PS [9], composite PS [13], and micro Fourier transform profilometry [14]. These approaches utilize each projected pattern for both wrapped phase calculation and absolute phase unwrapping, effectively reducing the number of patterns. Fourier transform profilometry (FTP) employs a single fringe pattern for 3D reconstruction but struggles with complex shapes due to spectrum aliasing [15]. Recent advancements in artificial intelligence have introduced deep neural networks (DNNs) [16, 17] to optical metrology [18]. Properly trained DNNs can retrieve phase [19] and 3D coordinates [20–23] using a single fringe pattern accurately for complex objects, pushing the 3D measurement speed to the upper limit that is the camera's speed for capturing two-dimensional (2D) images.

However, enhancing the camera's speed often comes at a cost, such as the decrease in pixel resolution and the signal-to-noise ratio (SNR) of captured images. Although high-speed cameras capture images at a high frame rate without reducing the resolution, the cost of the system will sharply increase. Moreover, the speed of 3D imaging is inherently hindered by the rate at which 2D images can be captured and processed. Therefore, we are facing a big challenge that is "*can affordable low-speed cameras be used to replace high-speed cameras and achieve high-speed 3D imaging without compromising image resolution*".

In recent years, we have witnessed the rapid progress of deep learning in computational imaging [24]. Meanwhile, the refresh rate of digital micro-mirror devices (DMDs) has significantly increased, reaching tens of thousands fps, while at an affordable price. This motivated us to combine computational imaging and deep learning to encode temporal information in space and break through the physical limits of camera hardware speed. Inspired by the concept of holographic multiplexing [25], for the first time to our knowledge, we introduce a novel approach termed deep-learning-enabled multiplexed FPP (DLMFPP). DLMFPP enables high-speed 3D imaging, surpassing the camera's acquisition rate by nearly an order of magnitude, while preserving spatial resolution. We employ a series of fringe images with varying tilt angles. When the speed of projector is higher than that of camera, we capture a multiplexed image overlaid with a sequence of fringe patterns. DLMFPP can decode the image into its original sequence by DNNs embedded with Fourier transform (FT), PS [26], and ensemble learning [27]. By harnessing each fringe pattern to record the scene at different time, it achieves up to 9x temporal super-resolution imaging beyond the camera's frame rate. In practice, the DLMFPP method can be implemented on almost any off-the-shelf FPP system, eliminating the need for complicated optical paths

and furnishing a high SNR and ready-to-use solution compared to conventional computational imaging techniques [28–30]. We validate the effectiveness and versatility of DLMFPP through experimental demonstrations on different types of transient scenes, including rotating fan blades and bullet fired from a toy gun, showcasing its ability to achieve high-speed kHz 3D imaging with low-speed cameras operating at around 100 Hz. By transcending the limitations of sensor frame rates, the DLMFPP allows slow-scan cameras to quantitatively study dynamic processes with both high spatial and temporal resolution.

Methods

The schematic of the DLMFPP approach is demonstrated in Fig. 1. The projector sequentially projects fringe patterns I_m^p with different directions onto the dynamic scene. The pattern sequence can be represented as

$$I_m^p(x^p, y^p) = a^p + b^p cos[\varphi_m^p(x^p, y^p)], \tag{1}$$

where (x^p, y^p) represents the pixel coordinate of projector, a^p is the mean value, b^p is the amplitude, and m denotes the pattern index $m = 1, 2, 3, ..., M$ (M is the total number of the patterns). The phase φ_m^p is assigned as

$$\varphi_m^p(x^p, y^p) = 2\pi \left(f_x^p x^p cos\theta_m + f_y^p y^p sin\theta_m \right), \tag{2}$$

$$\theta_m = (-1)^m \left(\frac{m}{2} + \frac{(-1)^m - 1}{4} \right) \theta, \tag{3}$$

where f_x^p and f_y^p are the frequency in x^p, y^p directions, respectively, and θ is a scalar characterizing the incline of fringes. After modulated by the object surface, the corresponding fringe images I_m (shown in Fig. 1) can be expressed as

$$I_m(x, y) = A_m(x, y) + B_m(x, y) \cos[\phi_m(x, y)], \tag{4}$$

where (x, y) indicates the pixel coordinate of camera, A_m is the average intensity, B_m is the modulation, and ϕ_m is the phase to be measured. Letters of "MULTIPLEX" in Fig. 1 represent a dynamic scene, and each I_m encodes the scene at different time t. Then, the camera captures a multiplexed image I_{LE} overlaid by the sequence of I_m with a long exposure time. After performing FT on I_{LE}, multiple fundamental frequency components (corresponding to I_m) are circularly distributed in the spatial spectrum $\mathcal{F}_{LE}$, occupying distinct locations. Specifically, we consider four principles when designing the pattern sequence I_m^p: (1) the fringe interval in each I_m^p is kept equal to guarantee the consistent defocusing level when capturing the binary pattern sequence; (2) the zero component in $\mathcal{F}_{LE}$ should be far away from the fundamental components to avoid spectrum overlap; (3) the fundamental components of these fringe patterns should be distributed in a circular pattern in $\mathcal{F}_{LE}$, which minimizes the harm of spectrum leakage; (4) fundamental components near f_y axis should be excluded as it is hard to employ this kind of near-horizontal fringe pattern to measure 3D shape for a conventional horizontally configured FPP system.

The flowchart of DLMFPP is shown in Fig. 2, where there are two steps to analyze the input multiplexed image. Step 1 is to decompose the multiplexed pattern into a fringe
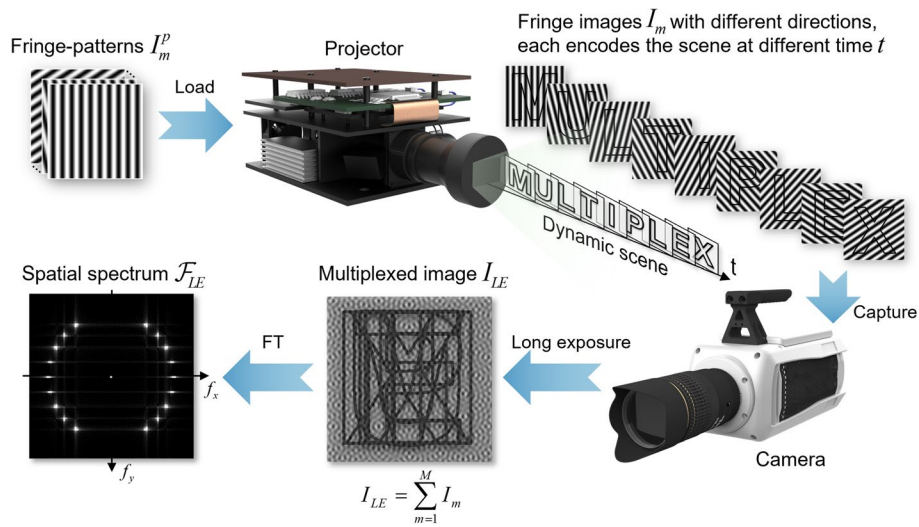
Fig. 1 Schematic of DLMFPP: The projector sequentially projects fringe patterns I_m^p [Eq. (1)] onto the dynamic scene, allowing the corresponding modulated fringe images I_m [Eq. (4)] to encode the scene at different time t. Then the camera captures a multiplexed image I_{LE} with a long exposure time, and the spatial spectrum $\mathcal{F}_{LE}$ (multiple fundamental components corresponding to I_m are circularly distributed) can be obtained by FT (pattern index $m = 1, 2, 3, ..., M$, M is the total number of the patterns). A synthetic scene composed of letters, "MULTIPLEX", is used to illustrate the principle

pattern sequence, each of which corresponds to the measured object at a moment. Step 2 is to analyze the decomposed fringe patterns for phase retrieval. To be specific, inspired by the rationalized deep learning framework [31], we propose a multiplexed pattern decomposing module (DNN1) that comprises three branches. The spatial decomposing (SD) branch is trained to extract the features of the multiplexed image I_{LE} and decompose it in the spatial domain. The frequency decomposing (FD) branch, which is parallel to the SD branch, incorporates the physical model of FT into the framework to analyze the multiplexed image as follows: (1) it obtains the spatial spectrum $\mathcal{F}_{LE}$ of I_{LE} by FT, and feeds its real and imaginary components into the FD branch [32]; (2) the branch then decomposes $\mathcal{F}_{LE}$ in frequency domain and outputs the real and imaginary parts of the separate spectrums as the branch output; (3) inverse FT (iFT) is performed to obtain separate fringe images. The feature ensemble (FE) branch is engineered to adaptively merge features learned by the SD and FD branches with the idea of ensemble learning [27]. This branch can incorporate features from both spatial and frequency domains and give the final outputs, i.e., separate fringe images $I_1 - I_9$ in Fig. 2. In Step 2, we design an augmented fringe pattern analysis (AFPA) module (DNN2) embedded with the physical model of PS to retrieve the phase from each fringe image. The module receives each separate fringe image I_m as input and predicts the corresponding numerator M_m and denominator D_m. Then, the wrapped phase ϕ_m in Eq. (4) is demodulated through an arctangent function

$$\phi_m(x,y) = \arctan \frac{cB_m(x,y)\sin[\phi_m(x,y)]}{cB_m(x,y)\cos[\phi_m(x,y)]} = \arctan \frac{M_m(x,y)}{D_m(x,y)}, \qquad (5)$$

where c is a constant determined by the phase demodulation approach, pattern index $m = 1, 2, 3, ..., 9$. After that, the absolute phase Φ_m can be acquired with the help of ϕ_m' from another camera via stereo phase unwrapping (SPU) [33], then 3D reconstruction
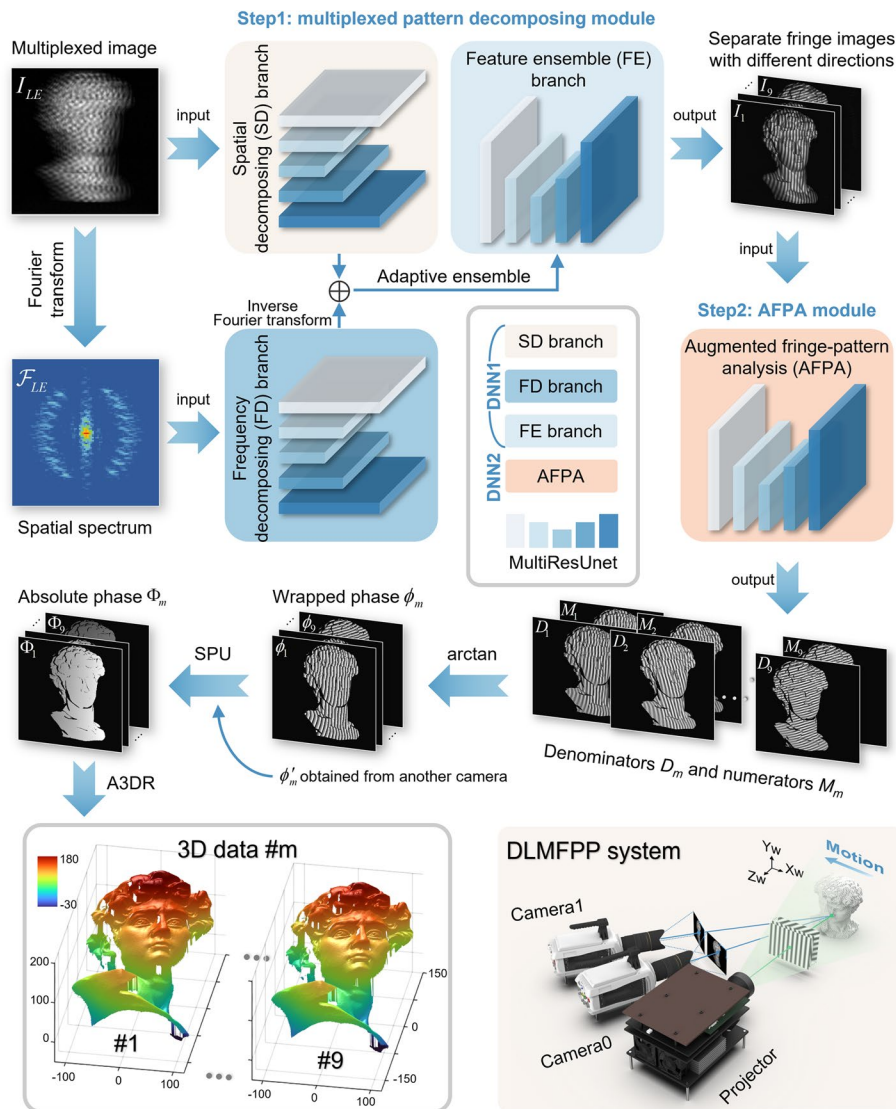
Fig. 2 Flowchart of DLMFPP. A multiplexed image I_{LE} and its spatial spectrum $\mathcal{F}_{LE}$ are fed into a multiplexed pattern decomposing module (DNN1) comprised of three branches. The DNN1 framework incorporates the physical model of FT and the idea of ensemble learning to decompose I_{LE} and output separate fringe images I_m. The AFPA module (DNN2) embedded with the physical model of PS receives each I_m to predict the corresponding M_m and D_m, enabling wrapped phase ϕ_m calculation via Eq. (5). The absolute phase Φ_m is then derived by SPU, and 3D data of #m can be reconstructed by the developed A3DR (pattern index $m = 1, 2, 3, ..., 9$). The insert shows the DLMFPP system configuration, consisting of a projector and two cameras. The projector sequentially projects nine fringe patterns with different directions onto a moving object, then the cameras capture the multiplexed image (shown as I_{LE}) with a long exposure time

can be performed. Notably, in a conventional horizontally configured FPP system, the mapping from phase to 3D coordinates is generally designed for vertical fringes. To cope with the case of arbitrarily oriented fringes in this work, we propose the augmented 3D reconstruction (A3DR) method. By creating a unique correspondence value $x^p \cos\theta_m + (f_y^p/f_x^p)y^p \sin\theta_m$ for every camera pixel coordinate (x, y), 3D reconstruction can be performed from Eq. (S13) with pre-calibrated parameters. For further details on system calibration and A3DR, see Supplementary Note 6.

The SD, FD, FE branches and the AFPA module are constructed by MultiResUnet [34], which is a novel architecture that combines MultiRes blocks and residual paths on the well-known U-Net framework [35], owing the advantage to reconcile features from different context size, alleviate the disparity between the encoder-decoder features, save memory and speed up network training (detailed in Supplementary Note 2 and Fig. S2). Network training for multiplexed pattern decomposition and phase retrieval is carried out in a supervised manner, and the process is elaborated in Supplementary Note 4 and Fig. S4. Moreover, for the objective functions of training, the SD and FD branches use joint losses containing data-based and physics-based loss, while the FE branch and the AFPA module use only the data-based loss. The combination of physical and data loss can effectively improve the recovered accuracy and generalization of the DNNs. Details related to the loss functions design are provided in Supplementary Note 5 and Fig. S5. By incorporating FT, PS and ensemble learning, DLMFPP embeds more physical prior knowledge in the network structure and loss functions to provide reliable phase recovery across various scenes and conditions, significantly improving the generalization ability of networks.

We developed the DLMFPP system shown in the insert of Fig. 2, composed by two CMOS cameras (Vision Research Phantom V611) and a customized projection system with an XGA resolution (1024×768) DMD. By functioning in binary (1-bit) mode, the DMD is manipulated to achieve a refresh rate of 1,000 fps. Meanwhile, the cameras are operated at an image resolution (640×440) with pixel depth of 16 bits. The projection system outputs a trigger signal every nine frames, thus the cameras work at a frame rate of ∼111.11 Hz. DLP development hardware is used for precisely triggering to ensure signal synchronization between the projector and the cameras. For more information about the system synchronization, see Supplementary Note 1 and Fig. S1. During the training stage, we photographed a variety of objects made of different materials (plastic, plaster, metal, ceramic, etc.) to generate diverse datasets. In this work, 1,200 groups of images were captured, of which 800 groups were used for training and 400 groups for validation. Details of training dataset generation can be found in Supplementary Note 3 and Fig. S3.

Results

To evaluate the contribution of each branch in DLMFPP, we measured three scenes to conduct an ablation study as shown in Fig. 3. The ground truths of separate fringe images were captured by setting the camera frame rate to 1,000 Hz (same as the DMD refresh rate). Then, the ground truths of phase were obtained by 12-step PS, as in Fig. 3e (detailed in Supplementary Note 3). Figure 3a shows multiplexed images modulated by the scenes (insets show the corresponding Fourier frequency spectrums, locally zoomed in for better visibility) and the phase errors of FTP. We can see substantial phase errors on the sharp edges of the measured surface, and the average mean absolute error (MAE) of these scenes is up to 0.4731 rad. Figure 3b-d show the separate fringe images decomposed by the SD, FD, and FE branches, respectively, and the corresponding phase errors of the reconstructed results demodulated by AFPA. From the fringe images in Fig. 3b, we can observe obvious noise. Meanwhile, blur fringes can be observed around the edges of the object as shown in Fig. 3c, which results in significant phase errors with an average MAE of 0.2091 rad. Contrastingly, in Fig. 3d, the FE branch harnesses the idea
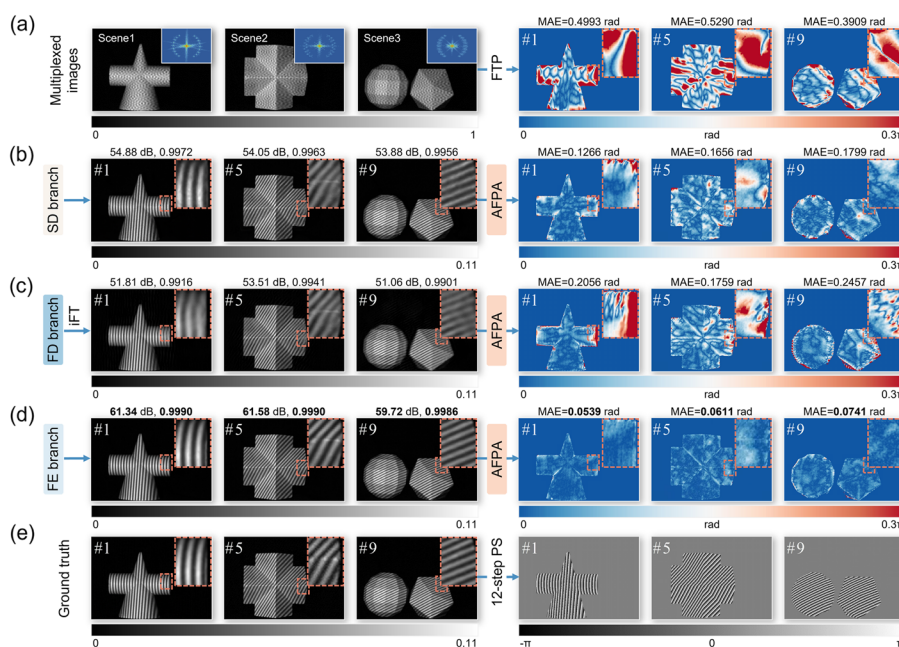
Fig. 3 Ablation study of DLMFPP: **a** Multiplexed images modulated by 3 different scenes [insets show the corresponding spatial spectrums (locally zoomed in)] and phase errors of FTP; **b**-**d** separate fringe images decomposed by SD, FD, and FE branches, respectively, evaluated by PSNR and SSIM, and phase errors of the reconstructed results demodulated by AFPA; **e** ground truths of separate fringe images and phase, obtained by setting the camera frame rate same as the DMD refresh rate (1,000 Hz) and 12-step PS (#*m* represents the *m*th pattern index of each scene, and $m = 1, 2, 3, ..., 9$)

of ensemble learning to integrate features from both the spatial and frequency domains, yielding a high-quality restoration of fringe images. The resultant average peak SNR (PSNR) ups to 60.88 dB and the average structural similarity index (SSIM) ups to 0.9989. By feeding these fringe images into AFPA, we can achieve high-accuracy phase recovery with the average MAE of 0.0630 rad.

For dynamic 3D measurements of moving objects, we applied DLMFPP to measure a fan with 4 rotating plastic blades. Figure 4a presents a particular frame of the multiplexed image I_{LE} and corresponding spectrum $\mathcal{F}_{LE}$ (locally zoomed in for better visibility). Although significant motion blur of the blades is observed in the multiplexed image, the proposed DLMFPP can still successfully reconstruct the 3D shape of the blades, as shown in Fig. 4b and e. It is noted that the motion blur in DLMFPP is not determined by the camera exposure time, but by the projection time, which is near-one-order of magnitude-lower than the exposure time of a single camera frame. This greatly reduced exposure time effectively handles the challenges of motion blur of dynamic scene changes, thus ensuring accurate 3D reconstruction. For more information on the discussion of motion blur in DLMFPP, see Supplementary Note 8 and Fig. S8. Figure 4c plots the displacement of z at 3 selected point locations within 90 ms [A, B, and C in Fig. 4b], revealing that the rotation period of the fan blades is 45 ms, i.e., the rotation speed is 1,333 rotations per minute (rpm). Figure 4d shows five fringe images (I_1, I_3, I_5, I_7, and I_9, corresponding to T = 27, 29, 31, 33, and 35 ms) decoded from the multiplexed image I_{LE} and the corresponding 3D model reconstructed by the proposed DLMFPP. Moreover, Fig. 4f displays two cross sections of the 3D reconstruction, one of which shows the tangential
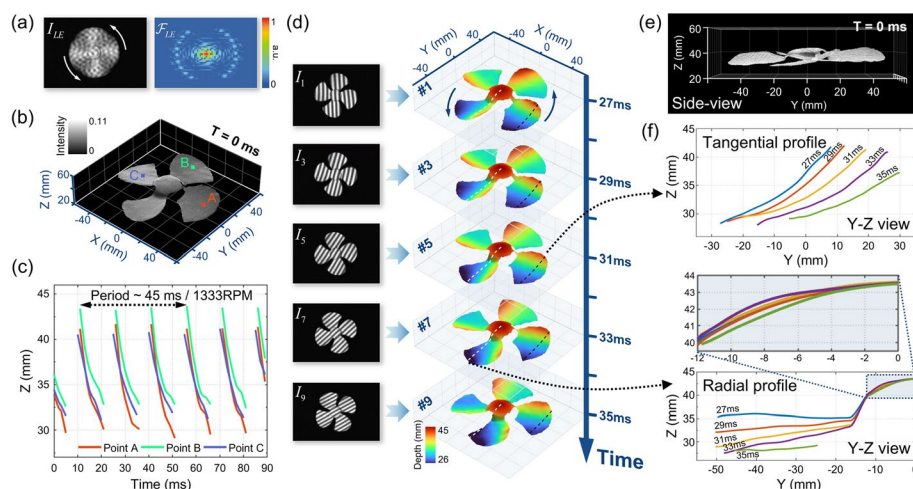
Fig. 4 Measurement of a rotating fan by DLMFPP. **a** The multiplexed image I_{LE} and corresponding spectrum $\mathcal{F}_{LE}$ (locally zoomed in). **b** 3D reconstruction of the fan at T = 0 ms. **c** Displacement of z at 3 selected point locations within 90 ms [A, B, and C in (**b**)]. **d** Five fringe images (I_1, I_3, I_5, I_7, and I_9, corresponding to T = 27, 29, 31, 33, and 35 ms) decoded from the multiplexed image I_{LE}, and the corresponding 3D model reconstructed by DLMFPP. **e** Side-view of (**b**). **f** Two cross sections of the 3D reconstruction, one of which shows the tangential profile (black dot line) and the other the radial profile (white dot line). The local zoomed-in view shows the profile of the centre hub

profile (black dot line) and the other the radial profile (white dot line). The profile of the centre hub is shown in the zoomed-in view. The corresponding 3D movie about the complete process of DLMFPP and 3D reconstruction results of the whole dynamic process of the rotating fan is further provided in Supplementary Movie S1. With this experiment, we can see that DLMFPP accurately retrieved nine 3D images with each multiplexed image I_{LE}, validating that 1,000 Hz high-speed 3D shape measurement has been achieved with cameras running at ~111.11 Hz. Additionally, we applied DLMFPP to image a running fascia gun for a supplementary experiment. It shows that the cyclic movement of the gun head has a period of about 35 ms, which corresponds to a speed of 1,714 rpm of the rotary motor inside the gun. More experimental results are provided in Supplementary Note 9, Fig. S9 and Supplementary Movie S3.

To verify the scalability of our DNNs, we developed another system consisting of two low-speed cameras (Basler acA640-750um) and the same projection unit. The cameras are equipped with zoom lenses that adjust the focal length, aperture size and degree of focus to make the field of view and brightness consistent with the existing datasets. So we can directly utilize the trained DNNs before. The projector operated at the rate of 1,080 fps and the camera at 120 fps. For the dynamic experiment, we measured a one-time transient event: a bullet was fired diagonally downward from a toy gun, and then rebounded from the ground. Representative 3D reconstruction results during the event are presented in Fig. 5a. The bullet began to appear near the muzzle at 11.1 ms. It flew straight forward until 59.3 ms and then hit the ground and rebounded upwards. Three points are selected to demonstrate the performance of DLMFPP [A, B, and C in Fig. 5a]. The displacements in z direction at selected locations are plotted in insets of Fig. 5a, indicating that DLMFPP has accurately recovered the profile of the fast moving bullet at different moments. Figure 5b shows the side-view (y-z) of the 3D reconstruction at T = 45.4 ms, and Fig. 5c shows the trajectory and the variation of the velocity of the bullet
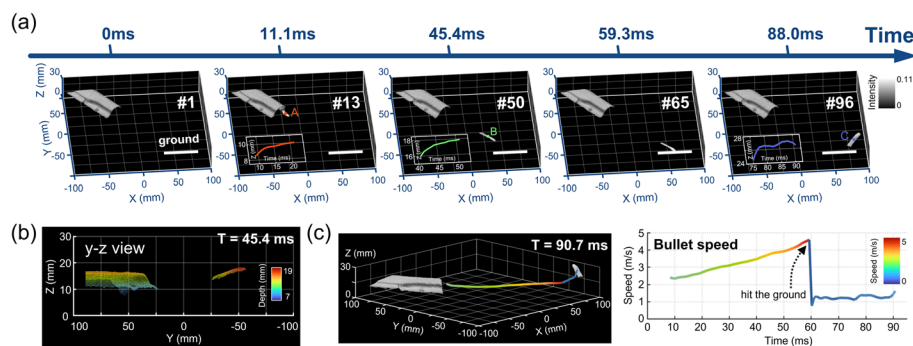
Fig. 5 Measurement of bullet fired from a toy gun by DLMFPP. **a** 3D reconstruction results at T = 0, 11.1, 45.4, 59.3, and 88.0 ms, with insets presenting displacements in z direction at A, B, and C locations. **b** The side-view (y-z) of the 3D reconstruction at T = 45.4 ms. **c** The 3D reconstruction of the scene at T = 90.7 ms, as well as the trajectory and the variation of the velocity of the bullet during the whole process

during the whole process. The initial speed of the bullet was 2.4 m/s at discharge. It accelerated uniformly to 4.6 m/s during the flight and then hit the ground with the speed decreased abruptly to 0.8 m/s (refer to Supplementary Movie S2 for more details). The experiment demonstrates the scalability of our DNNs for high-speed 3D imaging with low-speed cameras and the capability of DLMFPP to capture one-time transient events.

It should be noted that DLMFPP is the first temporally super-resolved 3D imaging technique proposed in FPP, while previous deep learning-based approaches were developed for single-shot 3D imaging [20–23]. The structure, training process, and loss function design of previous networks cannot meet the necessity for high-accuracy phase recovery and measurement in temporally super-resolved 3D imaging, therefore we proposed DLMFPP to address this challenge. To justify the progressiveness of DLMFPP, in Supplementary Note 7 and Fig. S6, we provide a comparative study and analysis between the proposed DLMFPP and two state-of-the-art deep learning-based approaches. This study demonstrates that DLMFPP solves the dilemma of the state-of-the-art methods in handling regions with large height variations and demodulates high-accuracy phase information from the multiplexed image. DLMFPP achieves the lowest phase error with the average MAE of 0.0495 rad, revealing the superior performance achieved from DLMFPP's advanced network design.

For the 3D imaging speed in DLMFPP, the increase of imaging speed depends on the number of overlapped images in a multiplexed image. The overlapping number is referred to as compression rate (CR). In this work, we employ $CR = 9$ when the marginal benefit between CR and recovered phase accuracy is highest (detailed in the comparative study of different CRs in Supplementary Note 7 and Fig. S7), allowing DLMFPP to achieve 9x temporal super-resolution. Practically, to trade off temporal resolution and spatial resolution accuracy, the DLMFPP approach is also flexible. If higher phase accuracy is required, CR can be reduced appropriately, and vice versa.

Discussion and conclusion

In this work, we have introduced a deep-learning-enabled temporally super-resolved 3D measurement approach by multiplexed FPP. By temporally embedding a sequence of fringe patterns with different tilt angles into a single multiplexed image, DLMFPP allows

to achieve high-resolution and high-speed 3D imaging at near-one-order of magnitude-higher 3D frame rate with conventional low-speed cameras. Experiential results demonstrate that kHz 3D imaging can be achieved by using cameras merely running at around 100 Hz without compromising the spatial resolution.

DLMFPP encodes multi-frame temporal information in the spatial dimension, which gives this compressive imaging modality the advantage of cost-effective, low bandwidth/memory requirements, and low power consumption [36]. Moreover, the modality breaks through the limitation of 3D imaging speed imposed by the intrinsic frame rate of the imaging sensor, allowing it to be further used for ultrahigh-speed imaging when combined with high-speed cameras. This new 3D imaging paradigm opens an avenue for the development of high-speed or ultra-high-speed 3D imaging capabilities, thereby pushing the boundaries of current 3D imaging technologies.

Compared to conventional computational imaging techniques [28–30], DLMFPP system eliminates the need for complex optical modulation hardware (e.g., a spatial encoder), avoiding complicated optical paths. Practically, DLMFPP can be implemented on almost any off-the-shelf FPP system. This simple optical path avoids photon losses and makes greater use of optical information, guaranteeing a high SNR in 3D imaging. Moreover, DLMFPP combines the physical models of FT and PS method, and harnesses the idea of ensemble learning to integrate features from both the spatial and frequency domains. This progressive architecture also ensures the high SNR in high-speed 3D imaging with low-speed cameras. From the perspective of space-time-bandwidth product (STBP), the multi-frame modulation mechanism of DLMFPP can rationally harness the spatio-temporal redundancy in fast changing scenes, thereby better utilizing the STBP of sensors compared to conventional single-frame recordings.

Despite promising results in high-speed 3D imaging, DLMFPP still faces challenges. For example, the exclusion of near-horizontal fringe patterns leaves the region near f_y axis in the multiplexed spatial spectrum unused, which exacerbates the harm of spectrum overlap, affecting the recovered phase quality. Moreover, due to the trade-off between CR and the information capacity of each fringe image, further increasing the multiple of temporal super-resolution results in a loss of final phase quality, and vice versa. It should also be noted that the maximum speed of DLMFPP is still constrained by the projection rate. The speed can be potentially further enhanced by using custom physical grating [7] or LED arrays [8, 9], which will be explored in our future research. Furthermore, there is an untapped potential of DLMFPP, as latest innovations in deep learning can be directly introduced into the method. For example, physics-informed learning can bring domain expertise to improve performance [37–40], and all-optical neural networks operating at the speed of light can accelerate computations [41–43].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43074-024-00139-2.

Supplementary Material 1: Supplementary Information.

Supplementary Material 2: Movie S1.

Supplementary Material 3: Movie S2.

Supplementary Material 4: Movie S3.

Availability of data and materials
The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
There is no ethics issue for this paper.

Consent for publication
All authors agreed to publish this paper.

Competing interests
The authors declare that they have no competing interests.

References
1. Malamas EN, Petrakis EGM, Zervakis M, Petit L, Legat JD. A survey on industrial vision systems, applications and tools. Image Vision Comput. 2003;21(2):171–88.
2. Ford KR, Myer GD, Hewett TE. Reliability of landing 3D motion analysis: implications for longitudinal analyses. Med Sci Sports Exerc. 2007;39(11):2021.
3. Tiwari V, Sutton MA, McNeill SR. Assessment of High Speed Imaging Systems for 2D and 3D Deformation Measurements: Methodology Development and Validation. Exp Mech. 2007;47(4):561–79.
4. Gorthi SS, Rastogi P. Fringe projection techniques: whither we are? Optics Lasers Eng. 2010;48(2):133–40.
5. Li B, Wang Y, Dai J, Lohry W, Zhang S. Some recent advances on superfast 3D shape measurement with digital binary defocusing techniques. Optics Lasers Eng. 2014;54:236–46.
6. Zuo C, Chen Q, Feng S, Feng F, Gu G, Sui X. Optimized pulse width modulation pattern strategy for three-dimensional profilometry with projector defocusing. Appl Opt. 2012;51(19):4477–90.
7. Heist S, Lutzke P, Schmidt I, Dietrich P, Kühmstedt P, Tünnermann A, et al. High-speed three-dimensional shape measurement using GOBO projection. Opt Lasers Eng. 2016;87:90–6.
8. Heist S, Mann A, Kühmstedt P, Schreiber P, Notni G. Array projection of aperiodic sinusoidal fringes for high-speed three-dimensional shape measurement. Opt Eng. 2014;53(11):112208.
9. Caspar S, Honegger M, Rinner S, Lambelet P, Bach C, Ettemeyer A. High speed fringe projection for fast 3D inspection. In: Optical Measurement Systems for Industrial Inspection VII. vol. 8082. SPIE; 2011. p. 298–304.
10. Feng S, Zuo C, Tao T, Hu Y, Zhang M, Chen Q, et al. Robust dynamic 3-D measurements with motion-compensated phase-shifting profilometry. Optics Lasers Eng. 2018;103:127–38.
11. Liu K, Wang Y, Lau DL, Hao Q, Hassebrook LG. Dual-frequency pattern scheme for high-speed 3-D shape measurement. Opt Express. 2010;18(5):5229–44.
12. Zuo C, Chen Q, Gu G, Feng S, Feng F, Li R, et al. High-speed three-dimensional shape measurement for dynamic scenes using bi-frequency tripolar pulse-width-modulation fringe projection. Optics Lasers Eng. 2013;51(8):953–60.
13. Tao T, Chen Q, Da J, Feng S, Hu Y, Zuo C. Real-time 3-D shape measurement with composite phase-shifting fringes and multi-view system. Opt Express. 2016;24(18):20253–69.
14. Zuo C, Tao T, Feng S, Huang L, Asundi A, Chen Q. Micro Fourier transform profilometry (μFTP): 3D shape measurement at 10,000 frames per second. Optics Lasers Eng. 2018;102:70–91.
15. Takeda M, Mutoh K. Fourier transform profilometry for the automatic measurement of 3-D object shapes. Appl Opt. 1983;22(24):3977.
16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
17. Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw. 2015;61:85–117.
18. Zuo C, Qian J, Feng S, Yin W, Li Y, Fan P, et al. Deep learning in optical metrology: a review. Light-Sci Appl. 2022;11(1):39.
19. Feng S, Chen Q, Gu G, Tao T, Zhang L, Hu Y, et al. Fringe pattern analysis using deep learning. Adv Photon. 2019;1(02):1.

20. Qian J, Feng S, Li Y, Tao T, Han J, Chen Q, et al. Single-shot absolute 3D shape measurement with deep-learning-based color fringe projection profilometry. Opt Lett. 2020;45(7):1842–5.
21. Qian J, Feng S, Tao T, Hu Y, Li Y, Chen Q, et al. Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement. Apl Photon. 2020;5(4):046105.
22. Li Y, Qian J, Feng S, Chen Q, Zuo C. Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3D shape measurement. Opto-Electron Adv. 2022;5(5):210021.
23. Li Y, Qian J, Feng S, Chen Q, Zuo C. Composite fringe projection deep learning profilometry for single-shot absolute 3D shape measurement. Opt Express. 2022;30(3):3424–42.
24. Barbastathis G, Ozcan A, Situ G. On the use of deep learning for computational imaging. Optica. 2019;6(8):921–43.
25. Shaked NT, Micó V, Trusiak M, Kuś A, Mirsky SK. Off-axis digital holographic multiplexing for rapid wavefront acquisition and processing. Adv Opt Photon. 2020;12(3):556.
26. Zuo C, Feng S, Huang L, Tao T, Yin W, Chen Q. Phase shifting algorithms for fringe projection profilometry: A review. Opt Lasers Eng. 2018;109:23–59.
27. Feng S, Xiao Y, Yin W, Hu Y, Li Y, Zuo C, et al. Fringe-pattern analysis with ensemble deep learning. Adv Photon Nexus. 2023;2(3):036010.
28. Gao L, Liang J, Li C, Wang LV. Single-shot compressed ultrafast photography at one hundred billion frames per second. Nature. 2014;516(7529):74–7.
29. Yuan X, Brady DJ, Katsaggelos AK. Snapshot compressive imaging: theory, algorithms, and applications. IEEE Signal Proc Mag. 2021;38(2):65–88.
30. He Y, Yao Y, Qi D, He Y, Huang Z, Ding P, et al. Temporal compressive super-resolution microscopy at frame rate of 1200 frames per second and spatial resolution of 100 nm. Adv Photon. 2023;5(2):026003.
31. Qiao C, Li D, Liu Y, Zhang S, Liu K, Liu C, et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. Nat Biotechnol. 2023;41(3):367–77.
32. Yin W, Che Y, Li X, Li M, Hu Y, Feng S, et al. Physics-informed deep learning for fringe pattern analysis. Opto-Electron Adv. 2024;7(1):230034–1.
33. Weise T, Leibe B, Van Gool L. Fast 3D Scanning with Automatic Motion Compensation. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis: IEEE; 2007. pp. 1–8.
34. Ibtehaz N, Rahman MS. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw. 2020;121:74–87.
35. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer; 2015. p. 234–241.
36. Zhang Z, Zhang B, Yuan X, Zheng S, Su X, Suo J, et al. From compressive sampling to compressive tasking: retrieving semantics in compressed domain with low bandwidth. PhotoniX. 2022;3(1):19.
37. Kellman MR, Bostan E, Repina NA, Waller L. Physics-based learned design: optimized coded-illumination for quantitative phase imaging. IEEE Trans Comput Imaging. 2019;5(3):344–53.
38. Wang F, Bian Y, Wang H, Lyu M, Pedrini G, Osten W, et al. Phase imaging with an untrained neural network. Light Sci Appl. 2020;9(1):77.
39. Bostan E, Heckel R, Chen M, Kellman M, Waller L. Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network. Optica. 2020;7(6):559–62.
40. Saba A, Gigli C, Ayoub AB, Psaltis D. Physics-informed neural networks for diffraction tomography. Adv Photon. 2022;4(6):066001.
41. Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M, et al. All-optical machine learning using diffractive deep neural networks. Science. 2018;361(6406):1004–8.
42. Liu J, Wu Q, Sui X, Chen Q, Gu G, Wang L, et al. Research progress in optical neural networks: theory, applications and developments. PhotoniX. 2021;2:1–39.
43. Luo Y, Zhao Y, Li J, Çetintaş E, Rivenson Y, Jarrahi M, et al. Computational imaging without a computer: seeing through random diffusers at the speed of light. ELight. 2022;2(1):4.

Publisher's Note