Optics and Lasers in Engineering 121 (2019) 416–427

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Micro deep learning profilometry for high-speed 3D surface imaging

Shijie Feng[a,b,c], Chao Zuo[a,b,c,*], Wei Yin[a,b,c], Guohua Gu[a,b], Qian Chen[a,b,*]

[a] School of Electronic and Optical Engineering, Nanjing University of Science and Technology, No. 200 Xiaolingwei Street, Nanjing, Jiangsu Province 210094, China
[b] Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing, Jiangsu Province 210094, China
[c] Smart Computational Imaging Laboratory (SCILab), Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

ARTICLE INFO

Keywords:
Deep learning
3D surface imaging
Structured light

ABSTRACT

How to obtain object information as rich as possible, with the highest possible speed and accuracy from recorded optical signals, has been a crucial issue to the pursuit of powerful imaging technologies. Nowadays, the speed of ultra-fast photography can exceed one quadrillion. However, it can record only two-dimensional images which lack the depth information, greatly limiting our ability to perceive and to understand the complex real-world objects. Inspired by recent successes of deep learning methods in computer vision, we present a novel high-speed three-dimensional (3D) surface imaging approach named micro deep learning profilometry (μDLP) using the structured light illumination. With a properly trained deep neural network, the phase information is predicted from a single fringe image and then can be converted into the 3D shape. Our experiments demonstrate that μDLP can faithfully retrieve the geometry of dynamic objects at 20,000 frames per second. Moreover, comparative results show that μDLP has superior performance in terms of the phase accuracy, reconstruction efficiency, and the ease of implementation over widely used Fourier-transform-based fast 3D imaging techniques, verifying that μDLP is a powerful high-speed 3D surface imaging approach.

1. Introduction

It is usually said that the first instance of what we would call high-speed photography nowadays was to settle the hot dispute "is there a moment in a horse's gait when all four hooves are off the ground at once?" in 1872 [1]. Eadweard Muybridge, a pioneer in the field of motion study, developed an imaging system that involved 12 cameras triggered by the legs of the horse through tripwires, successfully capturing photos on photographic glass plates at the shutter speed of near 2000 frames per second (fps) [2]. After that, the major development for high-speed photography came, as with scientific purposes, in the wake of the researches on nuclear weapons during the cold war. With applications of rotating mirror technologies, streak cameras, and rotating prism cameras [3], the imaging speed soared up to 100 million fps, i.e., Mfps. In the late nineteenth century, the high-speed imaging underwent a further advancement owing to the great breakthrough in electronic semiconductor devices, leading to film-based cameras replaced gradually by CCD or CMOS based cameras [4]. Nowadays, with the assistance of laser, e.g., the femtosecond laser pulse [5], the imaging speed can even exceed one quadrillion, i.e., 10^{15} fps. Benefiting from the ever-increasing power of the high-speed photography, many transient events, which happen at femtosecond to nanosecond time scale and reflect significant fundamental mechanisms, can be analyzed in-depth [6–11].

However, most high-speed cameras or imaging systems can record only two-dimensional (2D) images which lack the depth information. This fundamental restriction greatly limits our ability to perceive and to understand the complex real-world objects. The past several decades have witnessed tremendous development in three-dimensional (3D) imaging technologies in many fields including biomechanics [12], geomaterials [4], industrial manufacturing [13–15], driven by the rapid advances in sensors, optical engineering and computer vision [16–21]. In general, optical 3D surface imaging techniques can be classified into two categories: the passive approaches and the active ones. Stereo vision techniques, as the representative passive methods, capture inherent surface textures from two or more viewpoints and calculate 3D shapes through triangulation [22]. However, they are susceptible to uniform or periodic textures. Compared with the passive sensing, active methods encode test objects with predesigned signals, thus reducing the dependence of the object textures and increasing the accuracy of 3D reconstructions. Time-of-flight (ToF) techniques emit a modulated light ray onto test objects and collect the light scattered back. The distance is then estimated via multiplying the speed of light by the time delay of the light pulse [23]. As the 3D reconstruction of ToF is not based on triangulation, the system can be made very compactly for applications where portable equipment is preferred. Microsoft Kinect 2 exploits this technique for real-time 3D imaging and finds applications for

* Corresponding authors.
E-mail addresses: shijiefeng@njust.edu.cn (S. Feng), zuochao@njust.edu.cn (C. Zuo), chenqian@njust.edu.cn (Q. Chen).

https://doi.org/10.1016/j.optlaseng.2019.04.020
Received 14 February 2019; Received in revised form 22 April 2019; Accepted 23 April 2019
Available online 15 May 2019

human-computer interactions [24]. But, the depth precision of ToF is generally not high for short-range inspections as light travels too fast. As another extensively used active methods, the structured light techniques illuminate test scenes with 2D spatially varying intensity pattern. The 3D shape is extracted based on the information from the distortion of captured structured light patterns. Because of the advantages of favorable flexibility and versatility, 3D surface imaging based on the structured light illumination is receiving increasing attention, and becoming more and more important. The commercial success of these techniques includes Microsoft Kinect 1 [25], Intel RealSense [26], Apple iPhone X [27], and OPPO Find X [28]. Owing to advances of intelligent manufacturing, pilotless vehicle, and cloud imaging, the desire to developing real-time (~ 30 fps) or high-speed (> 10,000 fps) 3D imaging techniques has never been more apparent [29,30].

Rapid developments in high-frame-rate imaging sensors and digital projection technology are providing new avenues for the generation of powerful high-speed 3D surface imaging systems. Compared with high-speed cameras running at tens of thousands fps or even faster, however, projectors normally operate at a much lower rate that is often around 120 fps when gray-scale patterns are projected. Therefore, the defocusing techniques are developed, with which quasi-sinusoidal fringe patterns can be projected at the maximum allowed frame rate (typically more than 1000 fps) with binary dithering techniques and lens-defocused digital light processing projectors [31,32]. Once the limitation of the system hardware is overcome, the major concern focuses on the imaging theory, for which the key is to reduce the number of images required for a single 3D reconstruction. Intuitively, spatial-multiplexing or one-shot techniques, e.g., Fourier transform based profilometry (FT) [33–36], windowed Fourier transform technique (WFT) [37], wavelet transform technique [38], and intensity-correlation-based methods [39,40], are very suitable for scanning moving objects. As the codification can be condensed into a single pattern, these methods have ideal efficiency for high-speed 3D surface imaging. However, their spatial resolution and depth accuracy are not high for discontinuities, e.g., object edges, due to the inherent hypothesis of the continuity and the smoothness for local areas in these methods.

For high-accuracy 3D surface imaging, researchers typically prefer time-multiplexing or multi-shot techniques that can benefit from abundant information collected temporally. Some techniques project many patterns of random intensity to implement active high-speed stereo-vision 3D measurements [41,42]. However, the 3D reconstructions tend to compromise for rapidly moving objects since a relatively long sequence of images (usually > 9 frames) is required to extract a single 3D frame. In contrast, the phase-shifting profilometry (PSP) [43], which is one of the most widely used multi-shot approaches, can produce accurate 3D reconstructions by projecting a small-scale set of phase-shifting fringe images (minimum three images). Nevertheless, it is still sensitive to motion even with the minimum images. The reason is the object motion violates the nominal phase shifts of the raw fringe patterns, leading to artificial ripples on reconstructed surfaces [44]. Besides, the motivation to remove the phase ambiguity due to the periodic nature of sinusoids is also a challenge for time-critical PSP applications, which can easily double or even triple the size of the image sequence [45].

To reduce the size of the image sequence (captured in the time domain) while collecting comparable amount of information, some researchers suggest strengthening the encoding capability in the space domain. To reduce the images for phase unwrapping, one can have more than one viewpoints, e.g., using more cameras to capture structured-light patterns. Benefiting from the geometric constraint, the methods can discriminate the fringe order without capturing extra images [46–49]. But, the weakness is that the structure of the imaging system would become complex. Also, the cost would increase significantly because of the use of additional high-speed cameras. Alternatively, without resorting to more viewpoints, the spacial coding strategy can also be introduced into the time-multiplexing techniques by condensing two images into a single one or reusing the existing patterns with more than one purpose [50–54]. These approaches can remove the phase ambiguity without greatly increasing the projected images, but would suffer in the process of phase unwrapping when the projected fringe is very dense [55]. Recently, micro FTP (μFTP) was developed to measure 3D profiles for transient scenes at 10,000 fps [32]. Although the dynamic 3D shapes can be recovered from dense fringe patterns, several uniform images (i.e., pure white images) have to be projected along with the structured-light patterns for robust phase retrieval. Thus, the size of overall image sequence is still relatively large, making the 3D imaging sensitive to fast moving objects.

In this work, we present a novel micro deep learning profilometry (μDLP), which enables high-quality 3D shape reconstructions for transient scenes. The micro means small values for both the frequency variations and periods of fringe patterns, allowing highly-accurate phase measurement and high resistance to the global illumination. Deep learning is a powerful machine learning technique that has shown great success in numerous imaging and computer vision applications [56–61]. Thanks to the strength of machine learning, the proposed method shows superiority in three aspects to the state-of-art methods. The first one is the high efficiency. The phase information can be extracted from a single image via a properly trained neural network. Compared with μFTP, it only uses half of the images to obtain a 3D image. Then, the second advantage is the high-quality phase measurement. As indicated by our experiments, the phase error of μDLP is only one-third of those of FT and WFT and is almost half of that of μFTP. Further, with only three images our method can nearly reproduce the ground-truth 3D result that is calculated with the multi-shot phase-shifting method that uses 36 images. Last, the proposed method is easy to use. Different from Fourier-transform-based methods in which the phase measurement deeply relies on the fine tuning of parameters, e.g., the window size in FT, the sigma, the sampling intervals, and the frequency threshold in WFT, the presented μDLP is fully automatic once the neural network has been trained, which means the exhaustive search for the optimal parameters can be avoided. Experiments demonstrate that μDLP is a powerful high-speed 3D surface imaging approach that can reconstruct high-accuracy 3D shapes for transient scenes at 20,000 fps.

2. Theory

2.1. Phase retrieval through a deep neural network

In μDLP, the fringe image is captured with a system of structured light illumination, which consists of a projector and a camera typically. According to the schematic shown in Fig. 1, the projector emits a fringe image onto the measured object to encode the illuminated surface. The camera captures the image from a different viewpoint, from which the

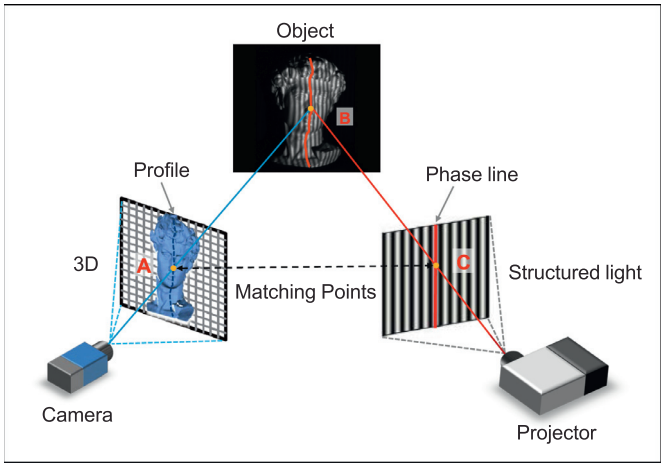

Fig. 1. Schematic of 3D surface imaging by structured light illumination.

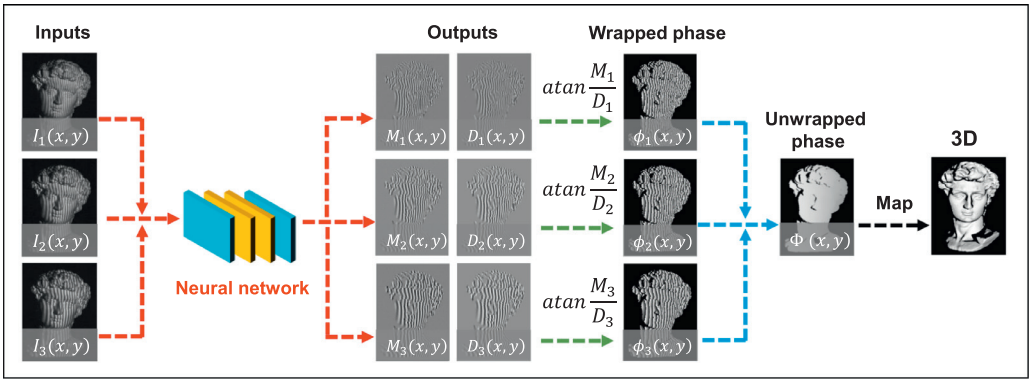

Fig. 2. Schematic of the proposed μDLP. With a few fringe patterns $I_1(x, y)$, $I_2(x, y)$, and $I_3(x, y)$, the neural network predicts the numerator $M_t(x, y)$ and the denominator $D_t(x, y)$ for each input fringe image. These intermediate results are then fed into the arctangent function to calculate the phase distribution $\phi_t(x, y)$. After phase unwrapping, an unwrapped absolute phase map $\Phi(x, y)$ is obtained and is further converted into the 3D reconstruction.

stripes are observed with distortion due to the depth variation of the object. The phase is then calculated from the captured fringe image, which works as a cue to compute the 3D information.

During the image projection, μDLP exploits several fringe patterns with slightly different wavelengths or fringe pitches $\{\lambda_1, \lambda_2, \ldots, \lambda_T\}$. For rapid projection, the sinusoidal patterns are generated in the binary mode and are projected by a defocused projector [62,63]. The wavelengths of projected patterns are carefully chosen by considering: First, the selected λ is supposed to be small enough, i.e., the frequency should be sufficiently high for high-quality phase retrieval. Second, the least common multiple (*LCM*) of the wavelengths should be larger than the horizontal or vertical resolution of the projector so that the phase ambiguity can be removed properly. In this work, we project vertical fringes, which means $LCM(\lambda_1, \lambda_2, \ldots, \lambda_T)$ should be greater than the width of projection plane. With the determined wavelengths, the intensity of projected patterns can be written as

$$I_t^p(x^p, y^p) = a + b\cos\left(\frac{2\pi x^p}{\lambda_t}\right) \tag{1}$$

where (x^p, y^p) is the pixel coordinate of the projector, and $t = 1, 2, \ldots, T$. Parameters a and b are the mean value and the amplitude, respectively.

Then, the generated patterns are projected and captured sequentially. The intensity of captured images can be represented as

$$I_t(x, y) = A(x, y) + B(x, y)\cos\phi_t(x, y) \tag{2}$$

where (x, y) is the pixel coordinate of the camera, $A(x, y)$ the background intensity, $B(x, y)$ the modulation, and $\phi_t(x, y)$ the phase to be recovered. In most phase measurement techniques, the wrapped phase map is often retrieved from an inverse trigonometric function:

$$\phi_t(x, y) = \arctan\frac{M_t(x, y)}{D_t(x, y)} = \arctan\frac{cB(x, y)\sin\phi_t(x, y)}{cB(x, y)\cos\phi_t(x, y)} \tag{3}$$

where $M_t(x, y)$ and $D_t(x, y)$ denote the numerator and the denominator of the arctan function, respectively. c is a constant that depends on the phase demodulation algorithm, e.g., $c = 0.5$ for FT and $c = \frac{N}{2}$ for N-step PSP.

To realize the process of phase retrieval with machine learning, we construct a deep convolutional neural network. As mentioned above, we prefer small sets of fringe images for high-speed 3D surface imaging. However, one or more assistant phase maps are required for robust phase unwrapping of dense fringe pattern [45]. Thus, we have a balance by totally employing three fringe patterns (i.e., $T = 3$) for 3D imaging that can produce three phase maps, one of which is used for 3D reconstruction and the rest for reliable phase unwrapping. Fig. 2 demonstrates the schematic of the proposed method. The neural network is trained to predict the numerator $M_t(x, y)$ and the denominator $D_t(x, y)$ for each input image $I_t(x, y)$. Each pair of numerator and denominator $\{M_t(x, y), D_t(x, y)\}$ is then fed into the arctangent function (Eq. (3)) to obtain the wrapped phase map $\phi_t(x, y)$. Next, an unwrapped phase distribution $\Phi(x, y)$ is obtained by the temporal phase unwrapping algorithm based on projection distance minimization. Finally, the 3D surface is calculated from the absolute phase map with calibrated mapping parameters between the camera and the projector.

Note that we presented a machine-learning-based fringe analysis method [56] that employs two neural networks to calculate the phase information. For applications of transient 3D measurements, some improvements have been made in this work. First, μDLP uses only one network for the phase retrieval, thus easing the learning process and saving the time cost of the training process. To compensate the influence of the absence of the background intensity, a more powerful three-scale data processing architecture is developed here to perceive the surface details and learn the phase extraction. Moreover, the neural network in μDLP can learn fringe patterns of different frequencies simultaneously and output the intermediate results for corresponding fringe patterns, which improves the measurement efficiency of the phase and 3D contours.

Fig. 3 shows the internal structure of the neural network in μDLP. The labeled dimension of each layer or block indicates the size of the output data. The inputs of the network are the fringe images $\{I_1(x, y), I_2(x, y), I_3(x, y)\}$. The size of each input image is $W \times H$ pixels, where W is the width and H is the height. Three data-flow paths are constructed to process the input images at different scales. In the first path which keeps the original size of input data, the fringe images are successively processed by a convolutional layer, a group of residual blocks and another convolutional layer. C is the number of filters used in the convolutional layer and equals the number of channels of output data. Each filter is used to extract a feature map (channel) for the output tensor. The same input data also undergoes similar but more sophisticated procedures in the second and the third paths where the data are first downsampled by $\times 2$ and $\times 4$ for high-level perceptions and then upsampled to match the original dimensions. Eventually, the results of each data-flow path are concatenated to produce the final outputs that feature three pairs of $\{M(x, y), D(x, y)\}$ corresponding to every input image $I_t(x, y)$. With the design of multi-scale data-flow paths, geometric details that the input images contain can be perceived precisely, ensuring the estimation of high-quality phase information. Note that it is difficult to output the wrapped phase directly with the input of the fringe image, since the sharp discontinuity at the 2π jump is hard to learn by the neural network. Therefore, for high-accurate phase estimations, the deep neural network is trained to calculate the intermediate results that vary continuously in space, i.e., the numerator and the denominator. Further details about the architecture of the network are provided in Appendix A.

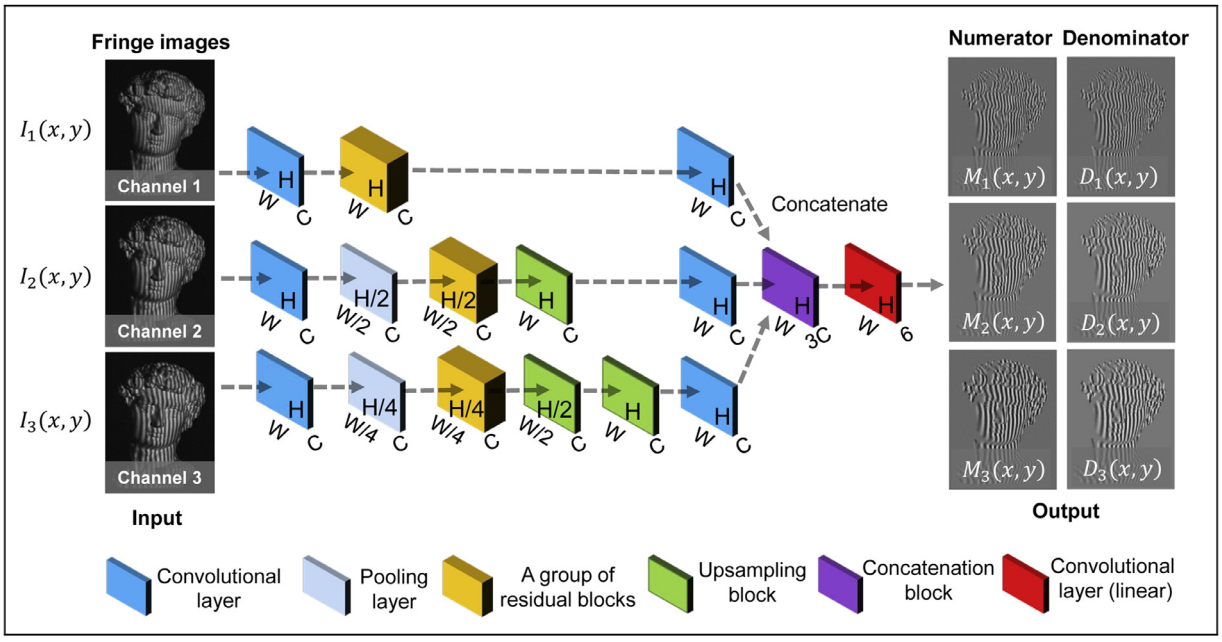

Fig. 3. Architecture of the proposed multi-scale deep neural network. The input data have three channels containing the three fringe images. The neural network has three data-flow paths that involve different kinds of layers/blocks, which can process the input data at different scales and extract useful information with downsampling rates of $\times 1$, $\times 2$ and $\times 4$, respectively. The outputs of the network are three pairs of numerator and denominator that correspond to each fringe pattern.

2.2. Phase unwrapping and 3D reconstruction

After feeding the estimated pair of numerator and denominator into Eq. (3), μDLP calculates wrapped phase maps $\phi_t(x, y)$ for each input fringe image. To remove the phase discontinuity of $\phi_t(x, y)$, we use the temporal phase unwrapping approach based on the projection distance minimization [32]. Given a vector of wrapped phase $\varphi = (\phi_1, \phi_2, \ldots, \phi_T)^{Trs}$ of the pixel (x, y), where Trs means the transposition, the vector of corresponding unwrapped phase $\mathbf{\Phi} = (\Phi_1, \Phi_2, \ldots, \Phi_T)^{Trs}$ can be expressed as

$$\mathbf{\Phi} = \boldsymbol{\varphi} + 2\pi\mathbf{k} \tag{4}$$

where $\mathbf{k} = (k_1, k_2, \ldots, k_T)^{Trs}$ is the vector of integer fringe order that we calculate for phase unwrapping. By taking the wavelengths into account, we have the following relationship

$$\Phi_1\lambda_1 = \Phi_2\lambda_2 = \ldots = \Phi_T\lambda_T \tag{5}$$

Eq. (5)reveals that the unwrapped phase $\mathbf{\Phi} = (\Phi_1, \Phi_2, \ldots, \Phi_T)^{Trs}$ of each pixel forms a line in space R^T. Therefore, with the mentioned constraint that $LCM(\lambda_1, \lambda_2, \ldots, \lambda_T) > W^p$ where W^p is the width of projection plane in pixel, there will be a unique qualified fringe order vector $\mathbf{k}$ that corresponds to the measurement range. In theory, the unwrapped phase $\mathbf{\Phi}$ of each pixel would align perfectly along the line expressed by Eq. (5). However, the unwrapped phase often scatters around the line due to the effects of random noise and non-sinusoidal fringe intensity in reality. Therefore, the distance between each candidate unwrapped phase and its projection onto this line is calculated. The desired $\mathbf{\Phi} = (\Phi_1, \Phi_2, \ldots, \Phi_T)^{Trs}$ is determined when the distance is minimized.

As a group of unwrapped phase maps is obtained after phase unwrapping, one of them is selected as $\Phi(x, y)$ for the 3D reconstruction. In the perspective of the camera, given the point (x^w, y^w, z^w) of test object is imaged by pixel (x, y), we have the following projection relationship in homogeneous coordinates

$$s^c\begin{pmatrix}x\\y\\1\end{pmatrix} = P^c\begin{pmatrix}x^w\\y^w\\z^w\\1\end{pmatrix} = \begin{pmatrix}p^c_{11} & p^c_{12} & p^c_{13} & p^c_{14}\\ p^c_{21} & p^c_{22} & p^c_{23} & p^c_{24}\\ p^c_{31} & p^c_{32} & p^c_{33} & p^c_{34}\end{pmatrix}\begin{pmatrix}x^w\\y^w\\z^w\\1\end{pmatrix} \tag{6}$$

where s^c is a scaling factor, P^c is the projection matrix of camera that is the product of the extrinsic parameter matrix and the intrinsic parameter matrix of the camera. In the other perspective of projector, there is a similar process when the projector is considered as an inverse camera

$$s^p\begin{pmatrix}x^p\\y^p\\1\end{pmatrix} = P^p\begin{pmatrix}x^w\\y^w\\z^w\\1\end{pmatrix} = \begin{pmatrix}p^p_{11} & p^p_{12} & p^p_{13} & p^p_{14}\\ p^p_{21} & p^p_{22} & p^p_{23} & p^p_{24}\\ p^p_{31} & p^p_{32} & p^p_{33} & p^p_{34}\end{pmatrix}\begin{pmatrix}x^w\\y^w\\z^w\\1\end{pmatrix} \tag{7}$$

where s^p is a scaling factor, P^p is the projection matrix of projector that is the product of the extrinsic parameter matrix and the intrinsic parameter matrix of the projector. Given the unwrapped phase of this pixel is Φ, the relationship between the camera pixel and its corresponding projector pixel can be expressed by

$$\Phi(x, y) = \frac{2\pi}{\lambda}x^p \tag{8}$$

Thus, the 3D coordinate can be calculated by combing Eqs. (6) and (7), giving

$$\begin{pmatrix}x^w\\y^w\\z^w\end{pmatrix} = \begin{pmatrix}p^c_{11} - p^c_{31}x & p^c_{12} - p^c_{32}x & p^c_{13} - p^c_{33}x\\ p^c_{21} - p^c_{31}y & p^c_{22} - p^c_{32}y & p^c_{23} - p^c_{32}y\\ p^c_{11} - p^c_{31}x^p & p^c_{12} - p^c_{32}x^p & p^c_{13} - p^c_{33}x^p\end{pmatrix}^{-1}\begin{pmatrix}p^c_{34}x - p^c_{14}\\ p^c_{34}y - p^c_{24}\\ p^c_{34}x^p - p^p_{14}\end{pmatrix} \tag{9}$$

The projection matrices of the camera and the projector can be obtained with the system calibration [46]. Note that gigabyte-scale image data are often recorded in applications of high-speed imaging. Although the 3D reconstruction can be carried out off-line, the time cost would be still very high. To increase the calculation speed, we suggest Eq. (9) to be implemented with a graphics processing unit [64] or several look-up tables [65], which can greatly save the time cost of the 3D reconstruction.

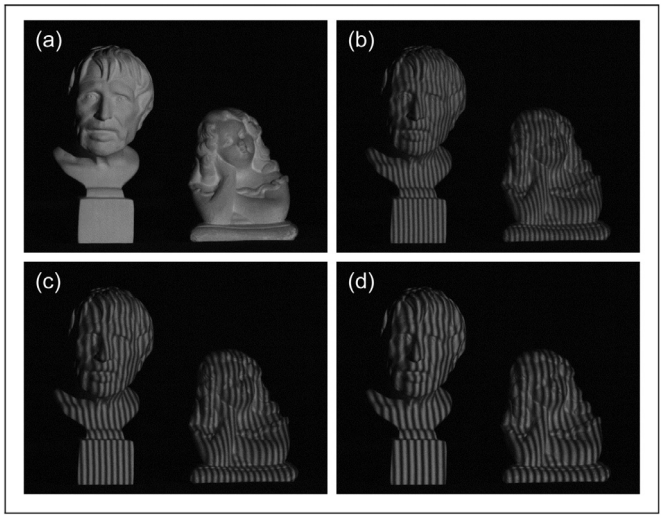

Fig. 4. Testing the trained network using a scene which is not present in the training phase. (a) The measured scene; (b) fringe image $I_1(x, y)$ with $\lambda_1 = 9$; (c) fringe image $I_2(x, y)$ with $\lambda_2 = 11$; (d) fringe image $I_3(x, y)$ with $\lambda_3 = 13$.

3. Experiments

To validate the proposed method, we built a structured light illumination system that consisted of a projector (DLP 4100, Texas Instruments) with resolution of 1024×768 and a high-speed camera (V611, Vision Research Phantom) with resolution of 640×440 and with pixel depth of 8 bits. The camera equipped with a lens of 24 mm focal length. The distance between the test object and the imaging system was about 1.5 m. The wavelengths of projected images were selected as $\{\lambda_1 = 9, \lambda_2 = 11, \lambda_3 = 13\}$, which provided unambiguous 3D reconstructions for the whole projection range (i.e., $LCM(9, 11, 13) = 1287 > 1024$).

The implementation of μDLP has two steps: training and testing. In the training stage, the training data were collected from different scenes. Analogous to traditional approaches of structured light illumination that require fringes with enough signal-to-noise ratio or without saturated pixels, μDLP also prefers the training objects without very dark or shiny surfaces. Otherwise, the training process would be damaged, since it is hard to obtain reliable ground truth data for these objects. Here, our training data set was collected from 45 scenes. With the 12-step phase-shifting method, we captured 1620 different fringe patterns and their corresponding ground-truth data for each wavelength (see Appendix B for more details on the collection of the training data). The neural network was implemented using TensorFlow framework (Google) and was computed on a GTX Titan graphics card (NVIDIA). To monitor during training the accuracy of the neural network on the data that it has never seen before, we created a validation set including 120 fringe images from 10 validation scenes which were separate from the training scenarios. With 120 epochs of training, the training loss and the validation loss of the network converged. And there is not overfitting to our training dataset. We provide further details of the training results in Appendix A.

3.1. The performance of μDLP for static scene

To test the performance of the trained neural network, we measured a static scenario that includes two isolated plaster models, as shown in Fig. 4(a). Note that our neural network never sees these models in the training stage. Fig. 4(b)–(d) are the captured fringe images $I_1(x, y)$, $I_2(x, y)$, and $I_3(x, y)$, respectively. With these images, the trained neural network predicted the numerator and the denominator for each of the input fringe image. The results are shown in the first two columns of Fig. 5. The estimated numerators and denominators were then fed into Eq. (3) to calculate the wrapped phase maps that are shown in the third column of Fig. 5. Finally, we calculated the unwrapped phase distributions that are displayed in the last column of Fig. 5. As we can see, the discontinuity have been removed completely for all of the wrapped phase.

We chose one of the unwrapped phase maps, i.e.,$\Phi_2(x, y)$, to investigate the quality of the phase estimated by μDLP. In the investigation, 12-step phase-shifting method was used to calculate a reference phase map which was unwrapped in the same way. Moreover, we also applied FT, WFT, and μFTP for comparison. Fig. 6 shows the phase error of each method. We can see the errors of WFT and FT are more significant than those of μFTP and μDLP. Further, μDLP shows better performance than μFTP due to less phase errors observed at the object edges. To compare the error maps in detail, we studied two recovered areas of complex surfaces, as can be seen in Fig. 7. The selected regions are the hair of the left model and the face of the right one. These two regions of interest (ROI) have rich details, which can be used to evaluate the capability of han-

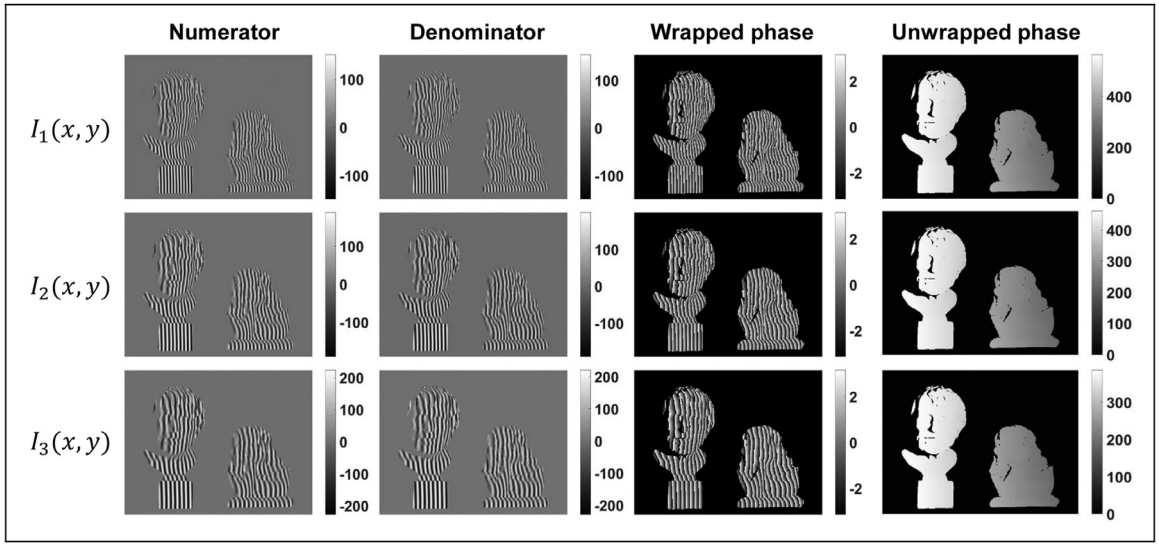

Fig. 5. Predicted results of the trained neural network. Each row shows the estimated numerator, denominator, wrapped phase, and unwrapped phase for each fringe image.

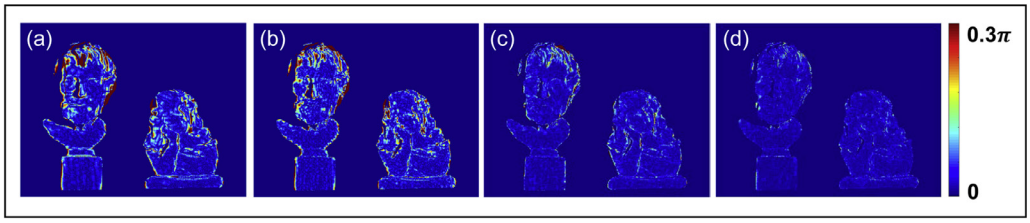

Fig. 6. Comparison of the phase error distribution for methods: (a) WFT, (b) FT, (c) μFTP, and (d) μDLP.

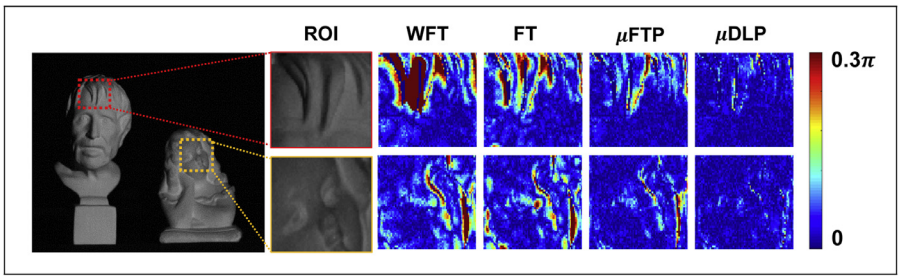

Fig. 7. Comparison of the phase error of two ROI. The first ROI is selected from the hair of the left model, and the second is picked from the face of the right model. The zoom-in phase error of different approaches are demonstrated for each region.

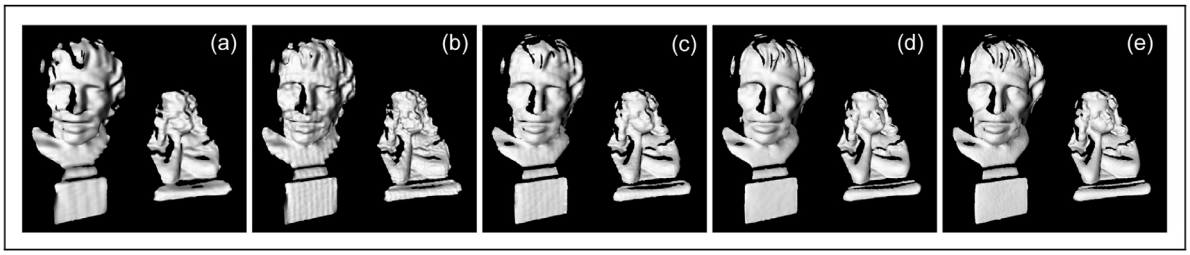

Fig. 8. 3D reconstructions of the methods: (a) WFT, (b) FT, (c) μFTP, (d) μDLP, and (e) 12-step phase-shifting method (ground truth).

Table 1
Quantitative comparison of the proposed μDLP with WFT, FT and μFTP in terms of MAE of unwrapped phase and the required number of images.

	WFT	FT	μFTP	μDLP
MAE (rad)	0.36	0.26	0.13	0.077
Images	3	3	6	3

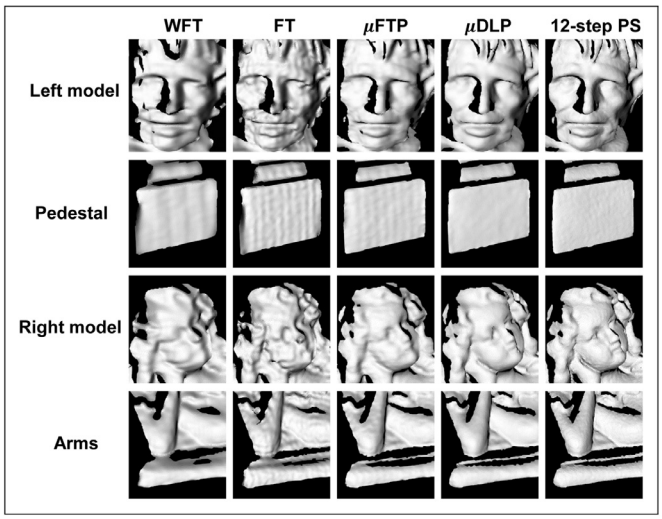

Fig. 9. Amplified views of the 3D reconstructions of four ROI: The face of the left model, the pedestal of the left model, the face of the right model, and the arms of the right model.

dling profiles with fine structures. In Fig. 7, we can observe WFT has the largest phase error, especially for the region of hair. By contrast, FT performed better than WFT as there are less errors at the reconstructed hair. But, it still failed to accurately retrieve the phase of the facial contour of the right model. In contrast to WFT and FT, μFTP shows increased but yet not high enough accuracy for these areas. As to μDLP, it has the least phase errors for both the hair of the left model and the details of the face of the right one. For quantitative evaluation, the mean absolute error (MAE) of unwrapped phase and the number of used images for the phase retrieval are shown in Table 1. Although the same images are used, the error of μDLP is smaller than one-third of those of WFT and FT. Compared with μFTP, μDLP only exploited half of the patterns while improved the phase accuracy by almost 50%.

Further, we converted the unwrapped phase maps into 3D rendered geometries, as shown in Fig. 8. Also, several ROI were selected for the detailed comparison. Fig. 9 shows the enlarged views of reconstructions of the face and the pedestal of the left model, and the face and the arms of the right model. From the result of WFT, the general profiles of these regions have been recovered but with significant loss of details compared with the reference that was reconstructed by 12-step phase-shifting method. From the 3D reconstruction of FT, the result features many grainy distortions that are mainly due to the inevitable spectra leakage and overlapping in the frequency domain. In contrast to WFT and FT, μFTP successfully retrieved some fine structures, e.g., the nose and the mouse of the right model. But, it still failed to preserve a few sharp edges. Finally, from the result of our method, we can see the deep-learning based approach yielded the highest-quality 3D reconstruction,

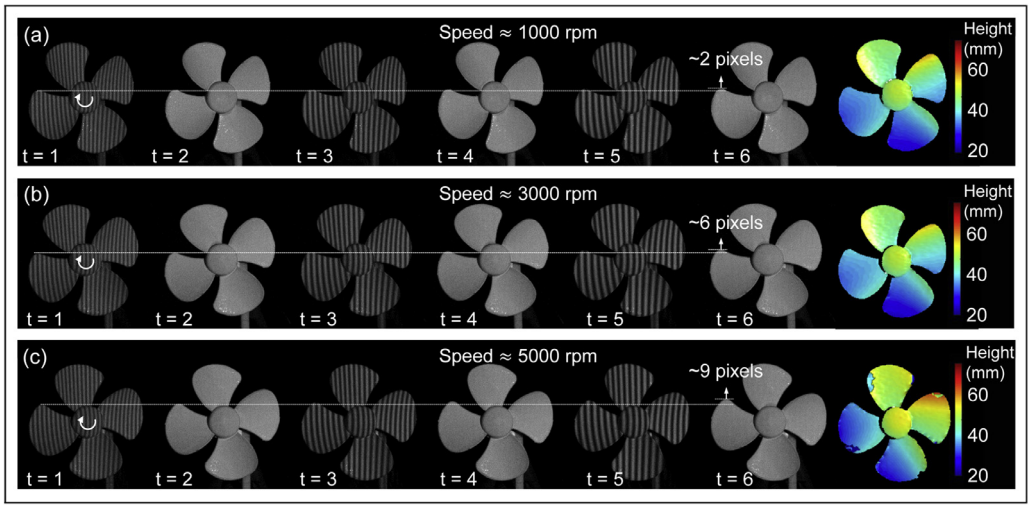

Fig. 10. 3D surface imaging of an electric fan rotating at different speeds by μFTP. (a)–(c) Images captured at 1000 rpm, 3000 rpm, and 5000 rpm with their corresponding 3D reconstructions.

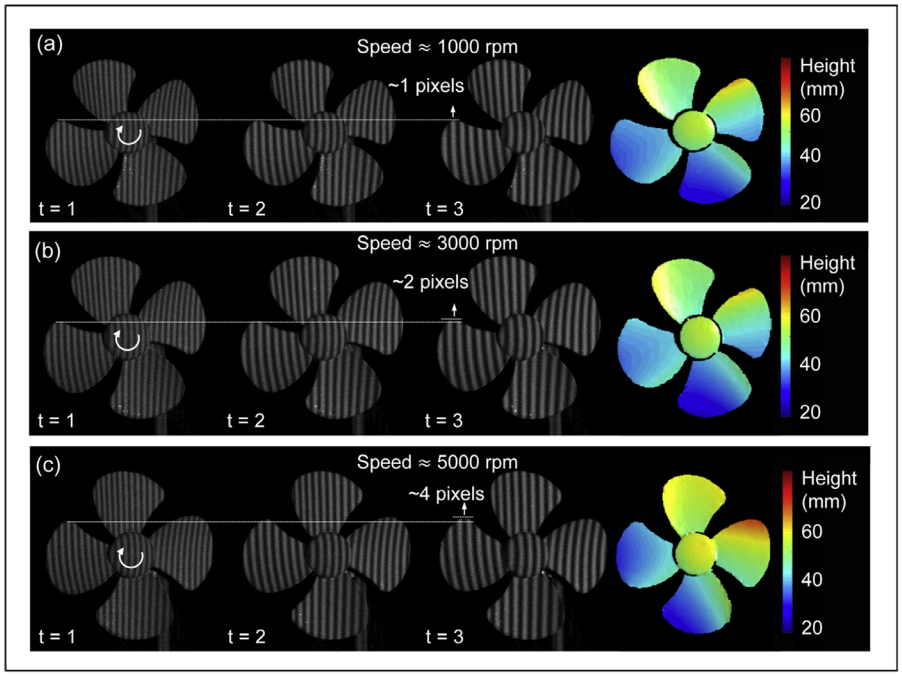

Fig. 11. 3D surface imaging of an electric fan rotating at different speeds by the proposed μDLP. (a)–(c) Images captured at 1000 rpm, 3000 rpm, and 5000 rpm with their corresponding 3D reconstructions.

which almost reproduced the reference 3D model. It is worthwhile to mention that only three images were used in our method while 12×3 images were employed by the 12-step phase-shifting method. This experiment verifies that μDLP can produce high-fidelity phase measurements and 3D reconstructions, and is superior to the state-of-art high-speed 3D surface imaging approaches regarding the accuracy and efficiency.

3.2. The performance of μDLP for dynamic scene

We measured an electric fan rotating at a high speed to show μDLP's performance of handling fast rotating objects. The radius of the fan is about 50 mm. For comparison, we also used the μFTP to test the same scene. By tuning the input current (from 0.3A to 5A), we let the fan rotate from 1000 rotations per minute (rpm) to 5000 rpm. Fig. 10 shows the images captured by μFTP and the corresponding surface reconstructions when the fan rotated at about 1000 rpm, 3000 rpm, and 5000 rpm, respectively. During the tests, the fan rotated clockwise, and the system kept capturing the images at 20,000 fps for both approaches. As the phase information was extracted from a pair of images (a fringe image and a plain image) in μFTP, it reconstructed the 3D surface at 10,000 fps. In Fig. 10, we can observe that within a period of 3D reconstruction the left blade shifted upward about two and six pixels respectively with the rotating speed of 1000 rpm and 3000 rpm. Under these conditions, μFTP successfully measured the contour of the blades. However, when the fan accelerated to 5000 rpm, several areas were retrieved with many errors as can be observed from Fig. 10(c). The reason lies in the fact that μFTP exploited six images to reconstruct a single 3D frame. When the speed reached up to 5000 rpm, the left blade moved $\sim$9 pixels during the capture of the six images. Because of the long period of the image capture, the 3D reconstruction becomes fragile for the object motion.

By contrast, μDLP can reconstruct 3D shapes at 20,000 fps with the fact that the height-related phase was measured from a single fringe image. Fig. 11 shows the captured images and the corresponding recovered 3D results of μDLP. Although the speed increased to 5000 rpm, our

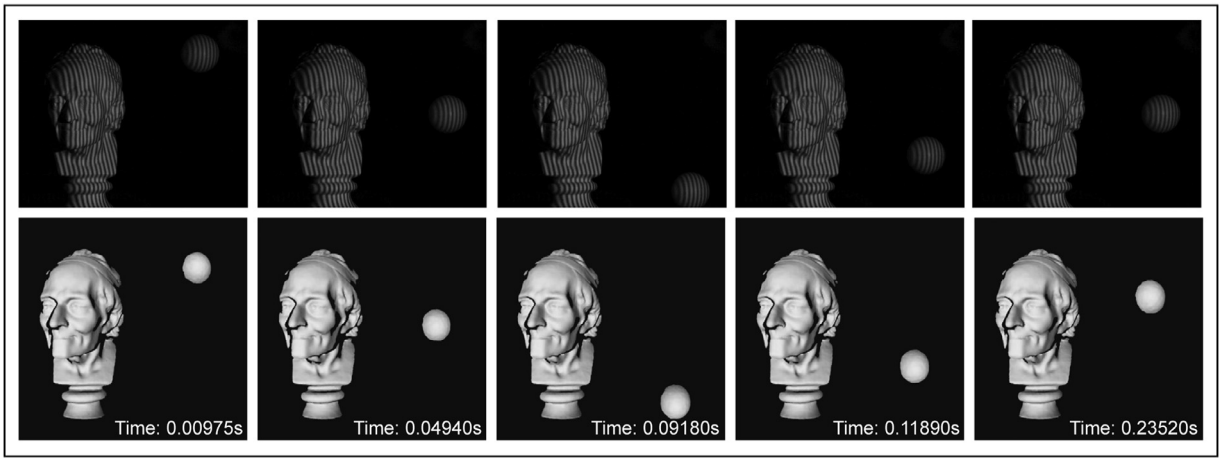

Fig. 12. Measurement of a dynamic scene that includes a static model and a falling table tennis, which are also not present in the training process. The first row shows captured fringe images at five different moments, and the second the corresponding 3D reconstructions obtained through μDLP.

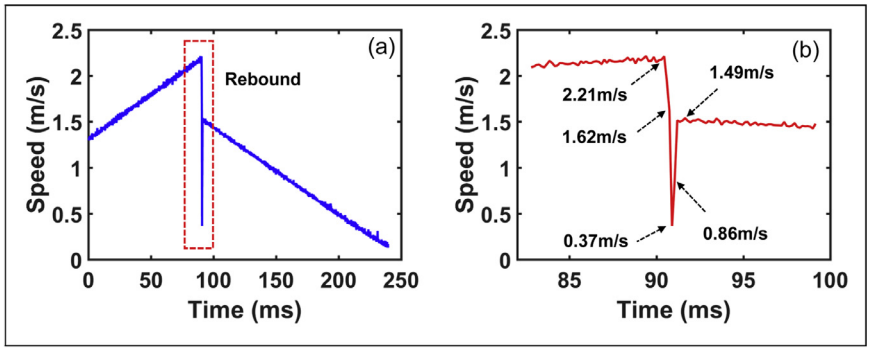

Fig. 13. Investigation of the speed of the table tennis. (a) The speed of the table tennis during the fall; (b) the amplified view of the red box in (a) showing the change of speed at the moment when the sphere hit the ground.

method can still measure the surface robustly. As fewer images were used by μDLP, the motion caused a shift of merely about 4 pixels as can be seen in Fig 11 (c), which did not affect the 3D reconstruction. From this experiment, thanks to the powerful computational capability of machine learning, the number of images can decrease significantly, which is favorable for overcoming the influence of object motion and dealing with fast moving objects.

Then, another dynamic scene was measured to further validate μDLP's capability of handling transient events. The scene consisted of a static plaster model and a falling table tennis. During the measurement, the fringe patterns were projected repeatedly onto the scene and the camera was synchronized with the projector at 20,000 fps. The first row of Fig. 12 shows the captured fringe images $I_2(x, y)$ at five different moments. We can see in this transient process the table tenuis gradually fell to the lowest point, and then bounced after hitting the ground. The dynamic process was retrieved by μDLP and is shown in Visualization 1. The 3D images corresponding to the selected moments are displayed in the second row of Fig. 12. We can observe that both the static model and the dropping sphere have been faithfully reconstructed with the deep-learning based technique.

Further, we analyzed the velocity of the falling table tennis using the retrieved geometry. First, the 3D point cloud of the table tennis was fitted to the function of sphere. Then, we estimated the center of the sphere, and calculated the speed by computing the displacement of the center between successive 3D frames. The velocity of the sphere during this transient event is shown in Fig. 13(a). As the measurement just started after the fall, the table tennis had an initial velocity which is about 1.36 m/s. As time went on, it moved faster due to the acceleration of gravity. When the velocity reached the maximum, the sphere hit the ground. Fig. 13(b) shows the speed of the sphere before and after the rebound. We can see the table tennis had the maximum velocity of 2.21 m/s before the hitting the ground. The speed began to decrease sharply right away after the hit. Within about one millisecond, the velocity reduced to 1.62 m/s and 0.37 m/s. Then, the speed went up instantly to 1.49 m/s due to the elastic potential energy. We can see the speed at this moment is smaller than the previous maximum velocity. The reason could be the fact that some of the energy was consumed to overcome the damping effect during the energy conversion. Next, the table tennis gradually raised but with a diminishing speed until it reached a point where the velocity came close to zero. From the overall process, we can see it happened in less than 0.25 s. Although the time period is very short, μDLP reconstructed the 3D shape of the falling sphere accurately and analyzed the velocity successfully with the geometry information. This experiment demonstrates that μDLP can not only reconstruct 3D shapes of the dynamic objects but also be applied to the study of some key physical quantities of the transient events.

3.3. Quantitative evaluation of 3D reconstruction accuracy

Last but not least, we measured a pair of gauge spheres made from ceramic to demonstrate the accuracy of 3D reconstruction quantitatively. The shape of the gauge spheres have been calibrated by a coordinate measurement machine. Fig. 14(a) shows the tested spheres whose radii are 25.398 mm and 25.403 mm, respectively, and their center-to-center distance is 100.069 mm. With the proposed method, we computed the 3D point cloud and fitted the 3D points into the sphere model. The reconstructed result is shown in Fig. 14(b), where the "jet" colormap is used to represent data values of reconstruction errors. The radii of reconstructed spheres are 25.449 mm and 25.470 mm, with the deviations of 0.051 mm and 0.067 mm respectively. The measured center-to-center distance is 100.134 mm with the error of 0.065mm. Further, Figs. 14(c)

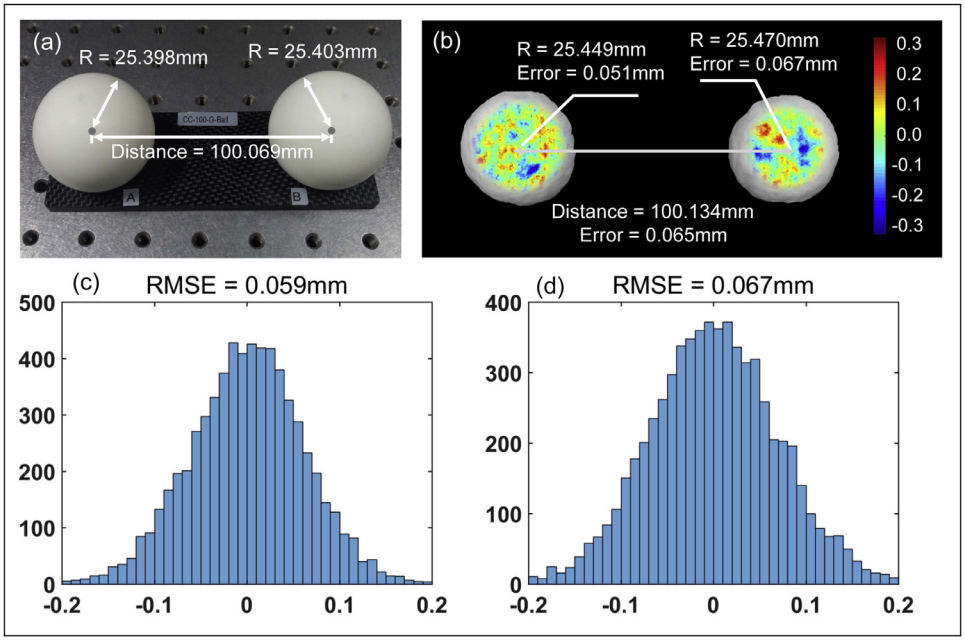

Fig. 14. Quantitative analysis of the reconstruction accuracy of μDLP. (a) Measured objects: a pair of gauge spheres; (b) 3D reconstruction with accuracy analysis; (c) histogram of the 3D error of sphere A; (d) histogram of the 3D error of sphere B.

and 14(d) show that the root-mean-square error (RMSE) of the spheres are 0.059 mm and 0.067 mm respectively. Since the measured shapes are very close to the ground truth, this experiment validates that our method can provide reliable phase information as well as high-accuracy 3D measurements.

4. Conclusion

In this work, we present a novel high-speed 3D surface imaging approach μDLP that can reconstruct dense and precise 3D shapes of transient events. Different from most of fast 3D imaging techniques using structured light illumination, μDLP can extract phase information from a single fringe image through a properly trained deep neural network. With only several fringe images of slightly different wavelengths, unambiguous high-quality 3D reconstructions can be obtained.

μDLP has three major advantages over the existing high-speed 3D imaging techniques. The first one is the high-accuracy phase retrieval. From our experiment, the phase error of μDLP is smaller than one-third of those of FT and WFT, and is almost half of that of μFTP. Moreover, μDLP can preserve details for fine structures or edges of test objects, resulting in the 3D reconstruction that is even comparable to that of 12-step phase-shifting method. Next, the second advantage of μDLP is the high efficiency. According to experimental results, μDLP leveraged only half of the patterns of μFTP but achieved nearly doubled phase precision. Also, μDLP used only three images to produce a high-quality 3D reconstruction that is close to that of 12-step phase-shifting method, by which, however, 36 fringe images were employed. Last, μDLP is easy to implement. Unlike the approaches based on Fourier transform, the performance of which heavily relies on tuning parameters, e.g., the window size for FT, the sigma, the sampling intervals, and the frequency threshold for WFT, μDLP is fully automatic and does not require a manual parameter search to optimize its performance once the neural network has been trained. Owing to these merits, μDLP can faithfully reconstruct 3D shapes of fast moving objects at 20,000 fps as demonstrated by the experimental result. The rate of 3D reconstruction can be further increased once more powerful equipment is in use. We believe the proposed μDLP could narrow the gap between the high-speed 3D imaging and the high-rate 2D photography, providing new insights for extensive studies and applications.

Funding

National Natural Science Foundation of China (61705105, 61722506, 11574152), National Key R&D Program of China (2017YFF0106403), Final Assembly “13th Five-Year Plan” Advanced Research Project of China (30102070102), Equipment Advanced Research Fund of China (61404150202), The Key Research and Development Program of Jiangsu Province (BE2017162), Outstanding Youth Foundation of Jiangsu Province (BK20170034), National Defense Science and Technology Foundation of China (0106173), “Six Talent Peaks” project of Jiangsu Province (2015-DZXX-009), “333 Engineering” Research Project of Jiangsu Province (BRA2016407, BRA2015294), Fundamental Research Funds for the Central Universities (30917011204, 30916011322, 30919011222).

Appendix A. Architecture and training of the neural network

The input fringe patterns are handled by three different data-flow paths, as demonstrated in Fig. 3. In the first path which keeps the original size of input data, the fringe images are successively processed by a convolutional layer, a group of residual blocks and another convolutional layer. Meanwhile, the same input data undergoes similar but more sophisticated procedures in the second and the third paths where the data are first downsampled by $\times 2$ and $\times 4$ for high-level perceptions and then upsampled to match the original dimensions. The downsampling is achieved through a max-pooling layer [66]. For each channel of the input, the pooling layer finds the maximum value in a 2×2 or 4×4 neighborhood. It then replaces the pixels in the 2×2 or 4×4 window with the found pixel of the maximum value. Therefore, the size of output is reduced by half/quarter for both the height and the width.

In the convolutional layers, the kernel size is 3×3 and the convolution stride is one. Zero-padding is used to control the spatial size of the output data, so that the input and output height and width are the same. The output of the convolutional layer is a three-dimensional (3D) tensor of shape (H, W, C), where H and W are the height and width in pixels of the input fringe pattern. C is the number of filters used in the convolutional layer and equals the number of channels of output data. Each filter is used to extract a feature map (channel) for the output tensor. Therefore with more filters, the convolutional network can perceive more details of measured surfaces. But the cost is that the network will

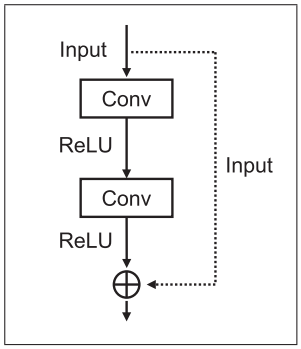

Fig. A1. Architecture of the residual block.

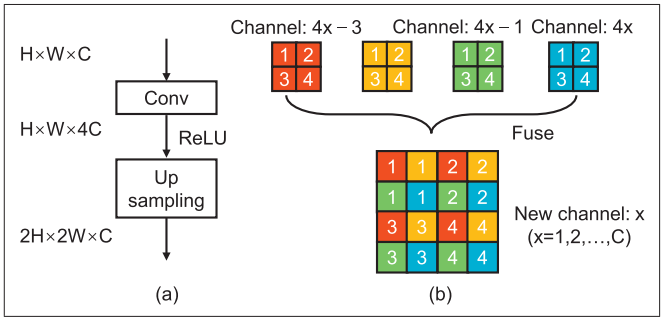

Fig. A2. (a) Architecture of the upsampling block; (b) diagram of the upsampling process.

consume more time during training. Thus, we have $C = 50$ filters in the work to achieve a balance. Except for the last convolutional layer which is activated linearly, the rest ones use the rectified linear unit (ReLU) as activation function, i.e., $ReLU(x) = max(0, x)$. Compared with other activation functions, e.g., sigmoid function [67], it has been demonstrated to enable better training of deeper networks [68].

In our network, we also used residual blocks whose architecture is shown in Fig. A1. The residual framework is composed of 2 sets of convolutional layer (Conv) activated by ReLU stacked one above the other [69]. It creates a shortcut between the input and output and can solve the degradation of accuracy as the network becomes deeper, thus easing the training process. To match the dimension of the original image, we upsample the output data from residual blocks using the upsampling block as shown in Fig. A2(a). The data first passes through a convolutional layer with ReLU activation. We then use quadruple filters to extract features from the input for providing rich information for the following upsampling, whose schematic is shown in Fig. A2(b). For the upsampled channel x, it is generated by original channels from $4x - 3$ to $4x$, thus allowing the output data with $\times 2$ spatial resolution. Next, the outputs of these three data flow paths are concatenated into a tensor with triple channels. Finally, the last convolutional layer yields a six-channel output datum which consists of three pairs of numerator $M(x, y)$ and denominator $D(x, y)$. The reason why we have the last convolutional layer to be linear is that the neural network is trained to predict the numerator and the denominator which can be negative.

To train the network, we minimize the mean-squared-errors of the output numerators and the output denominators with respect to the ground truth, which are obtained using the 12-step phase-shifting algorithm. The parameters of the network, i.e., the weights, bias and convolutional kernels, are trained using the backpropagation [70]. Thus, the loss function is computed as

$$Loss(\theta) = \frac{1}{H \times W} \sum_{t=1}^{3} \left(\left\| Y_t^M(\theta) - G_t^M \right\|^2 + \left\| Y_t^D(\theta) - G_t^D \right\|^2 \right) \tag{A1}$$

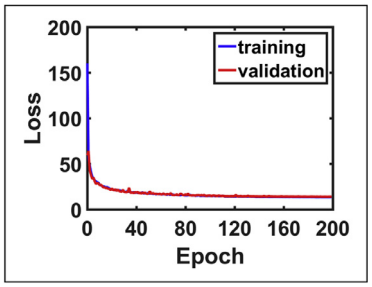

Fig. A3. Loss curve of the training and validation set for the neural network.

where G_t^M and G_t^D are the ground-truth numerator and denominator for the input fringe image I_t. $Y_t^M(\theta)$ and $Y_t^D(\theta)$ the numerator and denominator predicted by the network with the parameter space θ that includes the weights, bias and convolutional kernels.

During the training, the network uses the score of loss function as a feedback signal to adjust the parameters in θ by a little bit, in a direction that would lower the loss score. To this end, the adaptive moment estimation (ADAM) is used in our networks to tune the parameters for finding the minimum of the loss function [71]. In the implementation of ADAM, we start the training with a learning rate of 10^{-4}. We drop it by a factor of 2 if the validation loss has stopped improving for 10 epochs, which helps the loss function get out of local minima during training. To characterize the training, we plot the progression of the training and validation loss over training epochs, i.e., the number of iterations in the backpropagation over all of the dataset. Fig. A3 shows the loss curves converge after 120 epochs. From both curves, we can see there is not overfitting to our training dataset. As to the time cost, the training over 200 epochs took 3.16 hours.

Appendix B. Collection of training data

Prior to practical measurements, the developed neural network needs a training process in which the network learns to retrieve the phase. To obtain the ground-truth data used to train the neural network, we exploit the N-step phase-shifting method as it allows precise phase measurements. With this method, the captured phase-shifted fringe patterns with different wavelengths can be written as

$$I_n^t(x, y) = A(x, y) + B(x, y) \cos\left[\phi_t(x, y) - \delta_n\right] \tag{B1}$$

where $n = 0, 1, \ldots, N-1$ indicates the step of phase shift, and $t = 1, 2, 3$ implies the used wavelengths. δ_n is the phase shift that equals $\frac{2\pi n}{N}$. With the least square method, the ground-truth phase can be calculated by

$$\phi_t(x, y) = \arctan \frac{\sum_{n=0}^{N-1} I_n^t(x, y) \sin \delta_n}{\sum_{n=0}^{N-1} I_n^t(x, y) \cos \delta_n} \tag{B2}$$

According to Eq. (B2), the numerator and the denominator can be expressed as

$$M_t(x, y) = \sum_{n=0}^{N-1} I_n^t(x, y) \sin \delta_n \tag{B3}$$

$$D_t(x, y) = \sum_{n=0}^{N-1} I_n^t(x, y) \cos \delta_n \tag{B4}$$

Equations (B3) and (B4) are used to calculate the ground-truth numerator and denominator that are exploited to train the neural network.

In our experiments, three sets of 12-step phase-shifting fringe patterns with wavelengths $\{\lambda_1 = 9, \lambda_2 = 11, \lambda_3 = 13\}$ were generated according to Eq. (B1). These patterns were then projected onto different measured objects. The camera captured the reflected fringe patterns simultaneously at a different viewpoint and transferred them to our computer. In our experiment, we collected the training data from 45 different scenes including simple and complex objects. For each scene, we recorded 12×3 phase-shifting fringe patterns. Thus, 1620 fringe images were collected for all of the scenes. The captured training data are

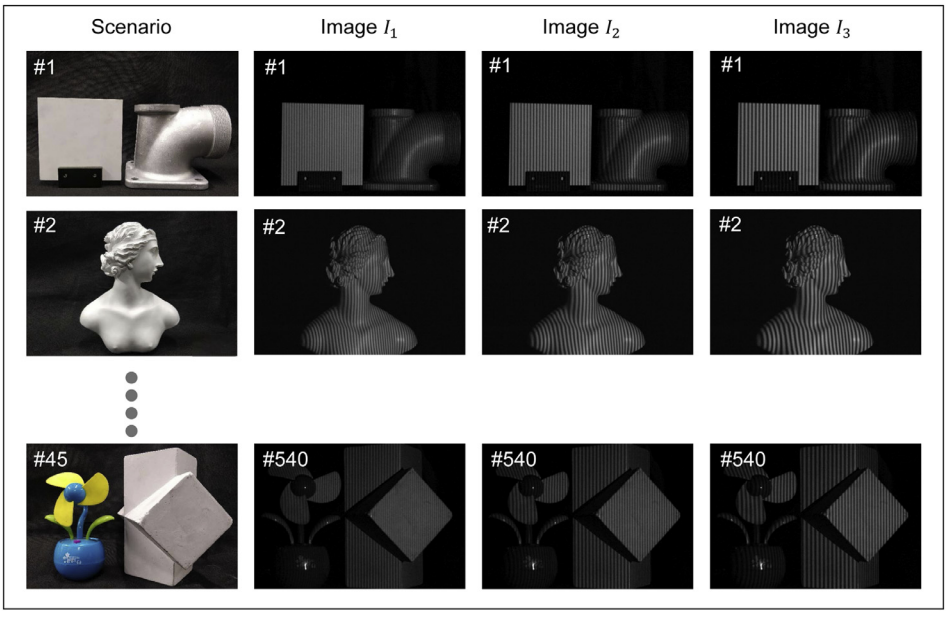

Fig. B1. The collected training data. The first column shows different tested scenarios. For each of them, we captured three sets of 12 phase-shifting fringe patterns and totally obtained 540 training input images for fringe images with three different wavelengths, as demonstrated in the second to the fourth column.

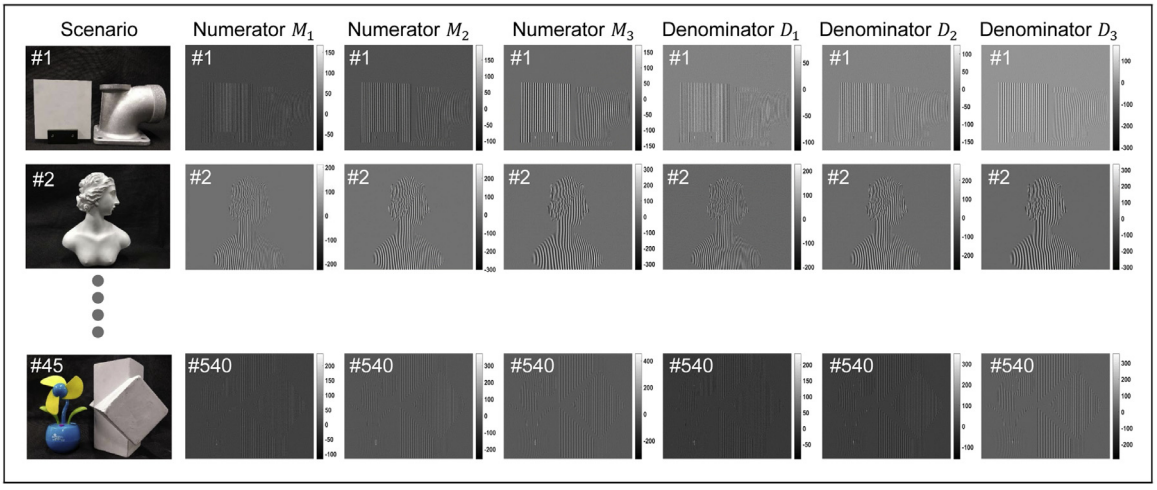

Fig. B2. Ground truth of the collected training data. The first column shows the tested scenarios. Within each set of fringe patterns of the same wavelengths, we calculated the ground-truth numerator and denominator by the 12-step phase-shifting algorithm. The second to the fourth columns displays the ground-truth numerator computed through Eq. (B3). The fifth to the seventh column shows the ground-truth denominator obtained through Eq. (B4).

demonstrated in Fig. B1. The first column shows the measured scenes. The second to the fourth column shows the captured fringe images with different wavelengths, respectively. Within each set of fringe patterns of the same wavelength, we calculated the corresponding ground-truth data by the 12-step phase-shifting algorithm. The results are shown in Fig. B2, where the second to the fourth column displays the ground-truth numerator, and the fifth to the seventh column shows the ground-truth denominator. It is noted that before being fed into the networks, the raw fringe images $\{I_1(x, y), I_2(x, y), I_3(x, y)\}$ were divided by 255 for normalization, which can make the learning process easier for the network. Moreover, for a preferable selection of training objects, one is suggested choosing objects without very dark or shiny surfaces to ensure captured fringe images with enough signal-to-noise ratio or without saturated points.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.optlaseng.2019.04.020 .

References

[1] Nicoletto G, Post D, Smith C. Moire interferometry for high sensitivity measurements in fracture mechanics. in: SESA/JSME Jt Conf Exp Mech; 1982. Oahu-Maui, HI
[2] Muybridge E.. The horse in motion. library of congress prints and photographs division. 2017.
[3] Stamper J, McLean E, Obenschain S, Ripin B, Thompson J, Luessen L. Fast electrical and optical measurements. New York: Martinus Nijhoff; 1986. 691
[4] Xing H, Zhang Q, Braithwaite CH, Pan B, Zhao J. High-speed photography and digital optical measurement techniques for geomaterials: fundamentals and applications. Rock Mech Rock Eng 2017;50(6):1611–59.
[5] Liang J, Wang LV. Single-shot ultrafast optical imaging. Optica 2018;5(9):1113–27.
[6] Goda K, Tsia K, Jalali B. Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. Nature 2009;458(7242):1145.
[7] Nakagawa K, Iwasaki A, Oishi Y, Horisaki R, Tsukamoto A, Nakamura A, Hirosawa K, Liao H, Ushida T, K Goda ea. Sequentially timed all-optical mapping photography (stamp). Nat Photonics 2014;8(9):695.
[8] Velten A, Lawson E, Bardagjy A, Bawendi M, Raskar R. Slow art with a trillion frames per second camera. In: in: ACM SIGGRAPH 2011 Talks. ACM; 2011. p. 44.
[9] Beurg M, Fettiplace R, Nam JH, Ricci AJ. Localization of inner hair cell mechanotransducer channels using high-speed calcium imaging. Nat Neurosci 2009;12(5):553.
[10] Gorkhover T, Schorb S, Coffee R, Adolph M, Foucar L, Rupp D, Aquila A, Bozek JD, Epp SW, Erk B. Femtosecond and nanometre visualization of structural dynamics in superheated nanoparticles. Nat Photonics 2016;10(2):93.

[11] 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.
[12] Ford KR, Myer GD, Hewett TE. Reliability of landing 3d motion analysis: implications for longitudinal analyses. Med Sci Sports Exerc 2007;39(11):2021–8.
[13] Jiang H, Zhao H, Li X. High dynamic range fringe acquisition: a novel 3-d scanning technique for high-reflective surfaces. Opt Lasers Eng 2012;50(10):1484–93.
[14] Malamas EN, Petrakis EG, Zervakis M, Petit L, Legat JD. A survey on industrial vision systems. ApplTools ImageVision Comput 2003;21(2):171–88.
[15] Pan B, Qian K, Xie H, Asundi A. Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. Meas Sci Technol 2009;20(6):062001.
[16] Cai Z, Liu X, Tang Q, Peng X, Gao BZ. Light field 3d measurement using unfocused plenoptic cameras. Opt Lett 2018;43(15):3746–9.
[17] Chen F, Brown GM, Song M. Overview of three-dimensional shape measurement using optical methods. Opt Eng 2000;39(1):10–22.
[18] Geng J. Structured-light 3d surface imaging: a tutorial. Adv Opt Photonics 2011;3(2):128–60.
[19] Xiong Z, Zhang Y, Wu F, Zeng W. Computational depth sensing: toward high-performance commodity depth cameras. IEEE Signal Process Mag 2017;34(3):55–68.
[20] Chen C, Gao N, Wang X, Zhang Z, Gao F, Jiang X. Generic exponential fringe model for alleviating phase error in phase measuring profilometry. Opt Lasers Eng 2018;110:179–85.
[21] Huang L, Idir M, Zuo C, Asundi A. Review of phase measuring deflectometry. Opt Lasers Eng 2018;107:247–57.
[22] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int J Comput Vis 2002;47(1–3):7–42.
[23] Cui Y, Schuon S, Chan D, Thrun S, Theobalt C. 3D shape scanning with a time-of-flight camera. In: in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE; 2010. p. 1173–80.
[24] Gonzalez-Jorge H, Rodríguez-Gonzálvez P, Martínez-Sánchez J, González-Aguilera D, Arias P, Gesto M, Díaz-Vilarino L. Metrological comparison between kinect i and kinect ii sensors. Measurement 2015;70:21–6.
[25] Smisek J, Jancosek M, Pajdla T. 3D With kinect. London: Springer London; 2013. 3–25
[26] Song S, Lichtenberg SP, Xiao J. Sun rgb-d: a rgb-d scene understanding benchmark suite. In: in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015.
[27] Apple iphone x. https://www.apple.com/iphone/">, Accessed: 24 December 2018.
[28] Oppo find x. https://www.oppo.com/en/smartphone-find_x/, Accessed: 24 December 2018.
[29] Jeught SVd, Dirckx JJ. Real-time structured light profilometry: a review. Opt Lasers Eng 2016;87:18–31.
[30] Zhang S. High-speed 3d shape measurement with structured light methods: a review. Opt Lasers Eng 2018;106:119–31.
[31] Lei S, Zhang S. Flexible 3-d shape measurement using projector defocusing. Opt Lett 2009;34(20):3080–2.
[32] 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. Opt Lasers Eng 2018;102:70–91.
[33] Su X, Chen W. Fourier transform profilometry: a review. Opt Lasers Eng 2001;35(5):263–84.
[34] Takeda M, Mutoh K. Fourier transform profilometry for the automatic measurement of 3-d object shapes. Appl Opt 1983;22(24):3977–82.
[35] Zhang Q, Su X. High-speed optical measurement for the drumhead vibration. Opt Express 2005;13(8):3110–16.
[36] Huang L, Kemao Q, Pan B, Asundi AK. Comparison of fourier transform, windowed fourier transform, and wavelet transform methods for phase extraction from a single fringe pattern in fringe projection profilometry. Opt Lasers Eng 2010;48(2):141–8.
[37] Kemao Q. Two-dimensional windowed fourier transform for fringe pattern analysis: principles, applications and implementations. Opt Lasers Eng 2007;45(2):304–17.
[38] Zhong J, Weng J. Spatial carrier-fringe pattern analysis by means of wavelet transform: wavelet transform profilometry. Appl Opt 2004;43(26):4993–8.
[39] Morita H, Yajima K, Sakata S. Reconstruction of surfaces of 3-d objects by m-array pattern projection method. In: in: 1988 Second International Conference on Computer Vision. IEEE; 1988. p. 468–73.
[40] Zhang Z. Review of single-shot 3d shape measurement by phase calculation-based fringe projection techniques. Opt Lasers Eng 2012;50(8):1097–106.
[41] Heist S, Lutzke P, Schmidt I, Dietrich P. P. kühmstedt, a. tünnermann, g. notni, high-speed three-dimensional shape measurement using gobo projection. Opt Lasers Eng 2016;87:90–6.
[42] Schaffer M, Grosse M, Harendt B, Kowarschik R. High-speed three-dimensional shape measurements of objects with laser speckles and acousto-optical deflection. Opt Lett 2011;36(16):3097–9.
[43] 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.
[44] Feng S, Zuo C, Tao T, Hu Y, Zhang M, Chen Q, Gu G. Robust dynamic 3-d measurements with motion-compensated phase-shifting profilometry. Opt Lasers Eng 2018;103:127–38.
[45] Zuo C, Huang L, Zhang M, Chen Q, Asundi A. Temporal phase unwrapping algorithms for fringe projection profilometry: a comparative review. Opt Lasers Eng 2016;85:84–103.
[46] 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.
[47] An Y, Hyun JS, Zhang S. Pixel-wise absolute phase unwrapping using geometric constraints of structured light system. Opt Express 2016;24(16):18445–59.
[48] Hyun JS, Chiu GTC, Zhang S. High-speed and high-accuracy 3d surface measurement using a mechanical projector. Opt Express 2018;26(2):1474–87.
[49] Li Z, Zhong K, Li YF, Zhou X, Shi Y. Multiview phase shifting: a full-resolution and high-speed 3d measurement framework for arbitrary shape dynamic objects. Opt Lett 2013;38(9):1389–91.
[50] Zuo C, Chen Q, Gu G, Feng S, Feng F. High-speed three-dimensional profilometry for multiple objects with complex shapes. Opt Express 2012;20(17):19493–510.
[51] Zuo C, Chen Q, Gu G, Feng S, Feng F, Li R, Shen G. High-speed three-dimensional shape measurement for dynamic scenes using bi-frequency tripolar pulse-width–modulation fringe projection. Opt Lasers Eng 2013;51(8):953–60.
[52] Zhang Y, Xiong Z, Wu F. Unambiguous 3d measurement from speckle-embedded fringe. Appl Opt 2013;52(32):7797–805.
[53] Wang Y, Liu K, Hao Q, Lau DL, Hassebrook LG. Period coded phase shifting strategy for real-time 3-d structured light illumination. IEEE Trans Image Process 2011;20(11):3001–13.
[54] Zhang Z, Towers CE, Towers DP. Time efficient color fringe projection system for 3d shape and color using optimum 3-frequency selection. Opt Express 2006;14(14):6444–55.
[55] Tao T, Chen Q, Feng S, Hu Y, Da J, Zuo C. High-precision real-time 3d shape measurement using a bi-frequency scheme and multi-view system. Appl Opt 2017;56(13):3646–53.
[56] Feng S, Chen Q, Gu G, Tao T, Zhang L, Hu Y, Yin W, Zuo C. Fringe pattern analysis using deep learning. Adv Photonics 2019;1(2):025001.
[57] Nah S, Kim TH, Lee KM. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: in: CVPR, vol. 1; 2017. p. 3.
[58] Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H, Ozcan A. Deep learning microscopy. Optica 2017;4(11):1437–43.
[59] Rivenson Y, Zhang Y, Günaydin H, Teng D, Ozcan A. Phase recovery and holographic image reconstruction using deep learning in neural networks. Light 2018;7(2):17141.
[60] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015;61:85–117.
[61] Sinha A, Lee J, Li S, Barbastathis G. Lensless computational imaging through deep learning. Optica 2017;4(9):1117–25.
[62] Floyd RW. An adaptive algorithm for spatial gray-scale. in: Proc Soc Inf Disp 1976;17:75–7.
[63] 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.
[64] Feng S, Chen Q, Zuo C. Graphics processing unit assisted real-time three-dimensional measurement using speckle-embedded fringe. Appl Opt 2015;54(22):6865–73.
[65] 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.
[66] Masci J, Meier U, Cireşan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In: in: International Conference on Artificial Neural Networks. Springer; 2011. p. 52–9.
[67] Cybenko G. Approximation by superpositions of a sigmoidal function, mathematics of control. SignalsSyst 1989;2(4):303–14.
[68] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: in: Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10. USA: Omnipress; 2010. p. 807–14.
[69] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: in: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. 770–778
[70] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436.
[71] Kingma D.P., Ba J.. Adam: a method for stochastic optimization. CoRR abs/1412.6980.