

Generative deep-learning-embedded asynchronous structured light for three-dimensional imaging

Lei Lu,^a Chenhao Bu,^a Zhilong Su,^{b,c,*} Banglei Guan,^d Qifeng Yu,^d Wei Pan,^e and Qinghui Zhang^a

^aHenan University of Technology, College of Information Science and Engineering, Zhengzhou, China

^bShanghai University, Shanghai Institute of Applied Mathematics and Mechanics, School of Mechanics and Engineering Science, Shanghai Key Laboratory of Mechanics in Energy Engineering, Shanghai, China

^cShaoxing Research Institute of Shanghai University, Shaoxing, China

^dNational University of Defense Technology, College of Aerospace Science and Engineering, Changsha, China

^eOPT Machine Vision Tech Co., Ltd., Department of Research and Development, Dongguan, China

Abstract. Three-dimensional (3D) imaging with structured light is crucial in diverse scenarios, ranging from intelligent manufacturing and medicine to entertainment. However, current structured light methods rely on projector–camera synchronization, limiting the use of affordable imaging devices and their consumer applications. In this work, we introduce an asynchronous structured light imaging approach based on generative deep neural networks to relax the synchronization constraint, accomplishing the challenges of fringe pattern aliasing, without relying on any *a priori* constraint of the projection system. To overcome this need, we propose a generative deep neural network with U-Net-like encoder–decoder architecture to learn the underlying fringe features directly by exploring the intrinsic prior principles in the fringe pattern aliasing. We train within an adversarial learning framework and supervise the network training via a statistics-informed loss function. We demonstrate that by evaluating the performance on fields of intensity, phase, and 3D reconstruction. It is shown that the trained network can separate aliased fringe patterns for producing comparable results with the synchronous one: the absolute error is no greater than 8 μm , and the standard deviation does not exceed 3 μm . Evaluation results on multiple objects and pattern types show it could be generalized for any asynchronous structured light scene.

Keywords: structured light; fringe pattern projection; asynchrony; deep learning; generative neural networks; three-dimensional imaging.

Received Jan. 19, 2024; revised manuscript received Jul. 3, 2024; accepted for publication Jul. 22, 2024; published online Aug. 13, 2024.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.AP.6.4.046004](https://doi.org/10.1117/1.AP.6.4.046004)]

1 Introduction

Structured light imaging is a key technology for acquiring three-dimensional (3D) information because of its advantages, such as nondestructiveness and high efficiency.^{1–3} Fringe pattern profilometry (FPP) is one of the most popular techniques implementing the structured light methods,^{4,5} showing powerful capabilities on the 3D sensing end in fields such as advanced manufacturing, medical inspection, and entertainment. A typical

FPP system consists of a projector and a camera,^{6,7} where the former projects a set of fringe patterns onto the object surface and then the latter captures the reflected fringe patterns from another view angle synchronously,⁸ as presented in Fig. 1(a). The captured fringe patterns are distorted due to the object height modulation, allowing for retrieving the 3D geometry by analyzing the distortion information encoded by phase values or binary coding.^{9,10} This sync-mechanism in existing FPP methods requires the camera to capture the image instantaneously after the projector projects the fringe pattern for avoiding fringe switching in the camera exposure time.

*Address all correspondence to Zhilong Su, szloong@shu.edu.cn

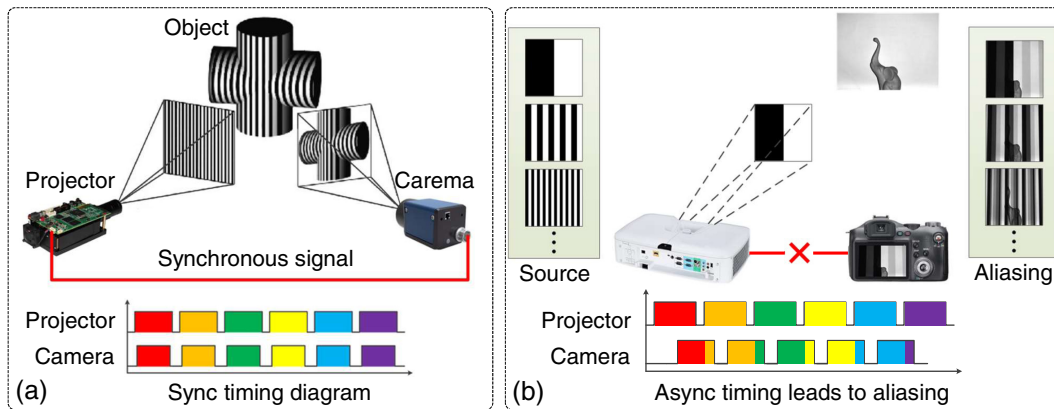


Fig. 1 Diagrams for (a) synchronous FPP and (b) asynchronous FPP systems.

When the synchronization is violated, the projector switches the fringe pattern in an unexpected way within the camera exposure period, leading to fringe pattern aliasing, as shown in Fig. 1(b), followed by reconstruction errors. To this end, several issues relevant to the synchronization of FPP systems should be paid attention to. Among them, the disadvantage of the high system complexity is the most notable one, since the dependence on the synchronization system highly relies on additional hardware or software coding.¹¹ In addition, it is challenging to synchronize the whole system with high accuracy, not least for the structured light array composed by multiple cameras and projectors.¹² On top of that, the limitations include its restricted applicability in specific scenarios, notably when the distance between the camera and projector exceeds certain thresholds, thereby hindering the proper functioning of the synchronous signal wire. Additionally, the system shows a susceptibility to electromagnetic interference, undermining its anti-interference capacity. Furthermore, equipment selection is constrained, as the employed camera and projector necessitate supplementary synchronization circuitry to accommodate the synchronous signal, which poses a challenge because many widely used consumer products lack this synchronization functionality.¹¹ In light of advancements in opto-electronic devices and imaging techniques, there has been a significant reduction in both the cost and dimensions of digital cameras. Consequently, imaging devices have become omnipresent in our daily lives. Notably, consumer products, including smartphones, tablets, and laptops, have attained remarkable levels of image resolution and frame rate, leading to a heightened potential demand for handy consumer-grade 3D reconstruction techniques, particularly those employing asynchronous FPP (async-FPP for short).

Almost all existing research employing FPP techniques is built upon the assumption of synchronization, resulting in the synchronization constraints being overlooked in the literature.^{13–15} Reference 16 classified the relationship between the camera and projector as hardware synchronization, soft synchronization, and no synchronization. The hardware synchronization utilizes a hardware circuit to transmit a trigger signal to achieve synchronization.^{17,18} When the projector completes a fringe pattern projection, the synchronization trigger signal is generated simultaneously and emitted to the camera side to invoke the instant image acquisition. The projector begins to project the next fringe pattern after the camera captures the current image.

It can achieve a high measurement rate of up to 5000 fps by the hardware-based synchronization.⁶ However, hardware circuits are only available in expensive industrial equipment, and affordable consumer products with popular applications often do not have modules or interfaces that can be used to achieve hardware synchronization. For the soft synchronization, a predefined waiting period is employed between the projection and acquisition processes.¹⁶ The projector projects one fringe pattern and then waits for a period for allowing the camera to capture the projected fringe pattern; the next fringe pattern is projected when the waiting period ends. To ensure the synchronization, the waiting period is always longer than the camera exposure time, leading to low frame rate and additional energy cost, which is advantageous for battery-powered or portable consumer devices.¹⁹ In addition, the synchronization controlled by coding does not work stably and accurately, leading to a potential risk of not being strictly synchronized. In asynchronous systems, the camera and projector do not have any synchronization mechanism but work independently to capture and project the fringe patterns, respectively, at a certain speed. Consequently, the projector can switch the projected fringe pattern within the exposure time of the camera. According to the frame rate relationship between the camera and projector, the asynchronous system can introduce duplicated captures, missed captures, and the toughest aliasing of multiple fringe patterns during the exposure time.

Though async-FPP-based 3D imaging has been a challenge, few works focus on this problem. Fujiyoshi et al. built a stereo vision system with a pair of asynchronous cameras,²⁰ where the two cameras captured images independently, to track the 3D positions of the object based on the Kalman filter and the most recent image in any of the cameras. Hasler et al. utilized multiple unsynchronized moving cameras to capture the motion of articulated objects.²¹ The static background and the positions of each camera are reconstructed based on structure-from-motion first; and then, the cameras are registered to each other using the background geometry. The audio signal is employed to achieve camera synchronization. Bradley et al.¹² proposed two approaches for solving the synchronization problem between camera arrays. The first approach is based on strobe illumination. It performs the exposure in each camera by first starting the cameras and then exciting the strobe lighting, thereby identifying the first synchronized frame. The second is based on optical flow vectors but is less accurate than the former. Moreno et al.¹¹

reported that the object shape can be reconstructed with asynchronous structured light system-based FPP. In the work, the binary fringe patterns were projected and captured independently at a constant speed, and an asynchronous decoding algorithm was introduced to generate a new image sequence equivalent to that obtained by the synchronized system, allowing for reconstructing the 3D model by adopting the existing binary code reconstruction algorithm. El Asmi et al.²² proposed an asynchronous scan method based on structured light by considering the captured image as a new reference pattern aliased from two consecutive projected patterns instead of regarding it as a partial exposure of them; with the local sensitive hashing algorithm, the first of the image sequences can be found to build the matching correspondence by utilizing the quadratic codes. In 2019, the same group proposed a subpixel asynchronous unstructured light algorithm to increase the reconstruction accuracy.²³ It is evident that all existing methods address other aspects of asynchrony or use simple binary patterns containing only two intensity values, 0 and 1.

Therefore, to the best of our knowledge, a notable gap in the field of structured light imaging is the lack of studies on async-FPP using widely adopted sinusoidal fringe patterns due to the challenges posed by complex fringe aliasing. This presents a significant practical limitation, as constructing a flexible structured light system for low-cost consumer-grade applications is difficult. Consequently, the development of flexible, accurate, and compatible methods for async-FPP measurement represents an important area for further research and advancement in structured light imaging. To this end, we model the asynchronous fringe imaging formation and thus, propose a novel aliased fringe pattern (AFP) separation network, which is referred to as APSNet for short, based on U-Net-like architecture to build an easy-to-use async-FPP imaging system. The *a priori* principles embedded in the async-FPP imaging underpin the APSNet to learn the intrinsic inductive bias of AFPs within a generative adversarial framework, which is supervised by a global similarity-informed loss function and guided with fringe pattern information. With the trained network, the synchronization constraint of the FPP is relaxed for allowing the camera and projector to work individually at a constant speed. Thereby, the well-established reconstruction model is still valid for an async-FPP to produce comparable results with the synchronous one. We experimentally show that our method heralds a vision that reconstructs the 3D shapes and geometries from sinusoidal and/or binary fringe patterns without relying on projector and camera synchronization.

2 Materials and Methods

2.1 Asynchronous Fringe Pattern Imaging

In the async-FPP system, the projector is allowed to switch the fringe pattern during the exposure time of camera, leading to AFPs. To describe the AFPs accurately, we first analyze the image formation of the camera component, which converts the light into the intensity responses of image pixels. Assume p and \bar{p} are the pixels in the camera and the projector, respectively. Let $P_{\bar{p}} \in [0, 255]$ be the intensity value of \bar{p} at time t and I_p be the obtained intensity value of p within the exposure time t_e . According to the complementary metal-oxide-semiconductor transistor imaging principle, I_p is the integration of the intensity contribution from all perceptive pixels \bar{p} in the exposure time.¹¹ As not all the \bar{p} contribute to I_p , $R(p, \bar{p}) \in [0, 1]$ is introduced to represent whether \bar{p} has contribution to p as shown in the following equation:

$$I_p = \int_0^{t_e} B_p B_r \sum_{\bar{p}} R(p, \bar{p}) P_{\bar{p}}(t) dt + \int_0^{t_e} C_p(t) dt + \int_0^{t_e} B_r C'_p(t) dt. \quad (1)$$

Equation (1) formulates the intensity response of the pixel p as three parts, including (1) the background light going to the camera directly $C_p(t)$, (2) the background light reflected by the object surface $C'_p(t)$, and (3) the projected fringe patterns reflected by the object surface $P_{\bar{p}}(t)$, where B_p is the relationship between the projector illumination and camera intensity response; B_r is the object surface reflectivity. The image formation model shows that the intensity of the captured image is related to the exposure time and the exposure light amount.

For the synchronous FPP, one fringe pattern is projected and captured by the camera in the exposure time, as shown in Fig. 2(a), extending Eq. (1) to the whole image to capture a fringe pattern expressed as

$$I_n := \int_{t_{n-1}}^{t_n} P_n(t) dt, \quad (2)$$

where $n = 1, \dots, N$, $P_n(t)$ is the n th projected fringe pattern, and I_n is the captured fringe pattern correspondingly in the exposure time range $[t_{n-1}, t_n]$. It is worth noting that in synchronous FPP systems, the background is assumed to be constant

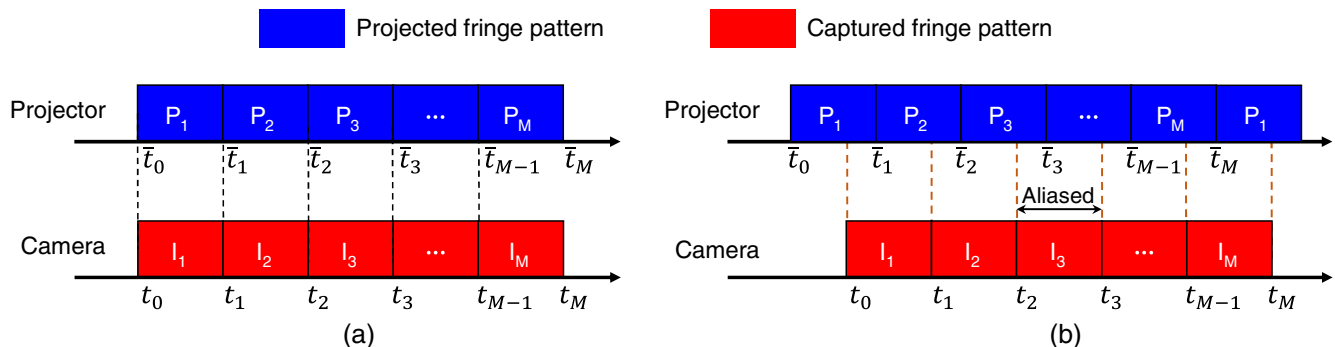


Fig. 2 Illumination and response relation between the projector and camera in (a) synchronous FPP and (b) async-FPP systems, respectively.

and thus the constant terms are omitted in Eq. (2) to focus on the projected fringe patterns.

However, for the async-FPP, as shown in Fig. 2(b), the projected fringe pattern can be switched during the exposure time and multiple fringe patterns are captured by the camera, leading to aliasing of the neighboring fringe patterns. To identify the projected patterns from the aliased observations, it is necessary to explore the aliasing formation in asynchronous projection mathematically.

Assume M fringe patterns are projected at a constant speed S_p and the camera captures N fringe patterns at a constant speed S_c . To capture all projected fringe patterns, S_c is required to be faster than S_p , and the captured frames should be able to cover the projected ones, i.e., $N > M$. It is obvious that, during the exposure time of the camera, the projected fringe pattern is switched in the async-FPP, leading to aliasing of the captured fringe patterns, which can be formulated as follows:

$$I_n = \int_{t_{n-1}}^{t_n} \sum_{m=1}^M r_m(t) P_m(t) dt, \quad (3)$$

$$r_m(t) = \begin{cases} 1 & \text{if } \bar{t}_{m-1} \leq t \leq \bar{t}_m, \\ 0 & \text{otherwise} \end{cases}, \quad (4)$$

where t_n ($n = 0, \dots, N$) and \bar{t}_m ($m = 0, \dots, M$) are the starting time for each captured and projected fringe pattern, respectively, and $r_m(t)$ indicates whether the projected fringe pattern $P_m(t)$ is within the camera exposure time or not. According to Eq. (3), the captured AFP is given as

$$I_n = \int_{t_{n-1}}^{t_n} \sum_{m=1}^M r_m(t) P_m(t) dt = \int_{t_{n-1}}^{t_n^{sw}} P_m(t) dt + \int_{t_n^{sw}}^{t_n} P_{m+1}(t) dt, \quad (5)$$

where t_n^{sw} is the moment when the projection is switched during the capture of I_n .

Equation (5) shows that, in async-FPP, the captured AFP I_n is the integral of all projected fringe patterns during the exposure time, as shown in Fig. 2(b), leading to a significant aliasing effect in the observed I_n and further reconstruction errors if the phase data are directly estimated from I_n . Because the starting time (t_n) and the fringe switching time (t_n^{sw}) are unknown and very difficult to measure, the individual contributions of time-dependent components $P_m(t)$ and $P_{m+1}(t)$ are not easily disentangled using conventional methods, especially for the complex sinusoidal fringe patterns. Despite this, the integral in Eq. (5) implies that aliasing has a global additive yet separable feature in the perspective of intensity, raising some graceful priors in geometry and statistics (to be discussed briefly in the next section). This allows us to learn a latent function with the help of the powerful feature representation capacity of deep neural networks to link the observed AFP with its nonaliased counterparts, which can be used to compute the phase map accurately. We formulate the process as a deep image generation problem by separating a pair of adjacent non-AFPs from an observed AFP.

2.2 Fringe Pattern Separation with Generative Deep Networks

Although the AFP I_n in Eq. (5) measures the intensity of the projected fringes at two neighboring instantaneous times,

P_m and P_{m+1} , in a global additive formation, the unknown-scale intensity aliasing poses an ill-posed ‘‘one-to-many (two)’’ inverse image generation problem that is still difficult to solve in traditional frameworks. However, the global additive aliasing potentially enables the couple of priors from geometry and statistics, as we mentioned before, inspiring us to propose an effective generative model for addressing the problem by aligning with recent advances in the line of generative deep-learning research, such as variational autoencoders,²⁴ autoregressive models,²⁵ and generative adversarial nets (GANs).²⁶ To this end, we present an APSNet with the geometry prior for separating AFPs and show how to train it within the framework of GANs supervised by the statistical prior.

Consider the two projected adjacent fringe images P_m and P_{m+1} , and the corresponding aliasing pattern measure $I_n \in \mathbb{R}^{H \times W}$, where H and W , the height and width of the measured image, respectively, define the aliased domain. Our goal is to learn an inverse mapping of Eq. (5) by constructing the APSNet, denoted by G , to retrieve the pair of projected fringe patterns, i.e., $[P_m, P_{m+1}] = G(I_n)$. A fundamental observation in Eq. (5) is that the aliasing is independent of spatial location but is only the global addition of pixel intensity of P_m and P_{m+1} . From a geometric point of view, the underlying grid structure of the fringe image is not deformed in the aliasing procedure and the APSNet for I_n defined on the aliased grid domain ($\mathbb{R}^{H \times W}$) is no longer required to account for pixel permutation invariance, providing a strong geometric prior that is translational symmetry. It turns out to be a graceful inductive bias of translation equivariance (or loosely referred to as shift-invariance) in geometric learning, assuming the aliased pattern separation should not be affected by the pixel location to improve the learning efficiency to those deep models that satisfy this prior. This geometric prior gives a global invariant support that can be proved by Fourier transformation to intimately connect the APSNet operations with convolutions, thus forwarding our model into a convolutional neural network. Though the convolutional APSNet is substantially strengthened by the global translational symmetry prior, it is not sufficient to overcome the curse of dimensionality posed in high-dimensional pixel space of fringe patterns. Our insight into solving this problem is to learn the multiscale representation by introducing maximum pooling operation in the encoding phase, as it provides efficient local invariance in cooperating with the global one that underpins the convolution.

Responding to the underlying inductive biases and operations in both global and local geometric invariances, we propose implementing the APSNet as a convolutional encoder–decoder architecture based on the existing U-Net model,²⁷ as shown in Fig. 3. This architecture serves as the backbone of the generator in our GAN framework in Fig. 4. The encoder of G is experimentally determined to build with four consecutive blocks. Each block has an identical structure consisting of a convolution block (CB for short) followed by a max-pooling (MP) layer with a stride of 2 to reduce the spatial dimension while preserving the salient response. The CB is built up by stacking a Conv2D layer twice, and each followed by batch normalization²⁸ and in-place ReLU activation.²⁹ When the pixel data flow through every encoding block in the forward pass, the fringe information aliased in the observed image is extracted and encoded in a latent space, abstracting out as a latent representation $\mathcal{Z} \in \mathbb{R}^{512 \times (H/d) \times (W/d)}$. For the decoder path, the source fringe pattern information is progressively disentangled by upsampling (US) and convolving the learned latent feature map 3 times,

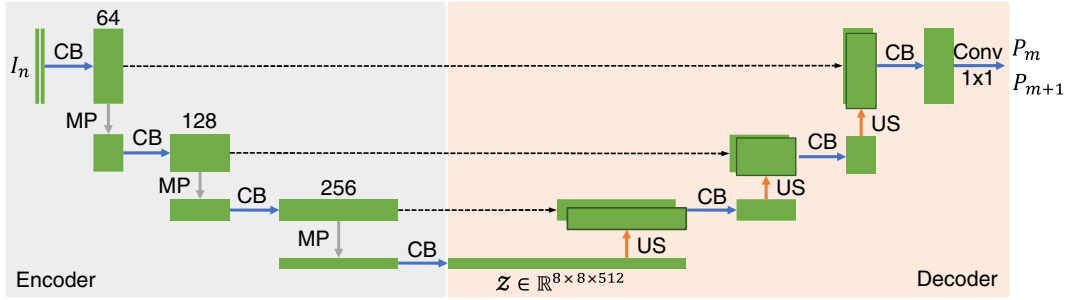


Fig. 3 APSNet with U-Net architecture and latent representation for AFP separation.

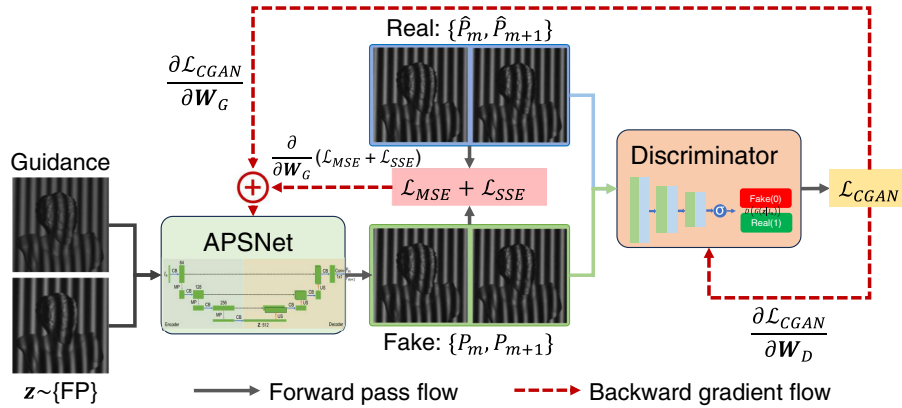


Fig. 4 Schematic description of training APSNet within the conditional GAN framework.

allowing us to generate the separated fringe patterns $\{P_m, P_{m+1}\}$ with a 1×1 convolution layer at the end. As the decoding is started in a deep layer, the skip connection is necessarily adopted to reintroduce the learned features from the corresponding encoder layer to avoid the deep network forgetting the detailed features for making the learning stronger. For details on the skip connection we refer to, see citation of ResNet.³⁰

The convolution and pooling operations in the architecture embed the geometric prior implicitly shown in Eq. (5), imposing an intrinsic inductive bias to APSNet. As a result, it presents a powerful generative capacity for any similar classes by encoding the aliased pattern I_n as the latent variable \mathcal{Z} and then decoding alternately to reconstruct the source patterns P_m and P_{m+1} with the help of the corresponding encoder-passed skip connection. Notably, the encoding part can be interpreted as a conditional autoencoder that allows our model to focus on the important pattern bits of fringe data in a low-dimensional latent space by giving a fringe pattern guidance. Compared to the original pixel space, this latent space representation is more suitable for likelihood-based generative networks, implying that we can supervise the APSNet learning by integrating a mean-squared loss term, as given in Eq. (7) later. Additionally, the symmetric structure and operations of APSNet align well with the strong geometric prior of translational symmetry, which, all combined, contribute to making our method more efficient in terms of computing and data resources, both of which are well-known important factors affecting the development of generative models. In Sec. 3, we show this architecture exactly learns the additive yet complex aliasing relationship in the high-dimensional pattern data with an affordable computational cost,

but also recovers high quality source patterns in the original space as a generative network does but requires few training data.

To train the APSNet G for separating the AFPs, we adopt a supervised adversarial learning by following the GAN strategy guided by an aliased fringe measure, as shown in Fig. 4. An additional network, known as the discriminator D , is introduced here to evaluate the generation quality of the model G in the training loop by taking the output of G as input, given the real data correspondences, i.e., $D(G(z|I_n)) = D(\{P_m, P_{m+1}\}|\{\hat{P}_m, \hat{P}_{m+1}\})$, where z is any of the measured fringe patterns. (We note the real data samples with $\hat{\cdot}$ operator from now on.) The architecture of discriminator D is a multiple layer perception (MLP) with three convolution layers followed by a sigmoid activation to output the probability of the generated fringe patterns. With the discriminator D , the fundamental training loss of the model G is governed as

$$\begin{aligned} \mathcal{L}_{CGAN}(G, D) = & \mathbb{E}_{P \sim p_{\text{data}}(P)} (\log D(\{P_m, P_{m+1}\}|\{\hat{P}_m, \hat{P}_{m+1}\})) \\ & + \mathbb{E}_{z \sim p_z(z)} (\log(1 - D(G(z|I_n))))). \end{aligned} \quad (6)$$

The loss function above frames the fringe pattern separation as an adversarial learning problem with two submodels: the generator G , i.e., APSNet, that we train to generate nonaliasing fringe patterns, and the MLP-based discriminator D that tries to classify fringes as either real (from the domain) or fake (generated). The two models are trained together in a zero-sum game until the discriminator model is fooled about half the time, meaning our APSNet is generating plausible fringe patterns

$\{P_m, P_{m+1}\}$ from the observed ones $\{I_n\}$. It is worth noting that, in contrast to the original adversarial loss in Ref. 26, the loss in Eq. (6) is conditioned by the observed AFP I_n , which imposes a strong guidance with the same modality and structure, making the training dynamics stabler and more predictable. Moreover, this conditioning can help reduce the so-called mode collapse, which limits the generative capacity of the generator in traditional GANs. Consequently, these two features of the adversarial loss term contribute to reducing the dependency of our training framework on the amount of data.

In addition, the principles across the global geometric prior in Eq. (5) also shed light on the statistical property of the fringe pattern aliasing; that is, the intensity distributions of the projected source patterns hold globally in the aliased pattern range. We found that it provides a strong statistical prior for model training: the separated fringe patterns $\{P_m, P_{m+1}\}$ we reconstructed from the aliased measures should maintain a statistical similarity with the projected source patterns $\{\hat{P}_m, \hat{P}_{m+1}\}$ in the original spatial domain. In cooperating with the statistical prior and the feature of aforementioned latent space representation, we propose to use the mean-squared error (MSE)³¹ as a loss term to measure the low-level intensity distribution correlation in original pixel space and then, generate a likelihood-controlled gradient flow that can be backpropagated to the latent space for learning the latent variable \mathcal{Z} implicitly in the backward pass. Additionally, we also respond to this prior by evaluating the loss of structural similarity error (SSE)³² at a higher level than the pixel intensity. The MSE and SSE loss terms are given in Eqs. (7) and (8), respectively,

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (P_i - \hat{P}_i)^2, \quad (7)$$

$$\mathcal{L}_{\text{SSE}} = \frac{(2\mu_{\hat{P}}\mu_P + c)(2\sigma_{\hat{P}P} + c)}{(\mu_{\hat{P}}^2 + \mu_P^2 + c)(\sigma_{\hat{P}}^2 + \sigma_P^2 + c)}, \quad (8)$$

where \hat{P} and P are, respectively, the real projected fringes and the generated fringes; $\mu_{\hat{P}}$ and $\sigma_{\hat{P}}^2$ are mean and variance of the sampled real fringes; μ_P and σ_P^2 are mean and variance of the corresponding generated fringes; $\sigma_{\hat{P}P}$ is the covariance of the sampled and generated fringes; and c is a small but nonzero constant to prevent the numeric instability potentially caused by near-zero $\mu_{\hat{P}}^2 + \mu_P^2$ and/or $\sigma_{\hat{P}}^2 + \sigma_P^2$.

We, finally, propose a loss function in Eq. (9) by summing the loss terms in Eqs. (6) and (7) together to train the generator G and the discriminator D alternatively, as follows:

$$\mathcal{L}(G, D) = \lambda_1 \mathcal{L}_{\text{CGAN}}(G, D) + \lambda_2 \mathcal{L}_{\text{MSE}}(G) + \lambda_3 \mathcal{L}_{\text{SSE}}(G), \quad (9)$$

where $\lambda_i, i = 1, 2, 3$ are three hyperparameters for controlling the contribution of each loss term. We experimentally found that $\lambda_1 = \lambda_2 = \lambda_3$ is the optimal configuration of the loss term contributions in this work; details are given in Sec. 3.1.

As the loss terms \mathcal{L}_{MSE} and \mathcal{L}_{SSE} impose a supervision from the intensity and structural similarity, the learning schematic shown in Fig. 4 could be regarded as a statistical prior-informed adversarial learning framework on the basis of the data-driven training loss $\mathcal{L}_{\text{CGAN}}$, enforcing the generator G , i.e., APSNet, to learn the underlying information of the projected source fringe patterns in considering both geometric and statistical priors uniformly. This helps to significantly improve the quality of the generated fringe patterns from the model training perspective while reducing the size of the training data set. The parameters w_G and w_D related to APSNet and the discriminator, respectively, are updated by the AdamW optimizer³³ through autodifferentiation-powered dual-stream gradient backpropagation,

$$\begin{cases} w_G \leftarrow w_G - \eta_G \frac{\partial}{\partial w_G} (\lambda_1 \mathcal{L}_{\text{CGAN}} + \lambda_2 \mathcal{L}_{\text{MSE}} + \lambda_3 \mathcal{L}_{\text{SSE}}), \\ w_D \leftarrow w_D - \eta_D \frac{\partial \mathcal{L}_{\text{CGAN}}}{\partial w_D}, \end{cases} \quad (10)$$

where η_G and η_D are learning rates; their adjustment policy is given in Sec. 3.1.

2.3 3D Imaging Pipeline with the Trained APSNet

With our trained APSNet, we now have access to the wrapped phase information in the generated nonaliasing fringe patterns and then reconstruct the final 3D geometries of the objects being measured. Figure 5 shows the overall pipeline that accepts three AFPs that are observed from an object surface and outputs the 3D shape of the object.

The pipeline consists of three independent phases. The first one is AFP separation, which is done by the APSNet trained in Sec. 2.2. After this phase, we have four successive non-AFPs with different phases. To this end, we propose to estimate the phase information via the well-established four-step phase-shifting algorithm in Ref. 3. Consider that the four phase-shifting patterns generated by our trained APSNet are P_{m+i} with $i = 0, 1, 2, 3$. The wrapped phase can then be estimated as

$\phi_w(\mathbf{u}) = -\arctan\left(\frac{\sum_{i=0}^3 P_{m+i}(\mathbf{u}) \sin \frac{\pi}{2} i}{\sum_{i=0}^3 P_{m+i}(\mathbf{u}) \cos \frac{\pi}{2} i}\right)$, where \mathbf{u} is the pixel coordinates in the observed fringe domain. By introducing the fringe order $k(\mathbf{u})$ at \mathbf{u} , we obtain the continuous phase map by mapping $\phi_w(\mathbf{u})$ to its unwrapped counterpart as $\phi(\mathbf{u}) = \phi_w(\mathbf{u}) + 2\pi k(\mathbf{u})$.

Finally, with the phase estimate $\phi(\mathbf{u})$, we can reconstruct the 3D shape of the object according to Eqs. (16) and (7) in the work.³

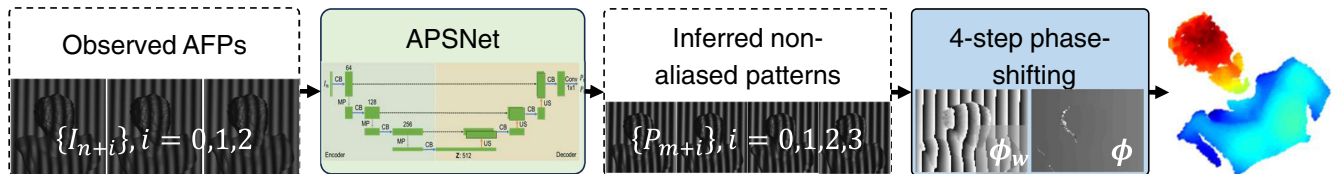


Fig. 5 Pipeline of 3D imaging with the trained APSNet.

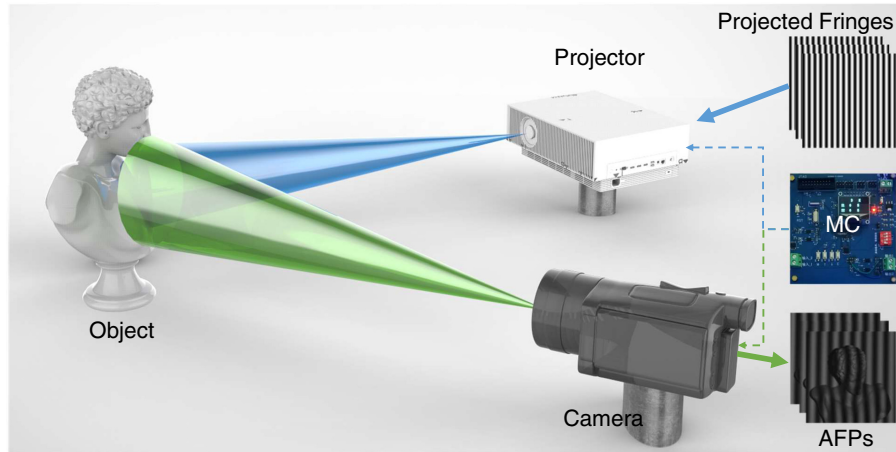


Fig. 6 Experimental setup of our async-FPP system for generating the data set.

3 Results

3.1 Data Set and Training Details

To train APSNet with high generalization capability, the data set is built with a real FPP system consisting of an industrial camera (S2 Camera TJ1300UM) with a resolution of 2048×1536 , a DLP projector with a resolution of 1280×1024 , and a self-developed microcontroller (MC), as shown in Fig. 6. In the MC, a trigger control circuit is designed according to the asynchronous imaging principle in Sec. 2.1 to control the camera and projector to generate sufficient and diverse AFPs for deep model training and testing.

With the async-FPP imaging system, the AFPs are obtained as follows. The projector projects a fringe first with an identical frequency and then sends a high-level signal to the MC. The built-in timer of the MC executes timing according to the pre-defined delay Δt , allowing the camera to observe and record the aliased fringes. (Note that the FPP system will capture nonaliased fringes once the delay Δt is adjusted to zero.) As shown in Fig. 7, if the projector shines the fringe patterns in Figs. 7(a) and 7(b) successively with delay $\Delta t = 0$ ms, the camera will synchronously record two normal fringes without any aliasing, while if the delay is set to $\Delta t = 10$ ms, an AFP

containing the information of in both Figs. 7(a) and 7(b) will be recorded in Fig. 7(c). The adjustable delay mechanism in the async-FPP allows us to make the best use of the capability to gradually increase the high-level signal sent to the camera followed by delayed imaging to build an incredible data set, avoiding the problem of sparsity for the data sampling from a real physical system.

For the source patterns to be projected, five sets of 256×256 sinusoidal fringes with 50, 60, 70, 80, and 90 periods are prepared; each of the source patterns has the phases of 0π , 0.5π , π , and 1.5π , respectively. To generate aliased fringes, the acquisition delay Δt corresponding to each set of source patterns increases from 1 to 10 ms with 1 ms increments. Following the asynchronous imaging, we finally obtain the data set composed of 1500 AFPs, each paired with a label containing two source patterns, to feed into the discriminator and to evaluate the MSE and SSE losses in Fig. 4. Among them, 1125 samples are used for training the proposed APSNet model and the remaining 375 are for model testing.

In the training stage, the AdamW optimizer is adopted to update the parameters of APSNet with batching by following the training loop in Fig. 4 for 500 epochs within which the learning rate is controlled by a dynamic decay strategy. The strategy is that, given an initial value of 1.0^{-4} for the learning rate, it will be reduced to 0.7 times the current value if the validation loss does

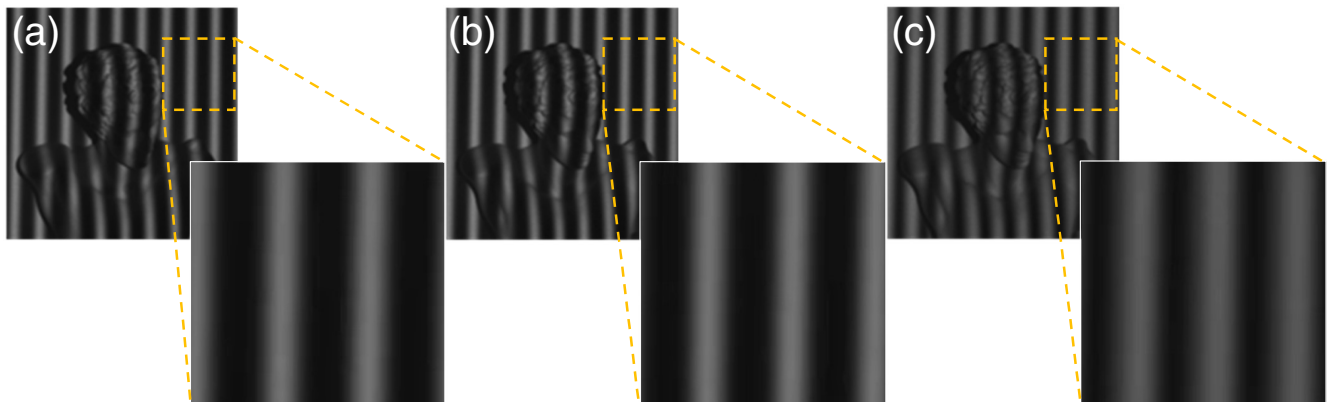


Fig. 7 (a) and (b) Two successively projected fringe patterns and (c) an observed AFP with 10 ms delay.

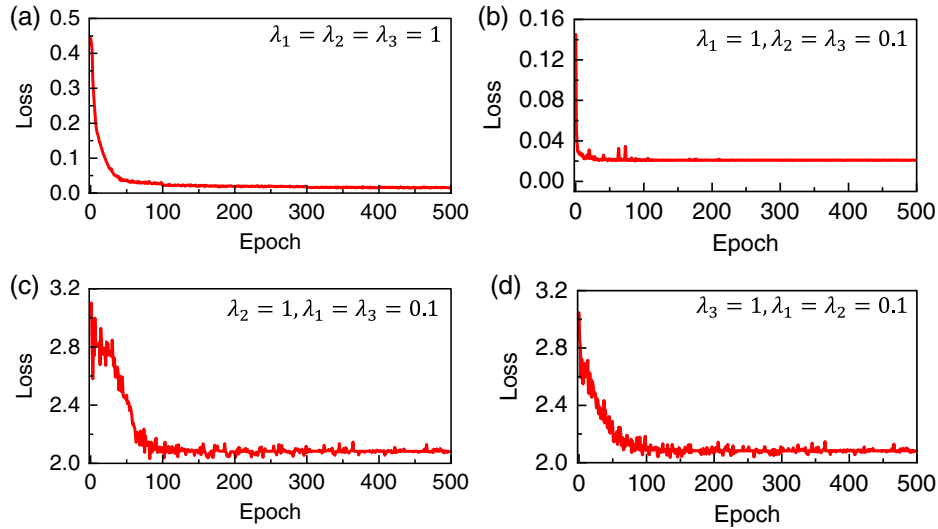


Fig. 8 (a)–(d) Convergence behavior of our APSNet model when training with different loss term contributions.

not decrease in 50 epochs and does not grow in 51 epochs. In addition, we ensure the learning rate is not lower than 1.0^{-7} to prevent slow convergence in the later stages of training. To avoid potential overfitting, l_2 normalization is introduced with the weight decay set to 0.01. Profiting from both the geometric prior embedding and the statistical prior supervision, our APSNet tends to convergence from approximately 100 epochs with coordinated training and testing losses. It should be noted that we have a limited learning domain size on our platform including Intel Xeon E5-2620 CPU (2.10 GHz), 72 GB RAM, and a GPU GeForce GTX 3060Ti, which, however, can be improved using more powerful computational resources.

To find an optimal configuration of the loss weighting hyperparameters in Eq. (9), we investigate the convergence behavior of the loss function for different combinations of λ_i by following the training strategy above. Though the combinations of λ_i are not enumerable, for an optimization (learning) problem based on gradient descent, it is the relative magnitudes of the weights of the loss terms, rather than their absolute values, that affect the convergence behavior (or the training dynamics) of the model, as shown in Eq. (10). Therefore, we first assume that the three loss terms in Eq. (9) have equal contributions by setting the values of λ_i to 1; then we fix one of them and reduce the other two to 0.1 to emphasize the individual contributions of the different loss terms. Following this strategy, we have four different combinations of λ_i , resulting in four training loss curves that reflect the various training behaviors or dynamics of our APSNet, as shown in Fig. 8.

For the case of equal contributions ($\lambda_1 = \lambda_2 = \lambda_3$) in Fig. 8(a), the training loss converges gradually to a small level close to zero by following an exponential decay pattern, as we expected in training most deep neural networks. The result implies that our model shows smooth convergence behavior and thus stable training dynamics under this combination of weights λ_i . For the configuration of $\lambda_1 = 1$ and $\lambda_2 = \lambda_3 = 0.1$ in Fig. 8(b), the training is dominated by the conditional adversarial loss $\mathcal{L}_{\text{CGAN}}(G, D)$. In such a case, we observe that the loss drops to a very low level after a few epochs, approximately around five epochs, and does not decrease any further, implying that the model's parameters are not adequately trained. This

problem aligns with the challenge of training GAN networks, namely, the difficulty in escaping saddle points. This is because $\mathcal{L}_{\text{CGAN}}(G, D)$ is known to exhibit a saddle-shaped loss surface, where gradients become zero at saddle points. For the remaining two cases, the training is dominated by the MSE loss $\mathcal{L}_{\text{MSE}}(G)$ or the SSE loss $\mathcal{L}_{\text{SSE}}(G)$; the training loss curves are shown in Figs. 8(c) and 8(d), respectively. It is obvious that for both MSE and SSE loss terms, the training of our model is volatile with noticeable fluctuations, and the loss converges with much higher values than that in Fig. 8(a). We could attribute this to the weak adversarial relationship between the generator and discriminator in such cases, resulting in the guidance information introduced by the adversarial loss term having almost no effect on the training of the generator. These results suggest that the configuration of $\lambda_1 = \lambda_2 = \lambda_3$ is optimal for training our networks within the framework in Fig. 4.

3.2 Performance Validation and Results

In this section, we evaluate the performance of our trained APSNet on the tasks of separating sinusoidal and binary AFPs for different objects, such as statues, a car model, and a sphere, which are not seen by the trained network in both training and validation.

First, we investigate the inference capacity of our network from fringe pattern separation, phase evaluation, and shape reconstruction levels on a set of new AFPs, assuming that the aliasing level is variant along the time dimension. Figure 9 shows the pipeline of AFP generation [Figs. 9(a)–9(c)] and separation [Figs. 9(c)–9(e)] via the async-FPP system and the trained APSNet, respectively, with evaluated discrepancy maps between the real and generated patterns attached in Figs. 9(f) and 9(g). By inspecting the inferred pattern pair, we can see a very small difference between the inferred results of APSNet and the corresponding source fringe patterns. To quantitatively evaluate the fringe separation quality, the average errors and associated standard deviations in both discrepancy maps are computed, with values of 1.69 and 2.13 for Fig. 9(f), respectively, and 1.94 and 2.88 for Fig. 9(g), respectively.

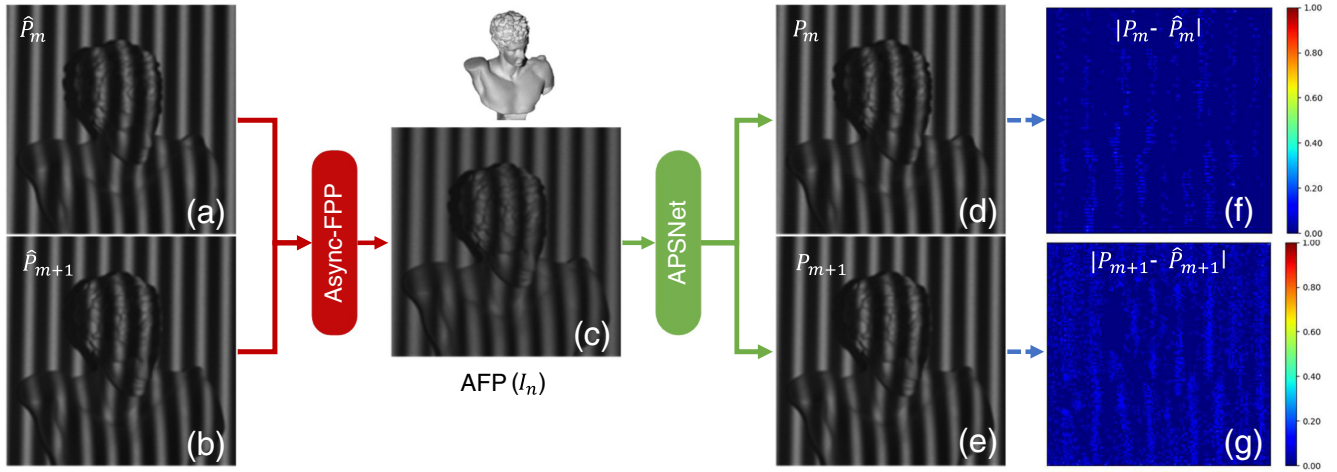


Fig. 9 Pipeline of AFP generation and separation with APSNet: (a) and (b) the source fringe patterns recorded by synchronized FPP, (c) the corresponding AFP recorded by async-FPP system, (d) and (e) are the separated results of (c), and (f) and (g) show the absolute discrepancy between the separated and source patterns.

Furthermore, adjusting the time delay Δt of the async-FPP system in Fig. 6 to 1, 5, and 9 ms, respectively, for simulating the different asynchronous relationship, resulting in three groups of AFPs with different aliasing spanning slight and middle to deep levels. By following a similar pipeline above, the captured AFPs are successfully separated by our APSNet; the resulting fringe patterns are shown in Fig. 10, together with the corresponding error maps. On close inspection of the error maps in both Figs. 9 and 10, we can see that the inference error is randomly distributed with small values (averages are less than 2.0 with variances below 3.0) in both spatial and delay dimensions, which we attribute to discretization errors. With these cases, the U-Net-based deep neural architecture with latent space representation proved to be expressive enough in intensity

field to predict the correct fringe pattern separation for samples from a class of objects by following our training loop, presented in Fig. 4.

Beyond the validation-in-intensity field, we then check the performance of our model by carrying out phase computation and shape reconstruction. To this end, a set of AFPs is separated by the trained APSNet to generate four successive fringe patterns without aliasing; meanwhile, four step synchronous fringe patterns are also captured correspondingly for comparison. By applying the four-step phase-shifting algorithm, we obtain the wrapped phase maps for both synchronous and asynchronous cases as shown in Figs. 11(a) and 11(b), respectively. One can see that these two phase maps are almost identical from the appearance view. To compare quantitatively, we sample

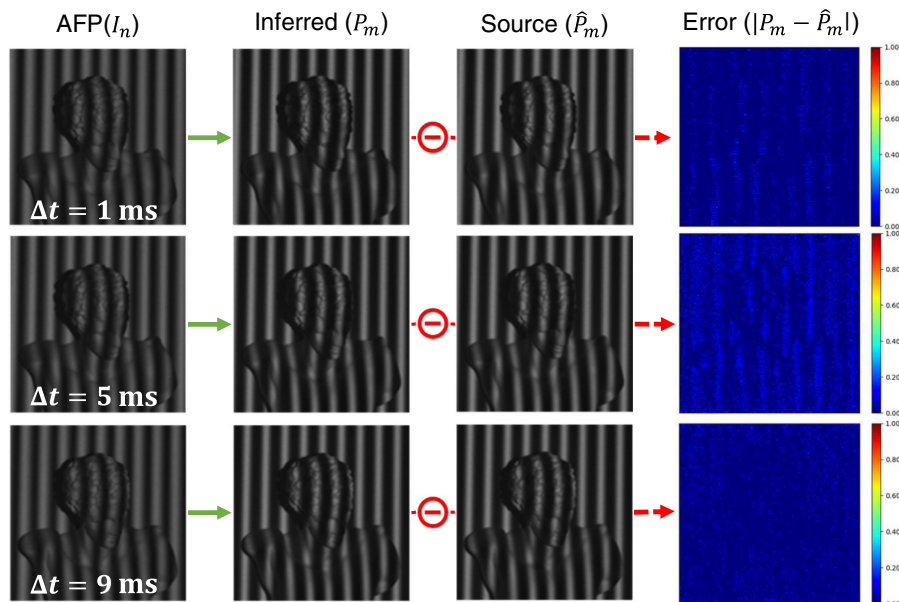


Fig. 10 Inference performance of APSNet on AFPs with different aliasing levels.

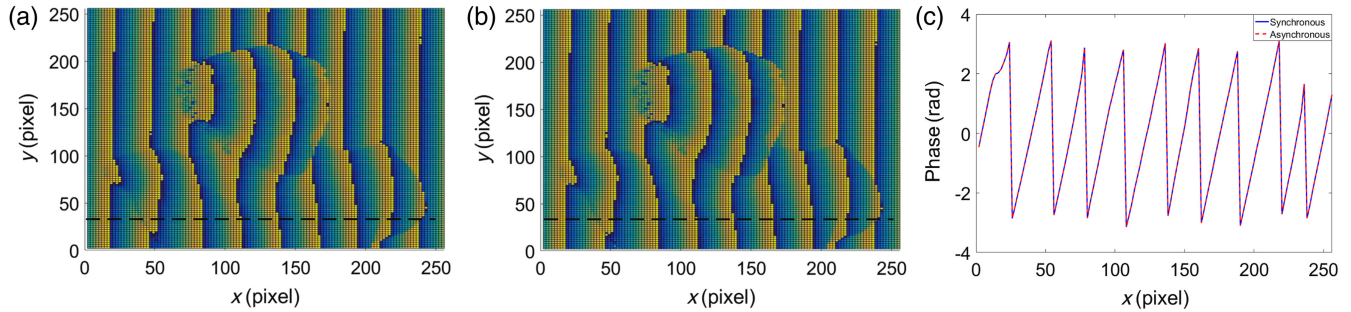


Fig. 11 Comparison of wrapped phase maps for synchronous and asynchronous cases.

the wrapped phase values along the line $y = 38$ pixels from both maps to obtain two phase curves and plot them together in Fig. 11(c). As a result, these two phase curves agree with each other fairly well with a maximum difference of 0.035 rad.

Followed by the phase domain evaluation, we compare the 3D shape reconstruction of the tested statue for the cases of synchronous, asynchronous, and our APSNet-generated fringe patterns. Results are shown in Fig. 12. In this test, we treat the reconstruction result of the synchronous fringe patterns in Fig. 12(a) as benchmark. It is observed that the 3D shape reconstructed from the fringe patterns separated by APSNet in Fig. 12(c) is remarkably close to the benchmark, while that reconstructed from the AFPs directly in Fig. 12(b) has a significant loss of shape information, which is undoubtedly caused by the fringe pattern aliasing. Moreover, the reconstructions remain quantitatively comparable. For that, we sampled a depth curve from each of the reconstructed models along the dashed lines shown in Fig. 12(a); the results are plotted in Figs. 13(a)–13(c), respectively. Let the reconstruction error be $e(d, \hat{d}) = d - \hat{d}$ with d the reconstructed depth in Figs. 13(b) or 13(c) and \hat{d} the corresponding ground truth in Fig. 13(a). We can obtain errors of the reconstructions from AFPs and our APSNet-generated fringe patterns; results are shown in Figs. 13(d) and 13(e), respectively. We can see that our method gets an absolute error that does not exceed 0.008 mm, with a standard deviation (STD) of 0.003 mm. In contrast, the errors in the reconstruction with AFPs range from -5.000 to 3.600 mm, with STD 1.713 mm, which is 3 orders of magnitude higher than our method. These results and the significant performance in computation accuracy highlight the APSNet potential for the 3D imaging applications of AFP-based structured light.

To show the ability of APSNet to be generalized to asynchronous 3D imaging of different objects, we perform a reconstruction task where our trained APSNet is used as a forward model to separate five sets of AFPs captured from objects of three statues, a standard sphere, and a car model. Results are shown in Fig. 14, where the leading row shows the test objects with the AFPs, and APSNet-inferred pattern pairs attached. It is worth noting that in these cases, we project sinusoidal fringe patterns onto two statues (see the first two columns) and binary fringe patterns on the sphere, the car, and the remaining statue (see the third to fifth columns, respectively). By comparing the reconstruction results from the asynchronous (third row) and APSNet-generated (fourth row) fringe patterns with the benchmark (second row), respectively, we found that our APSNet, trained using the proposed method on our data set in Sec. 3.1, successfully extrapolates to cases involving different objects and various fringe pattern types. Since these test objects and (binary) fringe patterns were not seen by the model during training, these reconstruction examples demonstrate the expected generalization capability of our method. We attribute this to the enhancement of image generation ability of the trained generative model, which is achieved through our appropriate architectural choice and supervision using the two statistical prior-informed loss terms in Eqs. (7) and (8).

Importantly, our approach could work by training on one scenario that with few samples (1500) and successfully predicting various other scenarios, showing the generative APSNet has a potential capacity to deal with the problem of asynchronous structured light imaging. To further demonstrate the generalization capacity trained on a smaller data set, Fig. 15 shows a demonstration of multi-object reconstruction, where the test objects

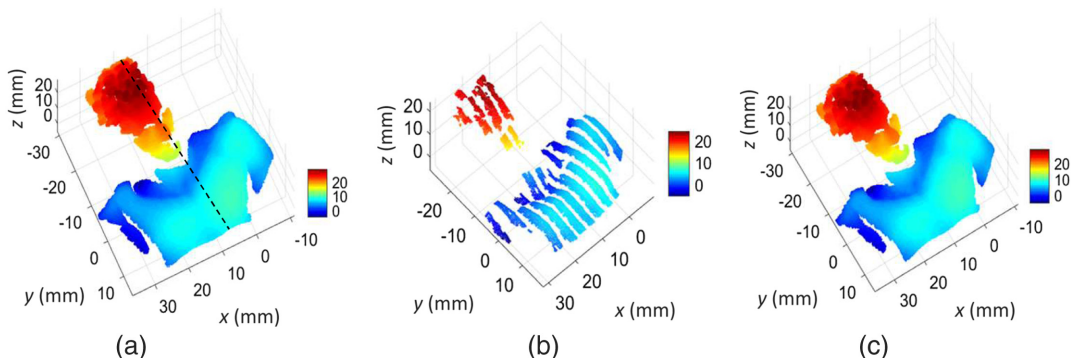


Fig. 12 Reconstructed results from (a) synchronous, (b) asynchronous, and (c) APSNet-generated fringe patterns, respectively.

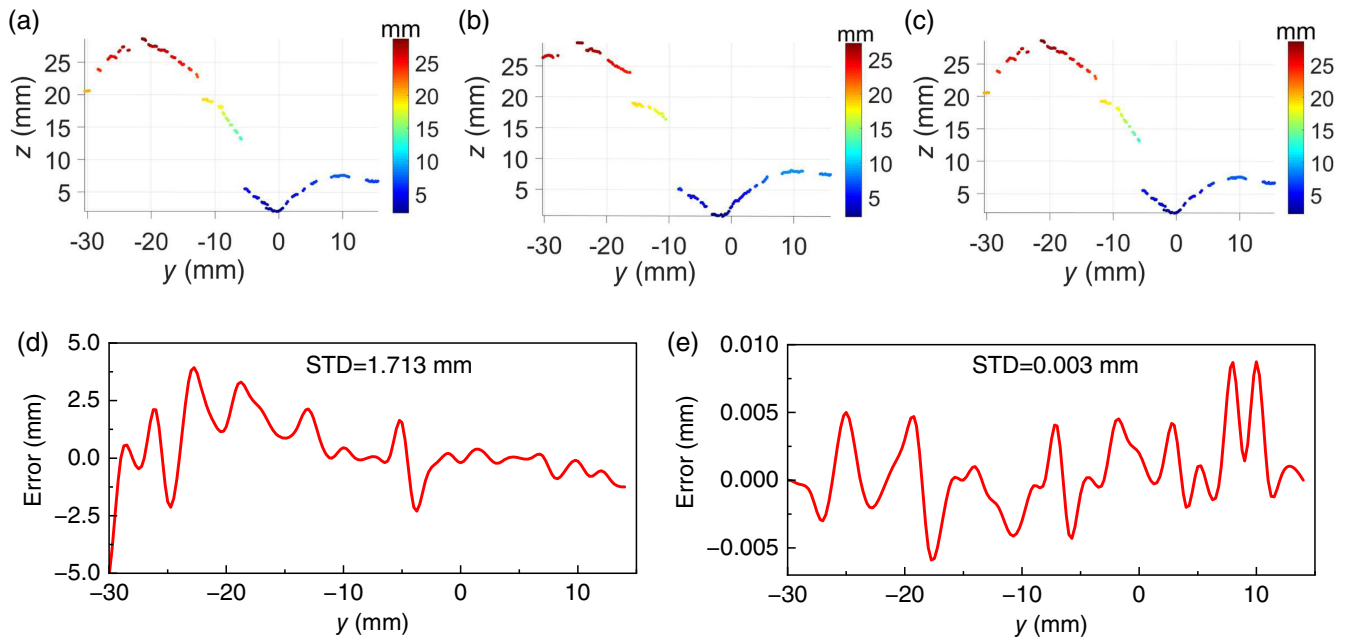


Fig. 13 (a)–(c) Depth curves sampled from Figs. 12(a)–12(c), respectively; (d) and (e) respond to the errors of (b) and (c) relative to (a).

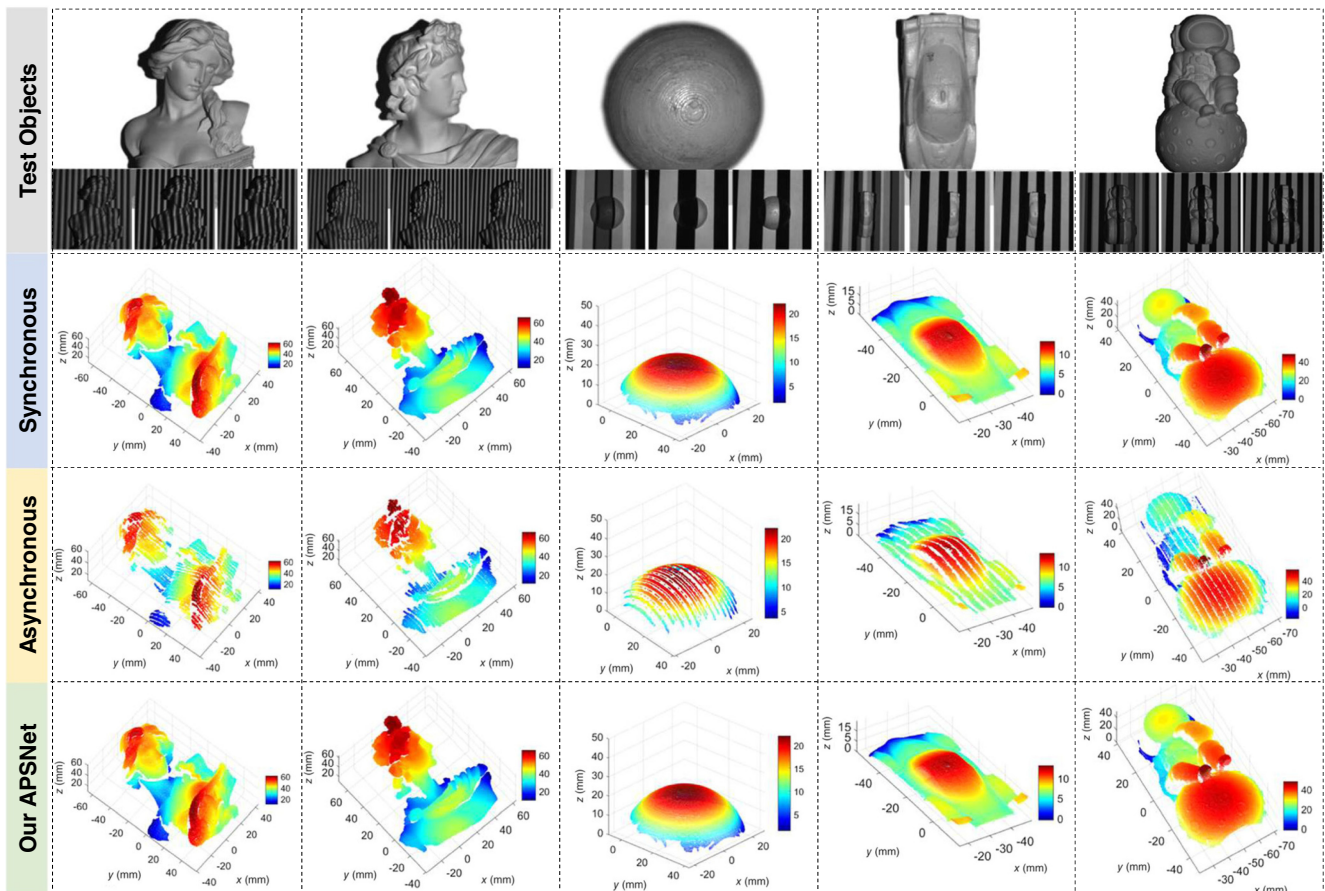


Fig. 14 Reconstruction results for different objects with synchronous, asynchronous, and APSNet-generated sinusoidal and binary fringe patterns, respectively.

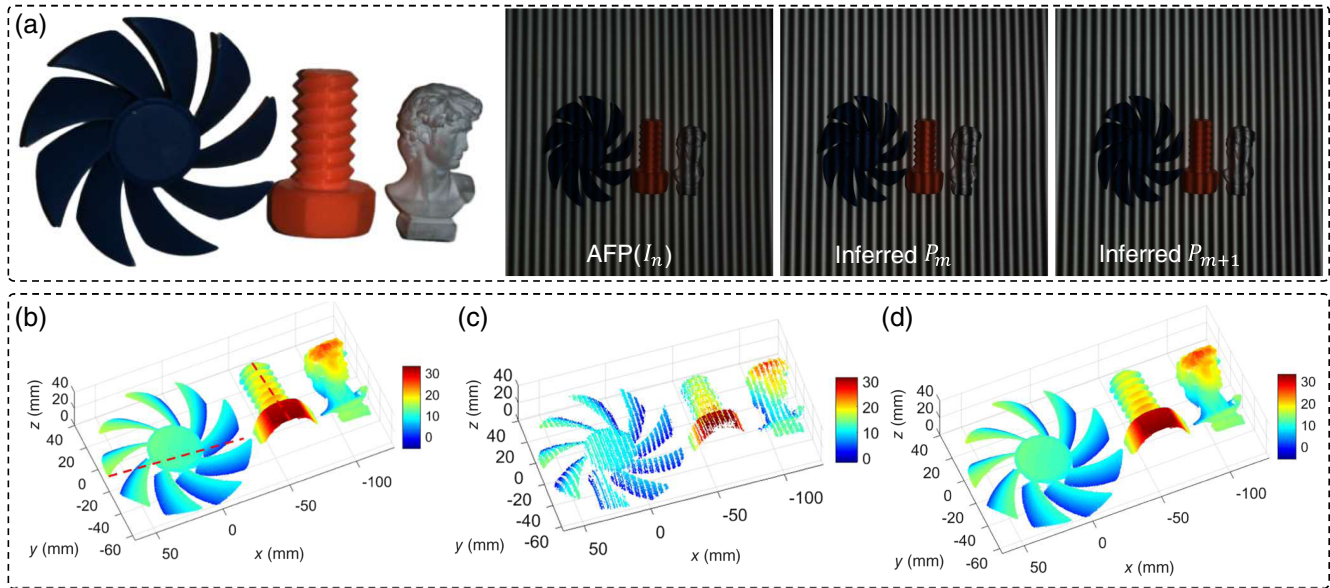


Fig. 15 Demonstration of generalization capability on multi-object asynchronous 3D imaging: (a) measured objects, with its fringe pattern examples before (I_n) and after (P_m, P_{m+1}) separation by our APSNet; (b)–(d) reconstructed results using synchronous, asynchronous, and APSNet-generated fringe patterns, respectively.

include a propeller prototype, a 3D printed bolt, and a statue. As shown in Fig. 15(a), despite the scene containing complex surface objects with different colors (dark blue, red, and gray) and varying reflective materials, our method can still well infer the source patterns (P_m and P_{m+1}) from the observed AFP (I_n). Similarly, we reconstruct the test objects using synchronous, asynchronous, and APSNet-generated fringe patterns, respectively. The results are shown in Figs. 15(b)–15(d), where the dashed lines in Fig. 15(b) indicate the sampling locations for subsequent quality evaluation.

To investigate the reconstruction quality of our method, we compare the reconstruction results of our method with that of the well-established synchronous method by computing the absolute errors of the data points sampled along the dashed line in Fig. 15(b). Note that our comparison here focuses on the propeller and bolt prototypes because they have large curvature surfaces that were not present in the previous validations. The resulting errors can be found in Figs. 16(a) and 16(b), respectively. We can see that the error range for both reconstructed models is consistent with the results shown in Fig. 13(e): the absolute errors for the reconstructed propeller and bolt models

are both less than 0.006 mm, with STDs of 0.002 and 0.001 mm, respectively. This demonstration and error investigation further evidence that our method could generalize to more complex scenes. We attribute this to the network’s intrinsic ability to learn the inductive bias corresponding to the underlying global translation equivariance in the aliasing pattern formation through appropriate loss terms, as we explained in Sec. 2.2.

4 Discussion and Conclusion

In summary, we proposed a statistical prior-informed APSNet that separates the AFPs captured by an async-FPP structured light system for implementing accurate 3D reconstruction as the synchronous ones do. Our network learns to respond to the geometry prior embedded in the fringe aliasing formation and is trained in a generative adversarial framework by minimizing a loss function supervised by global intensity and structural similarity. As a result, the training is allowed to perform directly on a few experimental data without any constraints of the considered async-FPP imaging system. Since the trained network learns the underlying features of fringe patterns from both

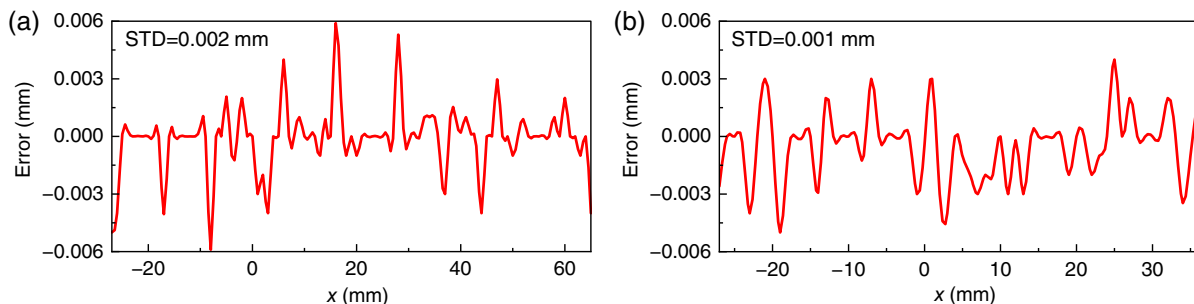


Fig. 16 Reconstruction errors of our method for (a) the propeller and (b) the 3D printed bolt models in Fig. 15.

geometry and statistics, it could generate correct non-AFPs from unseen AFPs before with enough accuracy for phase and 3D shape reconstruction. We showed that our learning approach can effectively extrapolate to the imaging cases of objects that are different from the one used in training. Especially, we found that our network, even though trained only on few sinusoidal patterns, could generalize to aliased binary fringe patterns that are statistically different from the training examples without any further special tuning. The results of 3D demonstrations show that our method achieves imaging results consistent with the synchronous methods and improves the reconstruction accuracy by 3 orders of magnitude relative to the reconstruction of directly using aliased fringes. The validation results relevant to generalization performance highlight that our approach overcomes the fundamental challenge in async-FPP problem, offering a pathway for exploiting synchronous structured light in a wider range of consumer-grade applications and can be potentially extended to other fields, e.g., 3D arrays imaging.

While our trained APSNet is capable of generating non-aliased source fringe patterns from the aliased images captured by the camera, the use of the proposed method can be questionable if there is a significant difference between the frame rates of the projector and the camera. As shown in Fig. 2, this work primarily addresses the aliasing problem involving two fringe patterns. If the projection rate is significantly higher than the camera's frame rate, the camera will capture more than two fringe patterns within the exposure time, leading to a more complex multiple-source aliasing problem, which can become a bottleneck for the direct use of the proposed method. Therefore, we assume that our model is somewhat limited in this respect. Despite this limitation, our method provides a baseline that inspires further advancement of generative deep-learning methods to address asynchronous structured light imaging with more complex aliasing problems.

Disclosures

The authors declare no conflicts of interest.

Code and Data Availability

Data and code underlying the results presented in this paper are not publicly available at this time. Interested researchers may request access to the data and code by contacting the corresponding author and access may be granted under specific conditions, such as signing a nondisclosure agreement.

Acknowledgments

Zhilong Su acknowledges funding from the National Natural Science Foundation of China (Grant Nos. 62375078 and 12002197), the Youth Talent Launching Program of Shanghai University, the General Science Foundation of Henan Province (Grant No. 222300420427), the Key Research Project Plan for Higher Education Institutions in Henan Province (Grant No. 24ZX011), and the National Key Laboratory of Ship Structural Safety.

References

1. A. G. Marrugo, F. Gao, and S. Zhang, "State-of-the-art active optical techniques for three-dimensional surface metrology: a review," *J. Opt. Soc. Am. A* **37**(9), B60–B77 (2020).
2. W. Jiang et al., "2,000,000 fps 2D and 3D imaging of periodic or reproducible scenes with single-pixel detectors," *Photonics Res.* **10**(9), 2157–2164 (2022).
3. S. Lv and Q. Kema, "Modeling the measurement precision of Fringe Projection Profilometry," *Light Sci. Appl.* **12**(1), 257 (2023).
4. S. S. Gorthi and P. Rastogi, "Fringe projection techniques: whither we are?" *Opt. Lasers Eng.* **48**(2), 133–140 (2010).
5. J. Geng, "Structured-light 3D surface imaging: a tutorial," *Adv. Opt. Photonics* **3**(2), 128–160 (2011).
6. S. Zhang, "High-speed 3D shape measurement with structured light methods: a review," *Opt. Lasers Eng.* **106**, 119–131 (2018).
7. J. Xu and S. Zhang, "Status, challenges, and future perspectives of fringe projection profilometry," *Opt. Lasers Eng.* **135**, 106193 (2020).
8. L. Lu et al., "Kinematic target surface sensing based on improved deep optical flow tracking," *Opt. Express* **31**(23), 39007–39019 (2023).
9. S. Feng et al., "Generalized framework for non-sinusoidal fringe analysis using deep learning," *Photonics Res.* **9**(6), 1084–1098 (2021).
10. S. Zhu et al., "Superfast and large-depth-range sinusoidal fringe generation for multi-dimensional information sensing," *Photonics Res.* **10**(11), 2590–2598 (2022).
11. D. Moreno, F. Calakli, and G. Taubin, "Unsynchronized structured light," *ACM Trans. Graphics* **34**(6), 178 (2015).
12. D. Bradley et al., "Synchronization and rolling shutter compensation for consumer video camera arrays," in *IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit. Workshops*, IEEE, pp. 1–8 (2009).
13. L. Lu et al., "Motion induced error reduction methods for phase shifting profilometry: a review," *Opt. Lasers Eng.* **141**, 106573 (2021).
14. P. Zhang et al., "High dynamic range 3D measurement based on structured light: a review," *J. Adv. Manuf. Sci. Technol.* **1**(2), 2021004 (2021).
15. S. Feng et al., "Fringe pattern analysis using deep learning," *Adv. Photonics* **1**(2), 025001 (2019).
16. T. Petković et al., "Software synchronization of projector and camera for structured light 3D body scanning," in *Proc. of 7th Int. Conf. 3D Body Scanning Technol.*, Lugano, pp. 286–295 (2016).
17. Y. Wang et al., "Enhanced phase-coding method for three-dimensional shape measurement with half-period codeword," *Appl. Opt.* **58**(27), 7359–7366 (2019).
18. W. Yongchang et al., "Period coded phase shifting strategy for real-time 3-D structured light illumination," *IEEE Trans. Image Process.* **20**(11), 3001–3013 (2011).
19. F. Paredes-Vallés et al., "Fully nonomorphous vision and control for autonomous drone flight," *Sci. Rob.* **9**(90), eadi0591 (2024).
20. H. Fujiyoshi et al., "Fast 3D position measurement with two unsynchronized cameras," in *Proc. 2003 IEEE Int. Symp. Comput. Intell. in Rob. and Autom. Comput. Intell. in Rob. and Autom. for the New Millennium*, Vol. 3, pp. 1239–1244 (2003).
21. N. Hasler et al., "Markerless motion capture with unsynchronized moving cameras," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 224–231 (2009).
22. C. E. Asmi and S. Roy, "Fast unsynchronized unstructured light," in *15th Conf. Comput. and Rob. Vision (CRV)*, IEEE, pp. 277–284 (2018).
23. C. E. Asmi and S. Roy, "Subpixel unsynchronized unstructured light," in *Proc. 14th Int. Joint Conf. Comput. Vision, Imaging and Comput. Graphics Theory and Appl. (VISIGRAPP 2019)*, Vol. 5, pp. 865–875 (2019).
24. D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Found. Trends Mach. Learn.* **12**(4), 307–392 (2019).
25. A. Van Den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *Int. Conf. Mach. Learn.*, PMLR, pp. 1747–1756 (2016).

26. I. Goodfellow et al., “Generative adversarial networks,” *Commun. ACM* **63**(11), 139–144 (2020).
27. O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
28. S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in *Int. Conf. Mach. Learn.*, PMLR, pp. 448–456 (2015).
29. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Adv. Neural Inf. Process. Syst.* **25** (2012).
30. K. He et al., “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 770–778 (2016).
31. Z. Wang and A. C. Bovik, “Mean squared error: love it or leave it? A new look at signal fidelity measures,” *IEEE Signal Process. Mag.* **26**(1), 98–117 (2009).
32. Z. Wang et al., “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
33. I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” arXiv:1711.05101 (2017).

Lei Lu received his BEng degree from Henan University of Science and Technology, China, in 2007; his MEng from Zhengzhou University, China, in 2011; and his PhD from the University of Wollongong, Australia, in 2015. He was a postdoctoral fellow in Singapore University of Technology and Design, Singapore, from 2015 to 2016. Now, he is an associate professor in Henan University of Technology, China. His research interests include 3D shape measurement, 3D printing, and 3D data processing.

Zhilong Su is currently an associate professor at Shanghai Institute of Applied Mathematics and Mechanics, Shanghai University, and the deputy director of the Department of Mechanics, School of Mechanics and Engineering Science, Shanghai University. He received his PhD from the Department of Engineering Mechanics, Southeast University, in 2019. His main research activities are devoted to geometric optical sensing, visual intelligence, generative and geometric deep learning, and photo-mechanics.

Banglei Guan is currently an associate professor at the National University of Defense Technology. His research interests include photo-mechanics and videometrics. He has published research papers in top-tier journals and conferences, including IJCV, IEEE TIP, IEEE TCYB, CVPR, ECCV, and ICCV.

Qifeng Yu is currently a professor at the National University of Defense Technology. He is an academican of the Chinese Academy of Sciences. He has authored three books and published over 200 papers. His current research fields are image measurement and vision navigation.

Wei Pan is currently working in OPT Machine Vision Corp. as a research leader in 3D algorithm development. Prior to that, he worked as a research fellow at Shenzhen University and South China University of Technology after he got his PhD from Singapore University of Technology and Design. His research interests include 3D imaging, 3D data processing, computer vision, machine learning, and computer graphics.

Biographies of the other authors are not available.