Research Article Vol. 33, No. 25 / 15 Dec 2025 / Optics Express 52084
Optics EXPRESS

Deep transformer-enhanced dynamic structured light for three-dimensional moving object imaging

PENG LI,[1] PENGKAI DOU,[2] ZHILONG SU,[3,5] LEI LU,[1,6] AND WEI PAN[4]

[1]Institute for Complexity Science, Henan University of Technology, Zhengzhou 450001, China
[2]College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China
[3]Shanghai Institute of Applied Mathematics and Mechanics, School of Mechanics and Engineering Science, Shanghai Key Laboratory of Mechanics in Energy Engineering, Shanghai University, 200444, Shanghai, China
[4]Department of Research and Development, OPT Machine Vision Tech Co., Ltd., 523860 Dongguan, China
[5]szloong@shu.edu.cn
[6]lulei@haut.edu.cn

Abstract: Dynamic three-dimensional (3D) shape measurement techniques based on phase-shifting profilometry (PSP) have achieved considerable progress in previous studies. However, for dynamic scenes with arbitrary directional motion, accurately measuring the 3D shape and/or motion of objects remains an unsolved challenge. When an object moves in any direction within 3D space, both its spatial positions and the height values of individual surface points change. Moreover, the height variations at each pixel are typically inconsistent, resulting in different unknown phase-shift errors across various locations. To address the challenge, this paper proposes a PSP dynamic 3D shape measurement method that combines transformer-based motion tracking and motion compensation. We propose to use the COTR network to achieve motion tracking of objects under coded fringe illumination. By integrating this tracking information into the motion error estimation process of PSP, the proposed method enables pixel-wise error compensation, allowing for high-precision recovery of phase information for objects with 3D motion. Experimental results validate the effectiveness of the proposed method, demonstrating a significant reduction in motion-induced errors and improved recovery of true surface geometry.

1. Introduction

Three-dimensional (3D) shape imaging with optical sensors has demonstrated remarkable usage in various fields, such as intelligent manufacturing, embodied vision, biomedicine, and virtual reality [1–4]. Among the existing methods, fringe projection profilometry (FPP) based on structured light is one of the most popular techniques due to its high accuracy and robustness [5–7]. According to different phase retrieval methods, FPP can be divided into two main categories: Fourier transform profilometry (FTP) methods [8] and phase shifting profilometry (PSP) methods [9]. The former obtains phase information by extracting the fundamental frequency component in the frequency domain, which typically requires one frame of the sinusoidal fringe pattern to obtain the phase map and is thus independent of object motion. However, it is susceptible to ambient noise and complex surface textures, which limit its stability and accuracy in intricate environments. In contrast to FTP, the latter PSP offers greater robustness to changes in ambient light and fluctuations in surface reflectivity. Nevertheless, it requires more image data and processing time and typically needs at least three phase-shifting fringe patterns for phase calculation. Additionally, the measurement object must remain stationary during the projection

#579307 https://doi.org/10.1364/OE.579307
Journal © 2025 Received 18 Sep 2025; revised 12 Nov 2025; accepted 12 Nov 2025; published 2 Dec 2025

Optics EXPRESS

and acquisition of multiple fringe patterns [10]. Any motion of the object during this process can introduce errors or lead to reconstruction failure. This sensitivity to motion has significantly hindered the application of PSP in dynamic measurement scenarios.

To deal with the problem of motion errors in the PSP, several researchers have attempted to eliminate these effects by increasing the hardware's projection and capture speeds [11–14]. Wang et al. [15] proposed a dual-frequency binary phase-shifting method, combining the error-diffusion dithering technique with OPWM, to measure the 3D shape of a living rabbit heart at a data acquisition rate of approximately 800 Hz. Heist et al. [16] employed an LED array projector for high-speed 3D shape measurement, achieving a 3D frame reconstruction rate of about 330 Hz at a binary fringe projection rate of 3 kHz during measurements of a white planar object. This demonstrated its potential for accurate, high-speed 3D shape measurement. However, the increase in speed will not only significantly increase the cost of the hardware, and the effects caused by object movement can still not be ignored.

At the same time, several motion error estimation algorithms have been proposed to mitigate the influence of motion [17–19]. Weise et al. [20] assumed that motion errors are linearly related to inter-frame delay and exhibit local smoothing characteristics. They proposed using a linear least squares fit within the local neighborhood of each pixel to estimate the positional offset, thereby correcting the phase map. Guo et al. [21] proposed a real-time motion error compensation method. This approach does not directly estimate the motion error; instead, it divides the four-step phase-shifted image into two groups: front and back. These groups are then averaged to partially cancel out the motion error. While this method effectively reconstructs the three-dimensional contour of a moving object, its reconstruction accuracy is limited, and it requires a high data acquisition rate from the measurement system. Feng et al. [22] categorized motion-induced errors into three types: motion ripples, phase unwrapping errors, and motion outliers. By introducing an error term into the phase-shifting fringe pattern, their method iteratively estimates the average phase-shifting error for a segmented object. Error compensation is then performed by projecting three fringe patterns. Similarly, Yu et al. [23] exploited the periodicity of the fringe pattern to average the phase error over one-dimensional T/2 adjacent pixels. These methods significantly enhance the measurement accuracy of dynamic objects by estimating and compensating for motion-induced phase shifts, leveraging the statistical properties of local phase variations. However, most of the above methods are limited to measuring movement in the depth direction and fail to address the issue of image projection point mismatches caused by object movement in other directions.

Tracking the position of the object during the fringe projection period is a direct and effective method for motion compensation [24]. Lu et al. [25] analyzed the relationship between changes in object motion and phase. They proposed that the rotation matrix and translation vector describing the object motion can be estimated to map to phase changes. The accurate phase pattern of a moving object was determined by deriving the fringe pattern expression influenced by two-dimensional motion. Lu et al.'s motion compensation model performs well for motion compensation within the X–Y plane but shows limited capability in handling Z-direction motion. Recently, Wang et al. [26] compensated for motion-induced errors by tracking the projected points. However, their method simplifies the modeling of phase errors introduced by object motion and requires the additional projection of speckle patterns during motion. When the moving object exhibits a complex surface contour, this approach may lead to relatively large residual errors. Li et al. [27] proposed an object alignment and motion estimation method based on principal component analysis, which achieves image alignment and phase error compensation by incorporating windowed phase correlation and singular value decomposition (SVD) to extract the global displacement of the image. This method demonstrates good real-time performance and measurement accuracy, making it suitable for dynamic 3D reconstruction scenarios. However, it

Optics EXPRESS

only estimates the global translation and motion error of the entire image, making it difficult to capture local motion differences at the pixel level.

Given the recent application of deep learning methods in image processing [28–32], Li et al. [33] developed a three-stream neural network to mitigate motion errors. The three phase-shifting fringe patterns, captured by the system, were separately input into each of the network's three interfaces, with the corresponding phase diagram directly output as the numerator and denominator. To generate the necessary dataset, they designed a virtual fringe projection system. The network model was optimized by automatically generating fringe patterns for various types of motion, enabling the model to handle moving objects. Tan et al. [34] leveraged the expressive power of deep learning techniques to address strongly non-linear problems and proposed a deep learning-based method for non-uniform motion error suppression. First, they simulated motion-induced errors based on physical principles, then used a deep neural network to extract high-level features from the error distribution. A pixel-by-pixel, hidden nonlinear mapping between phase and motion was established through data learning. Although these methods possess the adaptability advantage of data-driven approaches and have achieved error compensation for 3D moving objects, their generalization capability still relies on pre-trained models.

Motivated by the above drawbacks, the goal of this study is to reconstruct the 3D shapes of rigid objects with arbitrary 3D motion based on PSP. First, we analyze the phase variations induced by the 3D motion of the object, where different surface points experience distinct displacements, resulting in varying additional phase shifts in the captured fringe pattern. Then, a novel algorithm is proposed to estimate the 3D motion-induced phase shift and subsequently correct phase information for each pixel based on accurate object motion estimation. To this end, a learning-based pixel motion tracking method with a transformer architecture is introduced to achieve sub-pixel correspondence of object positions before and after motion under fringe-pattern illumination. Our contributions are summarized as follows:(1) we introduce a novel pixel-wise motion error estimation method to compute the additional phase-shift error at pixels of interest. (2) we propose a fine-tuned COTR network under fringe-pattern illumination, enabling sub-pixel-level correspondence matching of fringe images before and after motion; (3) we combine this tracking information with the motion compensation model, allowing for high-precision recovery of phase information for rigid objects with 3D motion; (4) Experimental results validate the effectiveness of the proposed method, demonstrating a significant reduction in motion-induced errors and improved recovery of true surface geometry.

The rest of the paper is organized as follows. Section 2 introduces the principle of the proposed method, including the fundamentals of phase-shift reconstruction under 3D motion (Section 2.1), fringe pattern correspondence estimation with transformers (Section 2.2), and shape reconstruction under 3D motion (Section 2.3). Section 3 presents the experiments and validation results. We conclude in Section 4.

2. Method

2.1. Phase-shift reconstruction in 3D motion

In the standard N-step PSP algorithm, the projector continuously projects a series of sinusoidal phase-shifted fringe patterns, and the camera captures the fringe patterns synchronously. The image intensities observed by the camera for both the reference plane and the object surface are described as follows:

$$S_n(x, y) = A + B\cos[\varphi(x, y) + \Delta\varphi_n], \tag{1}$$

and

$$I_n(x, y) = A + B\cos[\varphi(x, y) + \Phi(x, y) + \Delta\varphi_n], \tag{2}$$

Optics EXPRESS

where $n = 1, 2, \ldots, N$ is the index of projected fringe patterns, A is the background intensity, B is the degree of modulation, $\varphi(x, y)$ is the phase value on the reference plane, $\Delta\varphi_n$ is the theoretical phase shift, which is equal to $\frac{2\pi(n-1)}{N}$; $\Phi(x, y)$ is the phase change caused by the height modulation of the object. The wrapped phase on the reference plane and on the object can be calculated using the following formula:

$$\varphi(x, y) = -\tan^{-1}\left[\frac{\sum_{n=0}^{N-1} S_n(x, y)\sin(\Delta\varphi_n)}{\sum_{n=0}^{N-1} S_n(x, y)\cos(\Delta\varphi_n)}\right], \tag{3}$$

and

$$\varphi(x, y) + \Phi(x, y) = -\tan^{-1}\left[\frac{\sum_{n=0}^{N-1} I_n(x, y)\sin(\Delta\varphi_n)}{\sum_{n=0}^{N-1} I_n(x, y)\cos(\Delta\varphi_n)}\right]. \tag{4}$$

Note that the phase values estimated in Eqs. (3) and (4) are wrapped from $-\pi$ to π due to the existence of $\tan^{-1}(\cdot)$, leading to phase ambiguity. To obtain monotonic phase maps, phase unwrapping is introduced to remove the wrapped phase discontinuities and then, $\Phi(x, y)$ can be obtained by absolute phase subtraction for reconstructing the shape of the object being measured. In the process, a fundamental assumption for achieving accurate 3D measurement is that the applied phase shifts must be both precise and known. For static scenes, this requirement can be satisfied by projecting a sequence of fringe patterns with fixed phase step intervals. However, in dynamic scenarios, a point on the object's surface moves from position S_m to a new position S'_m, as illustrated in Fig. 1(a), and is encoded by the projected pixels at different positions, respectively. As a result, the phase changes observed between consecutive fringe pattern frames comprise both the intended ideal phase shift and additional phase shift errors caused by the object's motion.

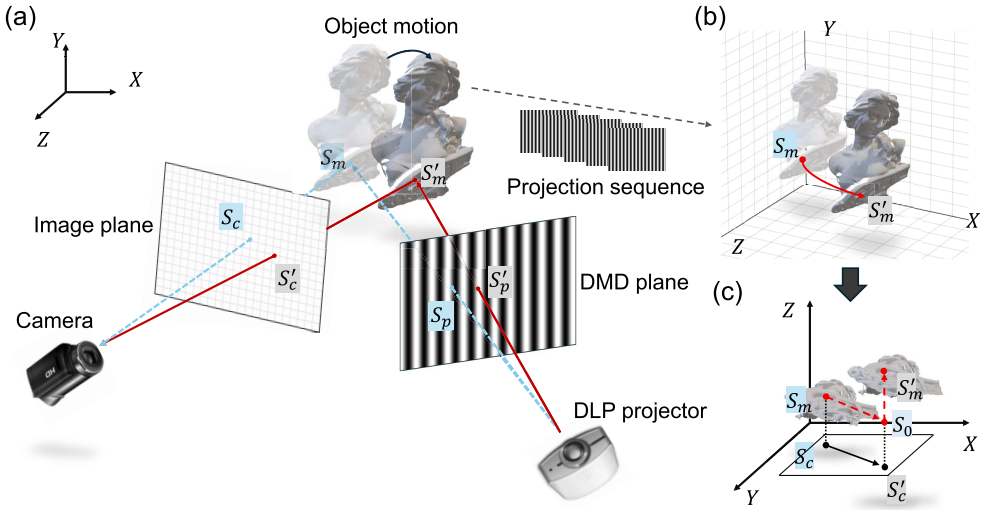

Fig. 1. Schematic diagram of dynamic structured light involving a moving object. (a) Phase variations induced by object motion. (b-c) Where the motion of a point on the surface is decomposed into in-plane motion within the X–Y plane and out-of-plane motion along the Z-direction.

As shown in Figs. 1(b) and (c), the 3D motion of a surface point from S_m to S'_m can be decomposed into two sequential components: (1) in-plane motion (X–Y) from S_m to S_0, and (2)

Optics EXPRESS

out-of-plane motion in Z-direction from S_0 to S'_m. Based on this decomposition, the total phase error $\Delta\phi_n(x, y)$ induced by object motion can likewise be divided into two distinct components: the phase errors $\Delta\phi_n^{x-y}(x, y)$, which is caused by the in-plane displacement, and $\Delta\phi_n^z(x, y)$, which is caused by the displacement in Z-direction, resulting in the following expression:

$$\Delta\phi_n(x, y) = \Delta\phi_n^{x-y}(x, y) + \Delta\phi_n^z(x, y). \tag{5}$$

Let $S_c(x, y)$ and $S'_c(x, y)$ denote the 2D projections in the fringe images of the same object point before and after motion in Fig. 1(c). Because of object motion, an additional displacement $(\Delta u, \Delta v)$ occurs between the two projections with $(u, v) = (x + \Delta u, y + \Delta v)$. Following Lu's motion method [25], the in-plane phase error caused by this displacement can be expressed explicitly as:

$$\Delta\phi_n^{x-y}(x, y) = \varphi(u, v) - \varphi(x, y). \tag{6}$$

Assuming that the background illumination $A(x, y)$ and modulation amplitude $B(x, y)$ remain constant during one projection period, the n-th fringe intensity under 3D motion can be written as:

$$\widetilde{I}_n(u, v) = A + B\cos\left[\varphi(x, y) + \Phi(x, y) + \Delta\varphi_n + \Delta\phi_n^{x-y}(x, y) + \Delta\phi_n^z(x, y)\right]. \tag{7}$$

Substituting Eq. (6) into Eq. (7), the fringe intensity after motion can be reformulated as:

$$\widetilde{I}_n(u, v) = A + B\cos\left[\varphi(u, v) + \Phi(x, y) + \Delta\varphi_n + \Delta\phi_n^z(x, y)\right]. \tag{8}$$

To describe this spatial correspondence more generally, a motion mapping function $M'_n(x, y)$ is introduced to describe the pixel displacement, such that $(u, v) = M'_n(x, y)$. Then Eq. (8) can be expressed in the following compact form:

$$\widetilde{I}_n(u, v) = \widetilde{I}_n(M'_n(x, y)) = A + B\cos\left[\varphi(M'_n(x, y)) + \Phi(x, y) + \Delta\varphi_n + \Delta\phi_n^z(x, y)\right]. \tag{9}$$

Furthermore, for arbitrary 3D motion, the displacement of each surface point along the Z-direction can be described as a distinct temporal trajectory:

$$z'_i(t) = f_i(t), \quad i = 1, 2, \ldots, N \tag{10}$$

where $f_i(t)$ denotes the temporal evolution of the Z-direction displacement for the i-th surface point. According to kinematics law in 3D space, the trajectory $f_i(t)$ of surface points are not identical and exhibit nonlinearity.

However, each trajectory can be locally approximated by its first-order Taylor expansion within a short temporal window $[t_0, t_0 + \Delta t]$:

$$f_i(t) \approx f_i(t_0) + v_{z,i}(t_0)\,(t - t_0), \tag{11}$$

where $v_{z,i}(t_0)$ represents the instantaneous velocity along the Z-direction at time t_0. Thus, the local displacement increment can be written as:

$$\Delta z_i = f_i(t + \Delta t) - f_i(t) \approx v_{z,i}(t_0)\,\Delta t. \tag{12}$$

This equation indicates that Δz_i remains approximately constant across different points within a sufficiently short time window. This first-order linearization allows nonlinear 3D trajectories could be represented as a sequence of locally linear segments, ensuring the system in Eq. (9) holds to simplify the motion compensation and phase error modeling process.

Assuming the captured image sequence is divided into a series of consecutive short time windows, the phase change $\Delta\phi_n^z(x, y)$ caused by Z-direction motion at the same point on the object can be considered constant within each window. Let ε denote this phase error. For

Optics EXPRESS

five consecutively captured fringe images $\widetilde{I}_n, (1 \le n \le 5)$, the phase errors introduced by the Z-direction motion of the object at a point (x, y) can be $\Delta\phi_n^z(x, y) = (n-3)\varepsilon(x, y)$, resulting in

$$\widetilde{I}_n(u, v) = A + B\cos\left[\varphi(M_n'(x, y)) + \Phi(x, y) + (n-1)\frac{\pi}{2} + (n-3)\varepsilon(x, y)\right]. \tag{13}$$

This equation establishes a fundamental relationship between the phase information and the observed fringe pattern, taking into account the phase error caused by object motion in 3D space. To recover the correct phase at each pixel (x, y), it is essential to estimate the phase error $\varepsilon(x, y)$ and then retrieve the corrected phase information.

It is worth noting that we chose five consecutive frames in our implementation. This is because, in the proposed pixel-wise phase recovery model, the unknowns include the background intensity A, fringe modulation B, object phase value Φ, and motion-induced phase error ε. Though A and B can be eliminated through linear combination, the remain two primary variables Φ and ε are coupled and can be solved if additional fringe frames exist. When the number of frames less than four, the system in Eq. (13) is underdetermined and the two variables are inseparable; when the number of frames is four, the system is solvable but its solution becomes highly sensitive to noise due to near-singular trigonometric denominators, which can be found in the subsequent relationship in Eq. (19). To overcome this problem, we employ a five-frame configuration that produces three independent equations and forms a well-conditioned system, effectively balancing solvability, noise robustness, and temporal consistency. If more than five consecutive frames are used, the assumption of the phase error constancy caused by the motion along Z-direction in Eq. (9) is not easy to hold.

Subsequently, we perform several simple operations between $\widetilde{I}_n$ based on the five-frame configuration, resulting in the following three equations:

$$E_1 = \widetilde{I}_4 - \widetilde{I}_2 = 2B\left[C_1\cos\Phi + S_1\sin\Phi\right], \tag{14}$$

$$E_2 = \widetilde{I}_5 + \widetilde{I}_1 - 2\widetilde{I}_3 = 2B\left[C_2\cos\Phi + S_2\sin\Phi\right], \tag{15}$$

$$E_3 = \widetilde{I}_5 - \widetilde{I}_1 = 2B\left[C_3\cos\Phi + S_3\sin\Phi\right], \tag{16}$$

where C_i and S_i are defined as:

$$\begin{cases} C_1 = \sin\left(\frac{\varphi_4+\varphi_2}{2}\right)\cos\left(\frac{\varphi_4-\varphi_2}{2} + \varepsilon\right), \\ S_1 = \cos\left(\frac{\varphi_4+\varphi_2}{2}\right)\cos\left(\frac{\varphi_4-\varphi_2}{2} + \varepsilon\right), \\ C_2 = \cos\left(\frac{\varphi_5+\varphi_1}{2}\right)\cos\left(\frac{\varphi_5-\varphi_1}{2} + 2\varepsilon\right) + \cos(\varphi_3), \\ S_2 = -\sin\left(\frac{\varphi_5+\varphi_1}{2}\right)\cos\left(\frac{\varphi_5-\varphi_1}{2} + 2\varepsilon\right) - \sin(\varphi_3), \\ C_3 = -\sin\left(\frac{\varphi_5+\varphi_1}{2}\right)\sin\left(\frac{\varphi_5-\varphi_1}{2} + 2\varepsilon\right), \\ S_3 = -\cos\left(\frac{\varphi_5+\varphi_1}{2}\right)\sin\left(\frac{\varphi_5-\varphi_1}{2} + 2\varepsilon\right). \end{cases} \tag{17}$$

Note that $\varphi_n = \varphi(M_n'(x, y))$ in above equations and the spatial coordinates are omitted for notation simplicity.

Combining Eq. (14) with Eqs. (15) and (16), respectively, we have

$$C_i\cos\Phi + S_i\sin\Phi = R_i \ (i = 1, 2, 3), \tag{18}$$

where $R_i = E_i/2B$. By applying Cramer's rule, the three equations in (18) lead us to the following relationship:

$$f(\varepsilon) = \tan\Phi = \frac{C_2R_1 - C_1R_2}{S_1R_2 - S_2R_1} = \frac{C_1R_3 - C_3R_1}{S_3R_1 - S_1R_3}. \tag{19}$$

Given that ε is small, applying the approximations $\sin\varepsilon \approx \varepsilon$ and $\cos\varepsilon \approx 1$, and neglecting higher order terms $o(\varepsilon^2)$, the above equation reduces to a linear algebraic equation:

$$K_1\varepsilon + K_0 = 0, \tag{20}$$

Optics EXPRESS

where K_1 and K_0 are coefficients determined by known intensity I_n and motion-related phase φ_n. For detailed derivation, see the Appendix in 4. The value of ε can thus be solved explicitly, allowing us to retrieve the phase based on Eq. (19) as follows:

$$\Phi = \arctan\left(f(\varepsilon)\right). \tag{21}$$

However, as shown in Eqs. (9) and (13), the phase value φ_n, on which K_0 and K_1 depend, is a function of the motion pattern associated with the corresponding pixel. Therefore, we need to estimate the correspondence between pixels in the observed fringe pattern after object motion and those in the initial fringe image, which will be introduced in the next subsection.

2.2. Fringe pattern correspondence estimation with transformers

To establish the correspondence between fringe patterns before and after object motion, we propose tracking the pixels of the initial fringe image across subsequent moving fringe images. This can be implemented by introducing a transformer-based deep learning method–COrrespondence TRansformer (COTR) [35].

In contrast to traditional image matching techniques, such as optical flow—which use local image features to establish correspondences between images without fringe patterns with high accuracy—COTR has been experimentally shown to match images with fringe patterns with considerable accuracy and density. We next provide a brief introduction to estimating fringe pattern correspondences using COTR.

In the context of fringe pattern matching, the architecture of COTR comprises three primary modular learning blocks — a backbone $\mathcal{E}(\cdot)$, a transformer $\mathcal{T}(\cdot)$, and a multiple layer perception (MLP) — as illustrated in Fig. 2. The backbone is implemented as a convolutional neural network (CNN) and is trained via a weight sharing policy to learn the underlying context representations with identical inductive bias for fringe images observed at various motion states, and the transformer is built in a convolutional encoder-decoder structure, where we denote the transformer encoder and decoder as $\mathcal{T}_{\mathcal{E}}(\cdot)$ and $\mathcal{T}_{\mathcal{D}}(\cdot)$, respectively. Taking the fringe images captured before and after object motion as the reference $\mathcal{I}_0$ and target $\mathcal{I}_t$, respectively, the backbone network learns to extract feature maps from each using shared weights, resulting in representative feature maps $\mathcal{E}(\mathcal{I}_0)$ and $\mathcal{E}(\mathcal{I}_t)$ of dimensions of $16 \times 16 \times 1024$. The two feature maps are then concatenated along the width dimension, resulting in a fused feature map as $[\mathcal{E}(\mathcal{I}_0), \mathcal{E}(\mathcal{I}_t)] \in \mathbb{R}^{16\times32\times1024}$. To ensure the following transformer to access the pixel locations across frames, a linear positional encoding $\mathcal{P}$ is applied to each spatial location of the fused feature map prior to the transformer encoding to retain spatial information. The positional encoding is defined as $\mathcal{P}(\mathbf{x}) = [p_1(\mathbf{x}), \ldots, p_{C/4}(\mathbf{x})]$ on a grid domain $\Omega(\mathbf{x})$ with $p_k(\mathbf{x}) = [\sin(k\pi\mathbf{x}^T), \cos(k\pi\mathbf{x}^T)] \in \mathbb{R}^4$. As a result, we obtain a context feature map with position embeddings as follows:

$$C = [\mathcal{E}(\mathcal{I}_0), \mathcal{E}(\mathcal{I}_t)] + \mathcal{P}(\Omega). \tag{22}$$

The resulting context feature map C is then fed into the transformer $\mathcal{T}$, along with a set of query points (denoted as $\{\mathbf{x}\}_0$) in the query fringe image. Without the need for an explicit cross-attention mechanism, self-attention enables the encoder $\mathcal{T}_{\mathcal{E}}$ to focus both intra-frame positional information and inter-frame motion dependencies. Subsequently, the transformer decoder $\mathcal{T}_{\mathcal{D}}$ is used to interpret the query-position-aware feature encoding, so that the MLP head could predict the corresponding point of each query in the frame after motion. The process can be expressed as:

$$\{\mathbf{x}\}_t = \mathrm{MLP}(\mathcal{T}_{\mathcal{D}}(\mathcal{T}_{\mathcal{E}}(C), \mathcal{P}(\{\mathbf{x}\}_0))). \tag{23}$$

We refer to see [35] for details of each modular in the COTR architecture.

COTR achieves matching point queries at arbitrary locations using the process described above. When multiple query points are provided as input, the network can perform dense

Optics EXPRESS

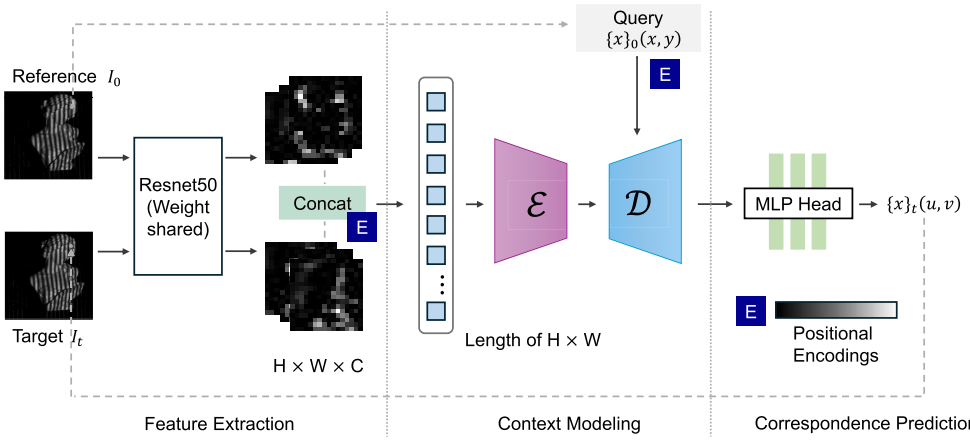

Fig. 2. COTR-based framework for tracking object motion in fringe images. Architectural details can be found in [35].

matching, predicting each target point independently. In this study, COTR is trained under supervised conditions by giving the ground truth location for each pixel after motion. Let the COTR neural network in Eq. (23) to be trained be $\mathcal{F}_\Theta$, with Θ the learnable parameters. The learning process is formulated by incorporating two loss measures: the sum of squared differences between the estimated correspondences and their ground-truth labels ($\mathcal{L}_{\text{ssd}}$), and the associated cycle-consistent error ($\mathcal{L}_{\text{cyc}}$):

$$\mathcal{L} = \mathcal{L}_{\text{ssd}} + \mathcal{L}_{\text{cyc}}, \tag{24}$$

where $\mathcal{L}_{\text{ssd}} = \sum_n ||\mathbf{x}_{t,n} - \mathcal{F}_\Theta(\mathbf{x}_{0,n}|(\mathcal{I}_0, \mathcal{I}_t))||_2^2$ and $\mathcal{L}_{\text{cyc}} = \sum_n ||\mathbf{x}_{0,n} - \mathcal{F}_\Theta(\mathcal{F}_\Theta(\mathbf{x}_{0,n}|(\mathcal{I}_0, \mathcal{I}_t))|(\mathcal{I}_0, \mathcal{I}_t))||_2^2$. In this work, the optimizer Adam is used to minimize the loss function for training on our dataset. The detailed training strategy and the evolution of the loss during training is discussed in Section 3.1.

2.3. Shape reconstruction algorithm under 3D motion

To illustrate the proposed moving-object measurement method, we present a COTR-based imaging framework, where the trained COTR model (Section 2.2) is integrated into the dynamic structured-light imaging process, as shown in Fig. 3. The measurement pipeline within this framework consists of five main steps, as outlined below.

Step 1: Based on the four-step PSP, the projector cyclically projects fringe patterns onto the moving object, while the corresponding deformed fringe image sets $\widetilde{I}_n(x,y)$, $n = 1, 2, \ldots, N$ are synchronously captured by the camera. we apply a sliding window mechanism of length 5 to partition the image sequence into short segments (referred to as windows).

Step 2: Within each window, a keyframe $\widetilde{I}_{key}$ (usually the middle image) is chosen as the reconstruction reference. COTR is then used to match $\widetilde{I}_{key}$ with the remaining images $\{\widetilde{I}_{sup_i}\}_{i=1}^4$ in the window, enabling accurate estimation of the object's motion.

Step 3: The estimated motion information $M'(x,y)$, obtained from COTR, defines a pixel-wise correspondence between the reference frame and subsequent fringe images, ie $(u,v) = M'(x,y)$. Before computing the phase error $\varepsilon(x,y)$, all fringe frames within each window are spatially aligned according to this mapping to ensure that the same surface point is tracked across frames. The aligned image frames $\tilde{I}_n(x,y) = I_n(M'_n(x,y))$ and the corresponding phase values $\varphi_n = \varphi(M'_n(x,y))$ are then substituted into Eq. (20) for solving $\varepsilon(x,y)$. This allows for accurate recovery of the correct phase Φ by Eq. (21).

Optics EXPRESS

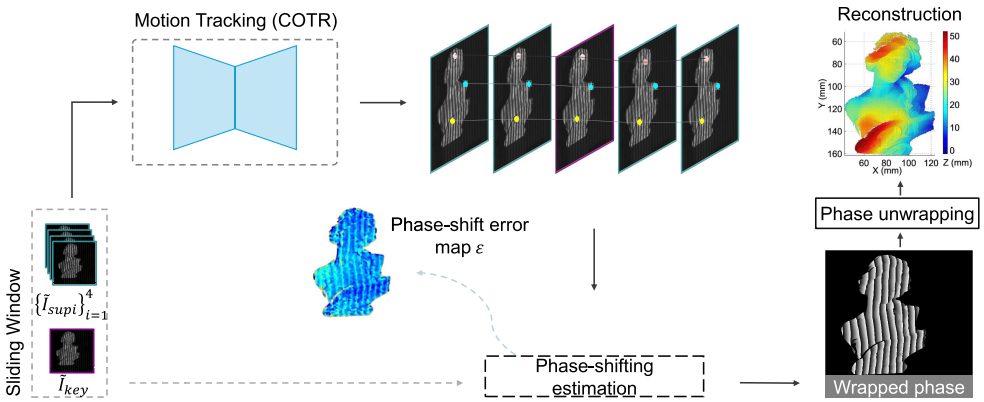

Fig. 3. Framework overview for measuring the shape of an object under 3D motion.

Step 4: In this step, the wrapped phase maps obtained from Step 3 are converted into continuous phase distributions using a widely adopted three-frequency phase-shifting algorithm. During imaging, two additional sets of lower-frequency fringe patterns are alternately projected along with the high-frequency ones. By ensuring that the wrapped phases across different frequencies remain spatially neighboring, possible phase offset and jump changes in dynamic measurement could be avoided as much as possible.

Step 5: The continuous phase is converted into the actual 3D surface shape using the calibrated camera–projector geometric model.

Step 6: The stride of the sliding window is set to 1, ensuring that each frame in the fringe image sequence is selected as a keyframe at least once. Each newly acquired fringe image is combined with the four preceding frames to form a new five-frame window. Subsequently, by repeating steps 2 to 5, a new 3D shape of the object can be reconstructed.

3. Experimental results

In this work, we build a fringe projection profiling system comprised by a high-speed camera (Revealer M120) with a resolution of 512×512 pixels at 30 fps and a DLP projector (Light Crafter 4500) with a resolution of 912×1140 pixels. The camera lens has a focal length of 20 mm and an aperture of f/1.5. With this structured light imaging system, we next present the dataset and training details for COTR in Section 3.1, the quantitative evaluation of reconstruction performance in Section 3.2, the evaluation on general moving objects in Section 3.3, the evaluation under complex 3D motion scenarios in Section 3.4, and the ablation studies in Section 3.5.

3.1. Dataset and training details

Dataset. To enable the trained COTR model to predict reliable correspondences for estimating motion between fringe frames observed at the initial and moving states, we construct a dataset by capturing fringe images of various moving objects using the above FPP system, where the projector sequentially projects fringe patterns with periods of 70, 64, and 59. The objects involved in the dataset include sculptures, toys, and industrial components. To observe general 3D rigid-body motion patterns, a sandbox (depth $\approx$ 5 cm) is placed 80 cm from the FPP system and then, each of the objects is placed on the sandbox to trigger continuous random rigid-body motion in 3D space manually. For each pose change, the projector sequentially projects 12 fringe patterns with the specified spatial periods, each containing four phase shifts ($0, 0.5\pi, \pi, 1.5\pi$), while the camera acquires the modulated fringe patterns, from which pairs of images corresponding to two distinct positions are extracted to represent the object's pre- and post-motion states.

Optics EXPRESS

Following this strategy, 45 dynamic scenes are simulated by arranging or combining the tracking target objects, with 60 distinct motion patterns in terms of displacement captured for each scene, yielding 2700 pairs of pre- and post-motion fringe images. Several demonstrations of the generated fringe pattern sequences are shown in Fig. 4. The first column illustrates representative motion scenes, while the second to fourth columns show fringe images at the initial, second, and final poses. For supervised training, both pre- and post-motion fringe images are input into the network along with ground-truth correspondence pairs (the last column). To obtain these correspondences, the projector additionally displayed a blank image, and robust local optical flow (RLOF) [36] method is applied to blank-image pairs to compute a dense displacement field across object positions. For each pair of fringe images, the ground truth consists of 100 high-quality correspondences, randomly sampled the dense field, which serve as the queries $\mathbf{x}_0$ and corresponding targets $\mathbf{x}_t$ for COTR training.

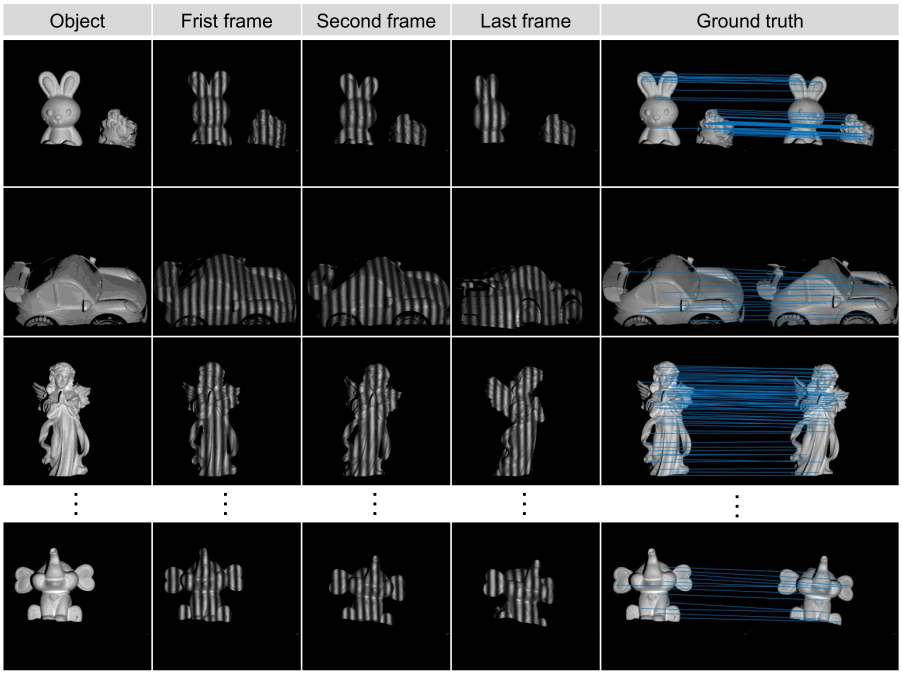

Fig. 4. Sample demonstrations in the constructed dataset.

Training details. We adopt a transfer learning strategy, initializing the COTR model with its original pre-trained weights and fine-tuning it on our dataset. During the training process, fringe images are cropped to 256×256 pixels to ensure computational efficiency and affordable training costs on our computation platform (Intel Xeon Silver 4215R CPU (3.20 GHz) with 32 GB RAM and two NVIDIA GeForce RTX 4090 GPUs). The training hyperparameters are as follows: the learning rate is 10^{-4} and the batch size is 84. It is worth noting that data augmentation, including random adjustment of brightness, normalization, and rotation, is used to improve data diversity.

In this study, 2400 image pairs from the dataset are used for training and 300 pairs for validation. The corresponding training and validation loss curves are presented in Fig. 5, showing convergence after approximately 1400 epochs, with the entire training process requiring about 42 hours.

To show the fringe image matching performance of the trained COTR model, several objects, such as statues and car models that are not seen in the training process, are used for testing. For each test object, 30 pairs of pre- and post-motion fringe images are acquired. From each pair,

Optics EXPRESS

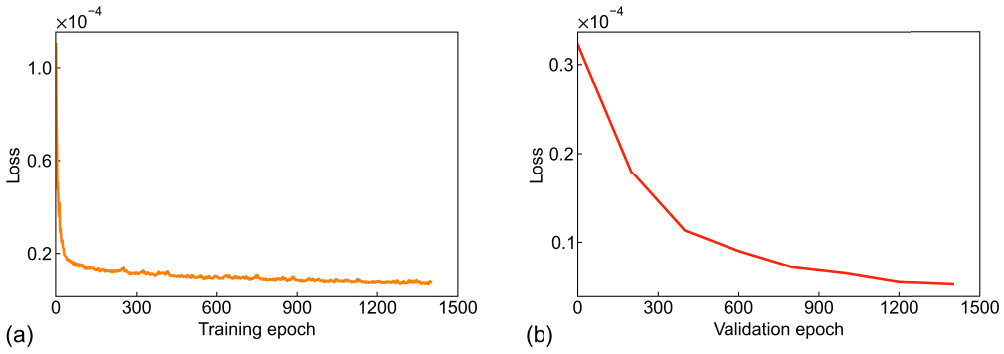

Fig. 5. Evolution of the model loss over epochs during (a) training and (b) validation.

100 query points are randomly sampled, and their corresponding positions in the post-motion image are estimated using the fine-tuned COTR model. In parallel, optical flow is applied to the same image pairs to provide comparison results. The predicted positions are then evaluated against the ground truth to compute the average endpoint error (AEPE) for each object, serving as a measure of matching accuracy. A total of five objects are tested, and the corresponding results are summarized in Table 1. As shown, the AEPEs obtained with the trained model are approximately two orders of magnitude lower than those achieved with the optical flow method. The experimental findings indicate that the fine-tuned COTR model attains subpixel-level motion-tracking accuracy, even in the presence of interference from projected fringe patterns.

Table 1. Quantitative comparison of optical flow method and the trained COTR model in terms of AEPE (unit: pixel)

Method	Object 1	Object 2	Object 3	Object 4	Object 5
Optical flow	14.73	12.32	21.06	15.03	17.12
Our method	0.49	0.47	0.72	0.83	0.54

3.2. Validation of 3D reconstruction accuracy

To evaluate the reconstruction performance of the proposed method, an experimental setup, shown in Fig. 6(a), was constructed using the same camera and DLP projector described earlier. A standard ceramic plate was mounted on a motorized translation stage, positioned approximately 80 cm from the system and oriented at a 40° angle relative to the camera's optical axis. Figure 6(b) illustrates the acquisition process of the fringe images. The translation stage advanced in 1 mm increments, while the system synchronously acquired a sequence of fringe images, indexed in chronological order as $\mathcal{I} = I_1, I_2, \ldots, I_6$. At t_3, we additionally projected a set of 12 step phase-shifted fringe images and used their reconstruction results as ground truth.

Subsequently, 3D reconstruction of the ceramic plate was performed using multiple methods, including the standard four-step phase-shifting method (PSP) [9], Guo's method [21], Lu's method [25], Wang's method [26], and our proposed method. Figure 6(c) presents the reconstruction results obtained by various methods, while Fig. 6(d) provides the error histograms for each method relative to the truth of the ground. Figure 6(e) shows the cross-sectional profiles extracted in $y = 325$ pixels for each method,

From Fig. 6, it can be observed that the reconstruction results from standard PSP exhibit obvious ripple distortions. Guo's method smooths some of the ripples caused by motion but still leaves significant residual motion errors. Lu's method introduces a motion compensation model. However, it only compensates for displacements within the X-Y plane and is ineffective

Optics EXPRESS

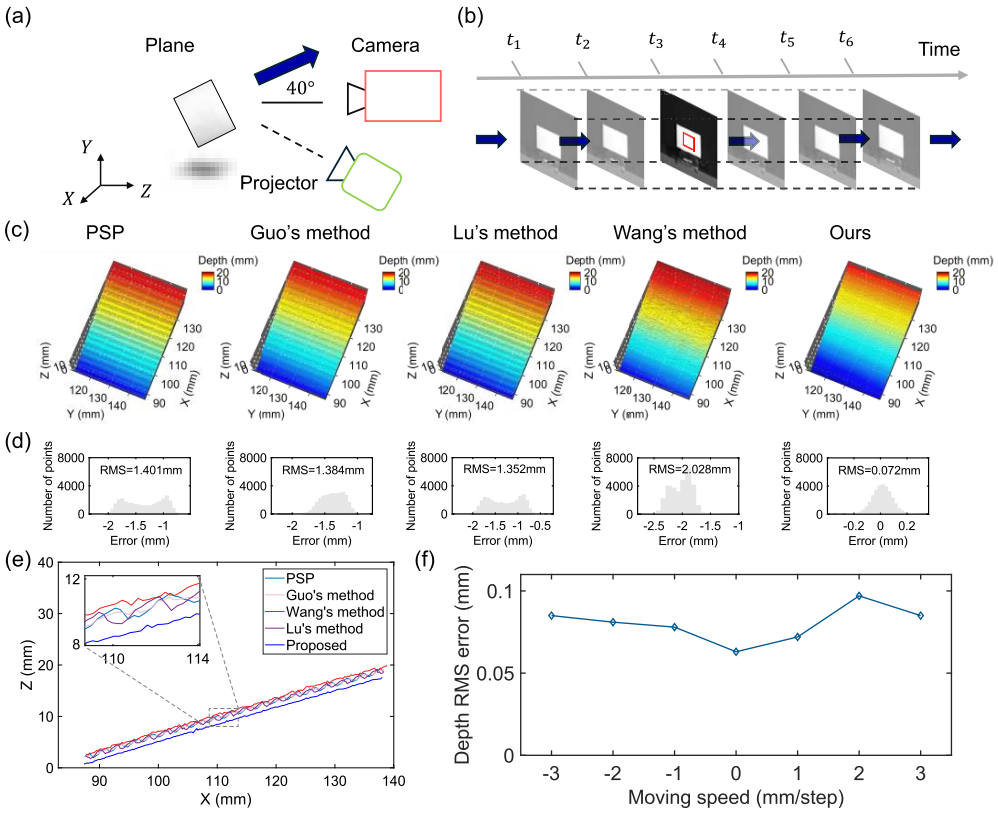

Fig. 6. Validation process and results with a standard ceramic plate. (a) and (b) are the schematic diagram of the experimental setup and the process of fringe image acquisition for the moving ceramic plate. (c) The reconstruction results of the standard four-step PS method (PSP) [9], Guo's method [21], Lu's method [25], Wang's method [26], and our method. (d) shows the histograms of reconstruction error for each method in (c) relative to the ground truth. (e) Cross-sectional profiles extracted at y=325 pixels for each method. (f) RMSE distribution for the proposed method under various motion amplitudes.

at correcting errors caused by motion in the Z-direction. Although Wang's method aligns object points on the camera plane through projection-point tracking, it simplifies the phase-shift error in its assumption, which leads to even larger phase errors. In contrast, our method produces high-quality reconstructed surfaces. The root mean square errors (RMSE) within the red ROI area are as follows: 1.401 mm for PSP, 1.384 mm for Guo's method, 1.352 mm for Lu's method, 2.028 mm for Wang's method, and a significantly reduced 0.072 mm for our method. As shown by the cross-sectional curves in Fig. 6(e), the results of our method closely match the ground truth, with almost no noticeable contour fluctuations.

Furthermore, to evaluate the robustness of the proposed method under larger motion amplitudes, the size of the platform movement step was varied to ±1 mm, ±2 mm, and ±3 mm per step. Reconstruction accuracy within the same ROI was analyzed, with RMS errors under each motion amplitude presented in Fig. 6(f). The experimental results demonstrate that, even under large object displacements, our method maintains strong robustness and high measurement accuracy.

Optics EXPRESS

3.3. Evaluation on general moving objects

To further evaluate the performance of the proposed method on more general scenes, three dynamic scenes with different states were tested. The first scene included an "Aphrodite" statue with complex surface shown in Fig. 7(a), which also underwent translational motion at a 40 ° angle relative to the camera's optical axis at a step size of 1 mm.

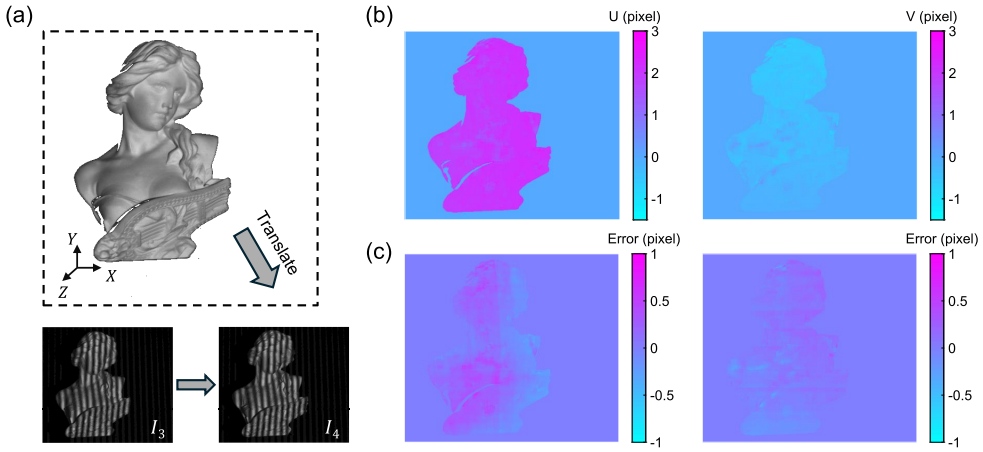

Fig. 7. Analysis of COTR-based motion tracking for the "Aphrodite" statue. (a) The statue of "Aphrodite" (translation at a 40° angle in the X-Z plane), with two temporally adjacent fringe images (I_3 and I_4) captured during the motion. (b) Estimated horizontal (U) and vertical (V) displacements of the object between I_3 and I_4 obtained from COTR tracking. (c) Corresponding displacement tracking errors in the U- and V-directions.

Using a pair of consecutively captured fringe images (I_3 and I_4) as an example, Fig. 7(b) illustrates the estimated motion vector fields in both horizontal and vertical directions, obtained using COTR. We also applied RLOF method to pure object images captured at the corresponding positions for comparison. Figure 7(c) shows the tracking error maps in the horizontal and vertical directions. According to the error distribution, most pixel errors are concentrated within the range of $[-1, 1]$, indicating that COTR effectively suppresses fringe interference and still achieves accurate object position tracking even in images with periodic stripe backgrounds.

Figure 8(a) presents the reconstruction results obtained by PSP, Guo's method, Lu's method, and the proposed method, as well as the ground-truth that was measured by 12-step PS algorithm. The PSP method suffers from severe stripe artifacts and boundary blurring under strong dynamic interference, with particularly noticeable discontinuities along the object's edges. Guo's method indirectly compensates for motion-induced errors and suppresses some artifacts to a certain extent. However, due to inaccurate boundary alignment, significant errors persist in the edge regions. Lu's method is able to reconstruct the overall shape of the object, but due to the absence of compensation for Z-direction motion errors, periodic ripple artifacts remain in the results. In contrast, our method integrates motion compensation with pixel-level tracking, enabling more accurate contour restoration and effectively suppressing geometric distortions caused by motion. The RMSE values within the ROI are as follows: 1.129 mm for PSP, 1.119 mm for Guo's method, 1.379 mm for Lu's method, and a significantly reduced 0.096 mm for our method. A cross-sectional profile was extracted at pixel row $y = 268$, and the corresponding depth variations are shown in Fig. 8(b). From the profile plot, it is evident that the contour reconstructed by the proposed method closely matches the ground truth, whereas the other methods exhibit varying degrees of deviation and fluctuation. Figure 8(c) further shows the phase error map ε induced by Z-direction motion, as estimated by the proposed method.

Optics EXPRESS

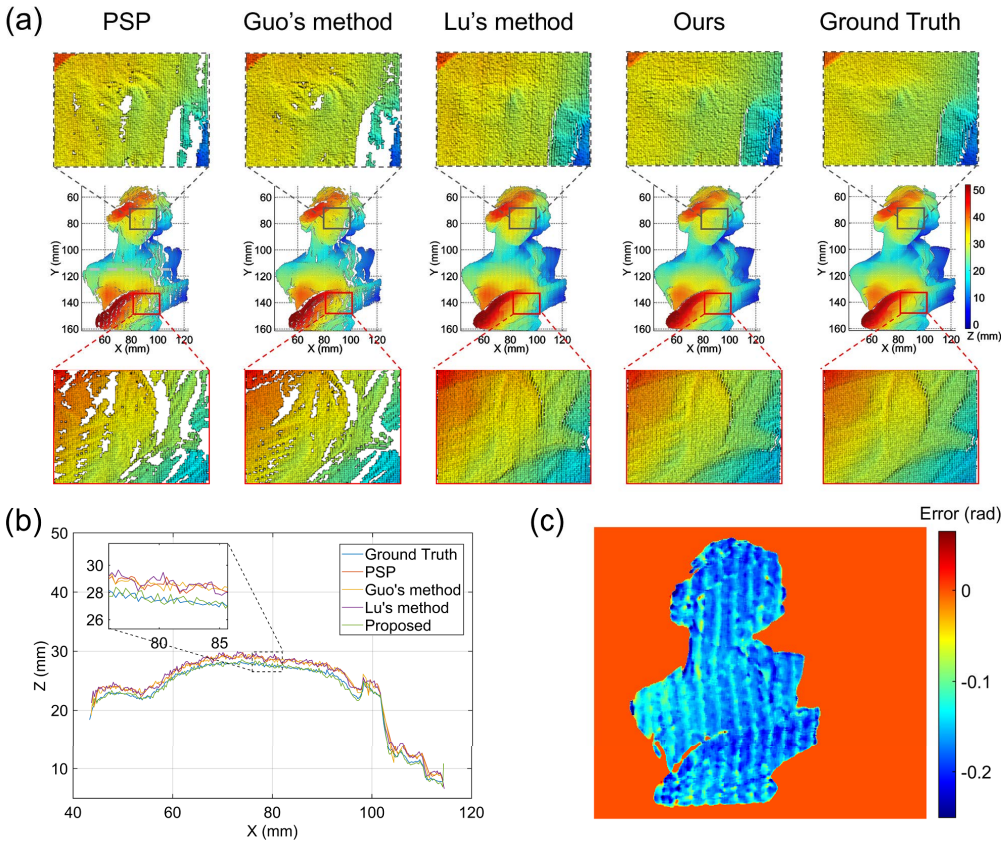

Fig. 8. (a) Surface reconstruction results obtained using the PSP, Guo's, Lu's, our method, and the ground truth. For each method, two regions of interest (ROI A and ROI B) are shown in enlarged views for detailed comparison. (b) Cross-sectional profiles extracted along pixel row $y = 268$, quantitatively comparing the reconstruction accuracy of all methods. (c) Distribution of the additional phase-shift error map ε (rad) induced by motion along the Z-direction.

For objects moving in arbitrary directions, the motion trajectories of individual surface points and the resulting phase errors may differ significantly, posing challenges to global motion estimation and compensation. However, the proposed method performs pixel-wise motion tracking and independently estimates the phase error introduced by each point's motion, providing a reliable foundation for accurate motion compensation. As shown in Fig. 9(a), the second experiment features a multi-object scene, consisting of a stationary "Astronaut" statue and a "Marseille" statue rotating around the y-axis, with an angular step of approximately 0.04 rad per step. This setup is used to valuate the algorithm's reconstruction performance in the presence of multiple objects and varying motion states. Figure 9(b) shows the motion tracking results between fringe images I_3 and I_4 based on COTR, while Fig. 9(c) presents the phase-shift error map ε (rad) induced by motion along the Z-direction. As shown in Fig. 9(d), the third scene involves a fan model with a smooth surface and weak texture, rotating around the x-axis at a similar angular speed of approximately 0.04 rad per step. The displacement estimates and phase error plots corresponding to this experiment are shown in Fig. 9(e) and (f), respectively.

The 3D reconstruction results obtained using PSP, Guo's method, Lu's method, and our method, as well as the ground truth, are shown in Fig. 10. Experimental results demonstrate that

Optics EXPRESS

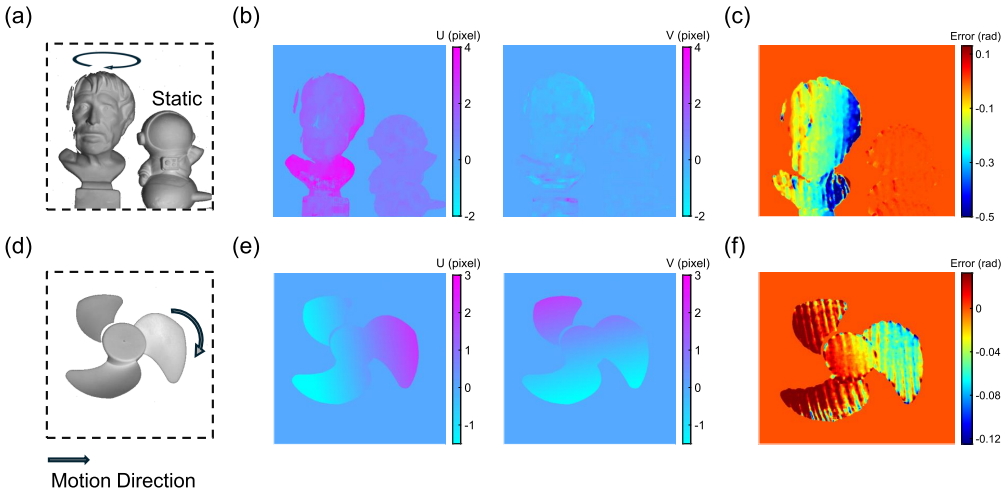

Fig. 9. 3D reconstruction results in complex dynamic scenes. (a) A multi-object dynamic scene consisting of a static "Astronaut" statue and a rotating "Marseille" statue, with the latter rotating around the y-axis at an angular velocity of approximately 0.04 rad/step. (b) Estimated horizontal (U-axis) and vertical (V-axis) displacements for the scene in (a) (unit: pixel). (c) Additional phase-shift error ε (rad) induced by the Z-direction motion in (a). (d) A rotating fan with a smooth surface and weak texture, rotating about the x-axis at a similar angular velocity. (e)-(f) Corresponding displacement estimates and motion error map for the scene in (d).

our method maintains strong adaptability and stability when applied to complex dynamic scenes, including multi-object configurations, multi-directional motion, and surfaces with weak texture.

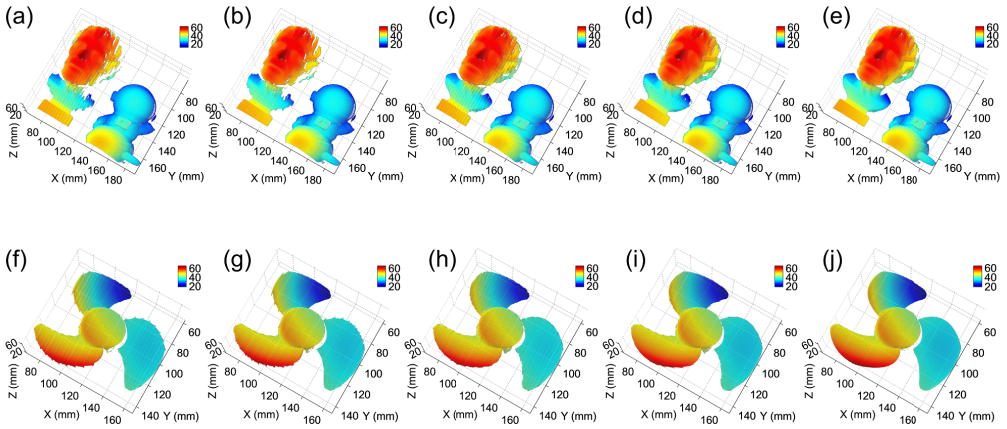

Fig. 10. The reconstruction results of two scenarios using different methods. (a–e) and (f–j) show the reconstruction results using different methods (PSP, Guo's, Lu's, ours) and the ground truth, respectively.

In the next experiment, we conducted 3D reconstruction on a continuously changing dynamic scene. The test object, a "David" statues, was mounted on a motorized rotating platform, which rotated continuously at an angular velocity of approximately 0.05π rad/s. Figure 11 presents a comparison of reconstruction results at multiple time points between the standard PSP and our method. The experimental results demonstrate that the proposed method significantly reduces the

Optics EXPRESS

reconstruction errors caused by motion of objects. (Complete visualization results are provided in the Visualization 1, where both camera capture and 3D playback were set to 30 fps.)

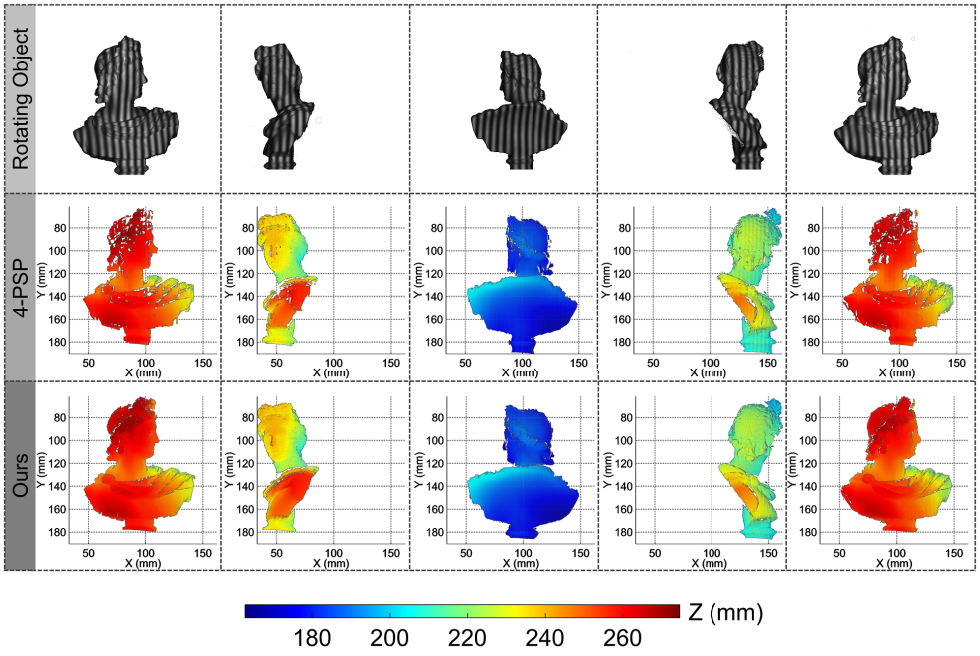

Fig. 11. Comparison of measurement results in rotating scenes at different times using standard PSP and our proposed method.

3.4. *Evaluation on complex 3D motion scenarios*

To further assess the capability of the proposed 3D motion-induced phase error compensation method, we designed two additional 3D motion experiments in which the object undergoes simultaneous translation and rotation. In the first experiment, the object undergoes translational motion along the X-axis at 1 mm per step, while the rotation stage rotates clockwise at 0.04 rad per step. In the second experiment, the object is translated along the Z-axis with the same step size and rotational rate as in the first experiment. After that, by applying the proposed method, we can reconstruct the 3D shapes of the object for each experiment. In addition, the traditional PSP method is also applied to obtain the 3D shapes for comparison.

The reconstructed results are shown in Fig. 12, where (a) and (d) depict the 3D shapes obtained using traditional PSP for the first and second experiments, respectively, and (b) and (e) show the corresponding 3D shapes reconstructed by our method. The estimated phase error maps for both experiments are shown in (c) and (f), respectively. It is obvious that the out-of-plane translation leads to large phase error that that of the in-plane translation. By comparison, it can be observed that the proposed method maintains accurate reconstruction for most surface areas, demonstrating the proposed tracking-plus-compensation framework maintains stable reconstruction performance even under large and complex 3D motion involving both translation and rotation. In addition, we found that the Z-direction displacement of surface points increases proportionally with their distance from the rotation axis as the rotation proceeds, introducing cumulative phase deviations of up to approximately 1 rad in the outer regions. For the second experiment with translation along Z-direction, the phase error estimation may still fail due to the excessively large displacement in Z-direction, as shown in the marked regions in Fig. 12(e) and (f). To mitigate this limitation, one may increase the projection rate or combine the proposed

Optics EXPRESS

framework with geometry-constraint-based phase unwrapping or Gray-code-based unwrapping techniques [37,38].

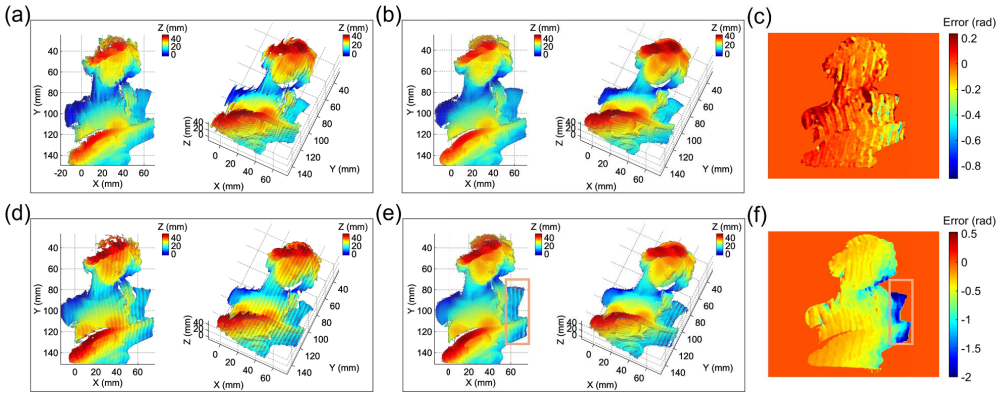

Fig. 12. Demonstration of experimental results for two complex motion scenarios. (a) and (b) show the reconstruction results of the first experiment using the traditional PSP method and our method, respectively, with (c) presenting the phase error map ε estimated by our method. (d) and (e) show the corresponding results for the second experiment, and (f) illustrates the phase error map ε estimated by our method for this experiment.

3.5. Ablation studies

To access the contribution of the COTR model to the reconstruction quality under fringe illumination, we compare to three representative optical flow models, including RAFT [39], FlowFormer [40], and GMFlow [41], by using the data from the first compound 3D motion experiment described above (see Fig. 12(a)–(e)). Each of these models is finetuned in our dataset and then is used to replace the COTR model in step 2 of Section 2.3 to obtain the comparison result. We evaluate the ablation performance by computing the MAE and RMSE of the reconstructed 3D shapes for each model relative to the ground truth. The results are listed in Table 2.

Table 2. Comparison to the representative optical flow models in term of 3D reconstruction accuracy. (Unit: mm)

Method	RAFT	FlowFormer	GMFlow	Ours
MAE	0.372	0.120	0.139	0.076
RMSE	0.490	0.157	0.181	0.092

The quantitative comparison demonstrates that our finetuned COTR model significantly outperforms the recent optical flow models in terms of 3D reconstruction accuracy. Specifically, for MAE, our method achieves 0.076 mm, which is substantially lower than RAFT (0.372 mm), FlowFormer (0.120 mm), and GMFlow (0.139 mm), indicating more accurate point-wise correspondence predictions. Similarly, in terms of RMSE, our approach attains 0.092 mm, markedly improving over RAFT (0.490 mm), FlowFormer (0.157 mm), and GMFlow (0.181 mm), reflecting reduced overall deviation from the ground truth. These results suggest that our model provides more precise 3D reconstructions, effectively capturing subtle geometric details compared to existing flow-based methods.

Optics EXPRESS

4. Conclusion

In this study, we proposed a novel method for precise 3D shape measurement of rigid objects moving in arbitrary directions. The approach begins with pixel-level motion tracking across a sequence of fringe images, enabling accurate displacement estimation. This motion information is then incorporated into a compensation framework to correct phase errors induced by object movement in the X–Y plane. Additionally, by estimating phase offsets caused by displacement along the Z-direction on a per-pixel basis, the method produces a corrected phase map that serves as the foundation for accurate. It worth noting that, this study mainly focuses on the task of reconstructing objects with 3D reconstruction. The effectiveness of the method was validated under real motion conditions through reconstruction experiments in representative dynamic scenes. Compared with existing techniques, our approach achieves higher accuracy and better preserves fine geometric details of moving objects in 3D space. Experiments across diverse scenarios further demonstrate its robustness and adaptability.

While this study primarily focuses on rigid-body motion, the proposed method is possible applied to 3D measurement of non-rigid objects, provided stable image motion tracking is achievable via suitable image correspondence models. For deformable objects, such as rubber-like materials, the motion-error compensation framework remains applicable when precise surface motion tracking is not required. It is also worth noting that, since the phase error estimation relies on a five-frame sliding window, the method works reliably for small inter-frame motions. Standard frame-rate setups are sufficient for slow to moderate motion, whereas high-frame-rate cameras and projectors are necessary for fast-moving objects to ensure accurate phase estimation.

Overall, this work establishes a solid theoretical and practical foundation for 3D motion reconstruction of rigid objects moving in arbitrary directions. From the application perspective, our method enhances existing PSP/FPP-based structured light systems by providing plug-and-play, physically interpretable, and scalable phase error compensation compatible with multiple fringe tracking algorithms. Future work will aim to reduce the number of required fringe patterns and expand the proposed method to measure deformable objects, improving measurement efficiency and enabling high-speed or real-time 3D measurement applications.

Appendix

Detailed derivation of Eq. (20) is presented as follows:

Given that ε is small, we can put $\sin\varepsilon \approx \varepsilon$ and $\cos\varepsilon \approx 1$. Then Eq. (19) can be further rewritten as:

$$\frac{C_2R_1 - C_1R_2}{S_1R_2 - S_2R_1} = \frac{C_1R_3 - C_3R_1}{S_3R_1 - S_1R_3}, \tag{25}$$

Optics EXPRESS

where

$$\begin{aligned}
C_1 &= \sin\left(\frac{\varphi_4+\varphi_2}{2}\right)\cos\left(\frac{\varphi_4-\varphi_2}{2}+\varepsilon\right)\\
&= \sin\left(\frac{\varphi_4+\varphi_2}{2}\right)\left[\cos\left(\frac{\varphi_4-\varphi_2}{2}\right)-\varepsilon\sin\left(\frac{\varphi_4-\varphi_2}{2}\right)\right]\\
S_1 &= \cos\left(\frac{\varphi_4+\varphi_2}{2}\right)\cos\left(\frac{\varphi_4-\varphi_2}{2}+\varepsilon\right)\\
&= \cos\left(\frac{\varphi_4+\varphi_2}{2}\right)\left[\cos\left(\frac{\varphi_4-\varphi_2}{2}\right)-\varepsilon\sin\left(\frac{\varphi_4-\varphi_2}{2}\right)\right]\\
C_2 &= \cos\left(\frac{\varphi_5+\varphi_1}{2}\right)\cos\left(\frac{\varphi_5-\varphi_1}{2}+2\varepsilon\right)+\cos(\varphi_3)\\
&= \cos\left(\frac{\varphi_5+\varphi_1}{2}\right)\left[\cos\left(\frac{\varphi_5-\varphi_1}{2}\right)-2\varepsilon\sin\left(\frac{\varphi_5-\varphi_1}{2}\right)\right]+\cos(\varphi_3)\\
S_2 &= -\sin\left(\frac{\varphi_5+\varphi_1}{2}\right)\cos\left(\frac{\varphi_5-\varphi_1}{2}+2\varepsilon\right)-\sin(\varphi_3)\\
&= -\sin\left(\frac{\varphi_5+\varphi_1}{2}\right)\left[\cos\left(\frac{\varphi_5-\varphi_1}{2}\right)-2\varepsilon\sin\left(\frac{\varphi_5-\varphi_1}{2}\right)\right]-\sin(\varphi_3)\\
C_3 &= -\sin\left(\frac{\varphi_5+\varphi_1}{2}\right)\sin\left(\frac{\varphi_5-\varphi_1}{2}+2\varepsilon\right)\\
&= -\sin\left(\frac{\varphi_5+\varphi_1}{2}\right)\left[\sin\left(\frac{\varphi_5-\varphi_1}{2}\right)+2\varepsilon\cos\left(\frac{\varphi_5-\varphi_1}{2}\right)\right]\\
S_3 &= -\cos\left(\frac{\varphi_5+\varphi_1}{2}\right)\sin\left(\frac{\varphi_5-\varphi_1}{2}+2\varepsilon\right)\\
&= -\cos\left(\frac{\varphi_5+\varphi_1}{2}\right)\left[\sin\left(\frac{\varphi_5-\varphi_1}{2}\right)+2\varepsilon\cos\left(\frac{\varphi_5-\varphi_1}{2}\right)\right]
\end{aligned} \tag{26}$$

Equation (25) is derived as follows:

$$R_1^2(C_2S_3-C_3S_2)-R_1R_3(C_2S_1-C_1S_2)+R_1R_2(C_3S_1-C_1S_3)=0 \tag{27}$$

Define:

$$\begin{cases}
w_1=\sin\left(\frac{\varphi_4+\varphi_2}{2}\right), & w_2=\cos\left(\frac{\varphi_4+\varphi_2}{2}\right), & w_3=\sin\left(\frac{\varphi_4-\varphi_2}{2}\right),\\
w_4=\cos\left(\frac{\varphi_4-\varphi_2}{2}\right), & q_1=\sin\left(\frac{\varphi_5+\varphi_1}{2}\right), & q_2=\cos\left(\frac{\varphi_5+\varphi_1}{2}\right),\\
q_3=\sin\left(\frac{\varphi_5-\varphi_1}{2}\right), & q_4=\cos\left(\frac{\varphi_5-\varphi_1}{2}\right).
\end{cases} \tag{28}$$

Then

$$\begin{cases}
C_2S_3-C_3S_2 = -(q_3+2\varepsilon q_4)(q_4-2\varepsilon q_3)-(q_3+2\varepsilon q_4)\left[\cos(\varphi_3)q_2+\sin(\varphi_3)q_1\right]\\
\qquad\qquad = L_1+L_2\varepsilon+o(\varepsilon^2),\\
C_2S_1-C_1S_2 = (q_4-2\varepsilon q_3)(w_4-\varepsilon w_3)(q_2w_2+w_1q_1)+(w_4-\varepsilon w_3)\left[w_2\cos(\varphi_3)+w_1\sin(\varphi_3)\right]\\
\qquad\qquad = L_3+L_4\varepsilon+o(\varepsilon^2),\\
C_3S_1-C_1S_3 = (q_3+2\varepsilon q_4)(w_4-\varepsilon w_3)(w_1q_2-q_1w_2)\\
\qquad\qquad = L_5+L_6\varepsilon+o(\varepsilon^2),
\end{cases} \tag{29}$$

where:

$$\begin{aligned}
L_1 &= -q_3\left[q_4+\cos(\varphi_3)q_2+\sin(\varphi_3)q_1\right],\\
L_2 &= -2q_4^2+2q_3^2-2q_4\left[\cos(\varphi_3)q_2+\sin(\varphi_3)q_1\right],\\
L_3 &= q_4w_4(q_2w_2+w_1q_1)+w_4\left[w_2\cos(\varphi_3)+w_1\sin(\varphi_3)\right],\\
L_4 &= -\left(q_4w_3+2q_3w_4\right)(q_2w_2+w_1q_1)-w_3\left[w_2\cos(\varphi_3)+w_1\sin(\varphi_3)\right],\\
L_5 &= q_3w_4(-q_1w_2+w_1q_2),\\
L_6 &= (2q_4w_4-q_3w_3)(w_1q_2-q_1w_2).
\end{aligned} \tag{30}$$

Optics EXPRESS

As $R_i = \frac{E_i}{2B}$, multiplying both sides of Eq. (27) by $2B$, and substituting Eq. (29) into it, we obtain an expression that can be further rewritten as:

$$K_1 + K_0\varepsilon + o(\varepsilon^2) = 0 \tag{31}$$

and

$$\begin{aligned} K_1 &= E_1^2 L_1 - E_1 E_3 L_3 + E_1 E_2 L_5, \\ K_0 &= E_1^2 L_2 - E_1 E_3 L_4 + E_1 E_2 L_6, \end{aligned} \tag{32}$$

where K_1 and K_0 are coefficients related to known parameters $\widetilde{I}_n, \varphi_n$. Considering that the high-order small terms $o(\varepsilon^2)$ can be neglected, then Eq. (20) is obtained.

Funding. National Natural Science Foundation of China (62375078, 62073123, 12572209); National Key Laboratory of Ship Structural Safety (NAKLAS2024KF016-S); Natural Science Foundation of Henan Province (242300421708); Key Research Project Plan for Higher Education Institutions in Henan Province (24ZX011); Science and Technology Innovation Project of Chinese Academy of Traditional Chinese Medicine (ZN2024A02); Training Plan for Young Backbone Teachers in Undergraduate Universities in Henan Province (2023GGJS058); Cultivation Programme for Young Backbone Teachers in Henan University of Technology; Open Fund of Institute for Complexity Science (CSKFJJ-2024-3).

Disclosures. The authors declare no conflict of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. S. Zhang, "Recent progresses on real-time 3d shape measurement using digital fringe projection techniques," Opt. Lasers Eng. **48**(2), 149–158 (2010).
2. B. Li, Y. An, D. Cappelleri, *et al.*, "High-accuracy, high-speed 3d structured light imaging techniques and potential applications to intelligent robotics," Int. J. Intell. Robot. Appl. **1**(1), 86–103 (2017).
3. S. Zhang, "High-speed 3d imaging with digital fringe projection techniques," in *Tribute to James C. Wyant: The Extraordinaire in Optical Metrology and Optics Education*, vol. 11813 (SPIE, 2021), pp. 226–233.
4. C. Zuo, J. Qian, S. Feng, *et al.*, "Deep learning in optical metrology: a review," Light: Sci. Appl. **11**(1), 1–54 (2022).
5. S. Zhang, "High-speed 3d shape measurement with structured light methods: A review," Opt. Lasers Eng. **106**, 119–131 (2018).
6. L. Lu, C. Bu, Z. Su, *et al.*, "Generative deep-learning-embedded asynchronous structured light for three-dimensional imaging," Adv. Photonics **6**(04), 046004 (2024).
7. Q. He, Z. Zheng, B. Zhang, *et al.*, "Three-dimensional surface-shape measurement of moving billets using phase-shifting profilometry," IEEE Trans. Instrum. Meas. **73**, 1–11 (2024).
8. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-d object shapes," Appl. Opt. **22**(24), 3977–3982 (1983).
9. J. Salvi, S. Fernandez, T. Pribanic, *et al.*, "A state of the art in structured light patterns for surface profilometry," Pattern Recog. **43**(8), 2666–2680 (2010).
10. S. Van der Jeught and J. J. Dirckx, "Real-time structured light profilometry: a review," Opt. Lasers Eng. **87**, 18–31 (2016).
11. H. Xing, S. She, J. Wang, *et al.*, "High-frame rate, large-depth-range structured light projector based on the step-designed led chips array," Opt. Express **32**(14), 24117–24127 (2024).
12. W. Yin, H. Zhao, Y. Ji, *et al.*, "High-resolution, wide-field-of-view, and real-time 3d imaging based on spatial-temporal speckle projection profilometry with a vcsel projector array," ACS Photonics **11**(2), 498–511 (2024).
13. S. Jeon, H. G. Lee, J. S. Lee, *et al.*, "Motion-induced error reduction for motorized digital fringe projection system," IEEE Transactions on Instrumentation and Measurement (2024).
14. J. Ai, Q. Zhang, and X. Su, "Fast modulation measurement profilometry based on phase-shifting and modulation ratio," Opt. Express **33**(10), 21707–21718 (2025).
15. Y. Wang, J. I. Laughner, I. R. Efimov, *et al.*, "3d absolute shape measurement of live rabbit hearts with a superfast two-frequency phase-shifting technique," Opt. Express **21**(5), 5822–5832 (2013).
16. S. Heist, M. Sieler, P. Kühmstedt, *et al.*, "High-speed 3d shape measurement using an array projector," in *Fringe 2013: 7th International Workshop on Advanced Optical Imaging and Metrology*, (Springer, 2014), pp. 835–838.
17. W. Guo, Z. Wu, Y. Li, *et al.*, "Real-time 3d shape measurement with dual-frequency composite grating and motion-induced error reduction," Opt. Express **28**(18), 26882–26897 (2020).
18. L. Yueyang, W. Zhoujie, and Z. Qican, "Phase error compensation technique based on phase-shifting fringe analysis: A review," Laser Optoelectron. Prog. **61**, 0211008 (2024).
19. W. Guo, Z. Wu, Q. Zhang, *et al.*, "Generalized phase shift deviation estimation method for accurate 3d shape measurement in phase-shifting profilometry," IEEE Transactions on Instrumentation and Measurement (2025).

Optics EXPRESS

20. T. Weise, B. Leibe, and L. Van Gool, "Fast 3d scanning with automatic motion compensation," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2007), pp. 1–8.
21. W. Guo, Z. Wu, Q. Zhang, *et al.*, "Real-time motion-induced error compensation for 4-step phase-shifting profilometry," Opt. Express **29**(15), 23822–23834 (2021).
22. S. Feng, C. Zuo, T. Tao, *et al.*, "Robust dynamic 3-d measurements with motion-compensated phase-shifting profilometry," Opt. Lasers Eng. **103**, 127–138 (2018).
23. J. Yu and S. Mai, "Quasi-pixelwise motion compensation for 4-step phase-shifting profilometry based on a phase error estimation," Opt. Express **30**(11), 19055–19068 (2022).
24. L. Lu, V. Suresh, Y. Zheng, *et al.*, "Motion induced error reduction methods for phase shifting profilometry: A review," Opt. Lasers Eng. **141**, 106573 (2021).
25. L. Lu, J. Xi, Y. Yu, *et al.*, "New approach to improve the accuracy of 3-d shape measurement of moving object using phase shifting profilometry," Opt. Express **21**(25), 30610–30622 (2013).
26. H. Wang, Y. Wang, Z. Chen, *et al.*, "Real-time motion-induced error reduction for phase-shifting profilometry with projection points tracking method," Measurement **239**, 115450 (2025).
27. Y. Li, M. Li, W. Chen, *et al.*, "mrpca-psp: Robust motion-resistant phase-shifting profilometry based on principal component analysis," Laser Photonics Rev. **19**(15), 2401938 (2025).
28. B. Wang, W. Chen, J. Qian, *et al.*, "Single-shot super-resolved fringe projection profilometry (sssr-fpp): 100,000 frames-per-second 3d imaging with deep learning," Light: Sci. Appl. **14**(1), 70 (2025).
29. T. Samavati and M. Soryani, "Deep learning-based 3d reconstruction: a survey," Artif. Intell. Rev. **56**(9), 9175–9219 (2023).
30. X.-F. Han, H. Laga, and M. Bennamoun, "Image-based 3d object reconstruction: State-of-the-art and trends in the deep learning era," IEEE Trans. Pattern Anal. Mach. Intell. **43**(5), 1578–1604 (2021).
31. A. Ioannidou, E. Chatzilari, S. Nikolopoulos, *et al.*, "Deep learning advances in computer vision with 3d data: A survey," ACM Comput. Surv. **50**(2), 1–38 (2018).
32. W. Yin, Y. Che, X. Li, *et al.*, "Physics-informed deep learning for fringe pattern analysis," Opto-Electron. Adv. **7**(1), 230034 (2024).
33. Y. Li, W. Guo, J. Shen, *et al.*, "Motion-induced phase error compensation using three-stream neural networks," Appl. Sci. **12**(16), 8114 (2022).
34. J. Tan, W. Su, Z. He, *et al.*, "Deep learning-based method for non-uniform motion-induced error reduction in dynamic microscopic 3d shape measurement," Opt. Express **30**(14), 24245–24260 (2022).
35. W. Jiang, E. Trulls, J. Hosang, *et al.*, "Cotr: Correspondence transformer for matching across images," in *Proceedings of the IEEE/CVF international conference on computer vision*, (2021), pp. 6207–6217.
36. T. Senst, V. Eiselein, and T. Sikora, "Robust local optical flow for feature tracking," IEEE Trans. Circuits Syst. Video Technol. **22**(9), 1377–1387 (2012).
37. W. Yin, C. Zuo, S. Feng, *et al.*, "High-speed three-dimensional shape measurement using geometry-constraint-based number-theoretical phase unwrapping," Opt. Lasers Eng. **115**, 21–31 (2019).
38. Z. Wu, W. Guo, and Q. Zhang, "High-speed three-dimensional shape measurement based on shifting gray-code light," Opt. Express **27**(16), 22631–22644 (2019).
39. Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *European conference on computer vision*, (Springer, 2020), pp. 402–419.
40. Z. Huang, X. Shi, C. Zhang, *et al.*, "Flowformer: A transformer architecture for optical flow," in *European conference on computer vision*, (Springer, 2022), pp. 668–685.
41. H. Xu, J. Zhang, J. Cai, *et al.*, "Gmflow: Learning optical flow via global matching," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (2022), pp. 8121–8130.