

Article

Motion-Induced Phase Error Compensation Using Three-Stream Neural Networks

Yueyang Li, Wenbo Guo, Junfei Shen, Zhoujie Wu  and Qican Zhang * 

College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

* Correspondence: zqc@scu.edu.cn

Abstract: Phase-shifting profilometry (PSP) has been widely used in the measurement of dynamic scenes. However, the object motion will cause a periodical motion-induced error in the phase map, and there is still a challenge to eliminate it. In this paper, we propose a method based on three-stream neural networks to reduce the motion-induced error, while a general dataset establishment method for dynamic scenes is presented to complete three-dimensional (3D) shape measurement in a virtual fringe projection system. The numerous automatically generated data with various motion types is employed to optimize models. Three-step phase-shift fringe patterns captured along a time axis are divided into three groups and processed by trained three-stream neural networks to produce an accurate phase map. The actual experiment's results demonstrate that the proposed method can significantly perform motion-induced error compensation and achieve about 90% improvement compared with the traditional three-step phase-shifting algorithm. Benefiting from the robust learning-based technique and convenient digital simulation, our method does not require empirical parameters or complex data collection, which are promising for high-speed 3D measurement.

Keywords: fringe projection profilometry; motion-induced error; deep learning



Citation: Li, Y.; Guo, W.; Shen, J.; Wu, Z.; Zhang, Q. Motion-Induced Phase Error Compensation Using Three-Stream Neural Networks. *Appl. Sci.* **2022**, *12*, 8114. <https://doi.org/10.3390/app12168114>

Academic Editors: Andrés Márquez and Edik U. Rafailov

Received: 25 May 2022

Accepted: 10 August 2022

Published: 13 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Three-dimensional (3D) measurement methods based on fringe projection profilometry (FPP) [1–3] has been widely used in computer vision [4], industrial inspection [5], and other fields [6] due to its advantages of non-contact, low cost, and high accuracy. A typical FPP system consists of a camera and a projector. By projecting specially designed fringe patterns onto a tested object, the depth information of the object's surface is modulated in the distorted fringe patterns. Then, the phase distribution can be retrieved through fringe analysis algorithms and converted to 3D geometry based on triangulation.

At present, there are two main approaches to extract phase value in FPP: Fourier transform profilometry (FTP) [7,8] and phase shifting profilometry (PSP) [9,10]. In FTP, only a single-frame high-frequency fringe pattern is required, and a suitable designed bandpass filter is applied to separate fundamental components in the frequency domain. However, with the limitation of filter operation, FTP is sensitive to surface variation and non-uniform reflection. In PSP, multi-frame (usually at least three frames) fringe patterns are utilized to calculate the phase map by a least square algorithm. At the expense of reducing speed, a more accurate and robust result can be obtained compared with FTP.

For the measurement of dynamic scenes, the moving object will no longer satisfy the assumptions of PSP, which include fixed position and known phase shift. More specifically, the location of a moving object changes over time between frames, while an extra value is introduced into phase shift on the same point of the object's surface [11]. The motion-induced error will cause periodical fluctuations in phase distribution, leading to the decrease of measurement accuracy.

To suppress the motion-induced error in PSP, researchers have developed various kinds of methods. One of the approaches is to combine the FTP with PSP. Cong et al. [12]

extracted phase map using FTP from single fringe pattern in three-step PSP, and performed phase subtraction to estimate the unknown phase shift. Qian et al. [13] used four fringe patterns to obtain the absolute phase with the stereo phase unwrapping (SPU) method, then developed a pixel-wised motion detection strategy to fuse the results of PSP and FTP. Guo et al. [14] proposed a dual-frequency composite grating method to identify the motion region using the phase of a virtual high frequency and replaced the phase map of PSP with FTP in corresponding region. However, due to the implicit drawbacks of FTP, these hybrid methods are unadaptable for dealing with complex scenes.

To avoid this issue, Lu et al. tracked the motion of an object by manually placed makers [15] or scale-invariant feature transform (SIFT) [16] and estimated the translation vector and rotation matrix to calculate phase map. Liu et al. [17] estimated the phase-shift error by averaging the difference of three adjacent phase maps, which are calculated from eight consecutive images of four-step phase-shift fringe patterns. Wang et al. [18] applied Hilbert transform to shift the phase of three-step phase-shift fringe patterns by $\pi/2$, constructed opposite distributional phase to compensate motion error. Guo et al. [19] divided four-step phase-shift fringe patterns into two groups to calculate two phase maps which have opposite phase error distribution, and the periodical motion-induced phase error can be compensated by averaging them. However, most of these methods may not be suitable if the object undergoes non-uniform motion [11].

Recently, plenty of studies have introduced deep learning technique into 3D measurement. Yu et al. [20] proposed a deep learning-based modulation-enhancing method that transforms two low-modulation fringe patterns into a set of three-step phase-shift fringe patterns. Zhang et al. [21] extracted accurate phase information from three-step phase-shift fringe patterns of low signal-to-noise ratio (SNR) and saturation by using convolutional neural network (CNN). Feng et al. [22] utilized U-Net [23] architecture to suppress phase error in non-sinusoidal patterns resulting from gamma distortion, defocus, saturation, and couples of these factors. Nguyen et al. [24] extracted multiple triple-frequency phase-shift grayscale fringes from single color fringe pattern and reconstructed 3D shape accurately.

As discussed before, the FTP-based methods will lead to unreliable results in complex variant surface, and PSP-based methods are usually limited in resolving complex objects with different types of motion. In order to solve these problems, we intend to reduce the motion-induced phase error in PSP using deep learning technique. Firstly, the phase error model of three-step PSP is derived to understand the characteristics of motion-induced error. Secondly, a method based on three-stream neural networks is proposed to process three different orders of the three-step phase-shift fringe patterns in the time series, and effectively suppress the phase error. Thirdly, a virtual FPP system is constructed to perform the measurement of dynamic scenes and provides the synthetic data for network training, which shows the potential of our method in dynamic scene analysis. Experimental results prove that the method can accurately reconstruct a moving object with different motion types and achieve significant improvement compared with traditional three-step phase-shifting algorithm.

2. Principle

2.1. Motion-Induced Error in Three-Step Phase-Shifting Algorithm

The intensity distribution of N -step phase algorithm can be expressed as:

$$I_n(x, y) = A(x, y) + B(x, y) \cos[\phi(x, y) - \delta_n(x, y)], \quad (1)$$

$$\delta_n(x, y) = 2\pi(n - 1)/N, n = 1, 2, \dots, N. \quad (2)$$

where x, y donates the pixel coordinate, A, B, ϕ represents the background intensity, intensity modulation, and phase map, respectively. δ_n is the theoretical phase shift, and n is the phase shift number. The desired phase map can be calculated by least square algorithm:

$$\phi(x, y) = \arctan \frac{M(x, y)}{D(x, y)} = \arctan \frac{\sum_{n=1}^N I_n(x, y) \sin \delta_n(x, y)}{\sum_{n=1}^N I_n(x, y) \cos \delta_n(x, y)}. \tag{3}$$

where $M(x, y)$ and $D(x, y)$ represent the numerator and denominator of the arctangent function, respectively. When the measured object is moving, the actual phase shift δ'_n becomes:

$$\delta'_n(x, y) = \delta_n(x, y) + \varepsilon_n(x, y). \tag{4}$$

where $\varepsilon_n(x, y)$ is the additional unknown phase shift caused by motion in n th fringe pattern. Similarly, the actual phase map ϕ' can be calculated by:

$$\phi'(x, y) = \arctan \frac{\sum_{n=1}^N I_n(x, y) \sin \delta'_n(x, y)}{\sum_{n=1}^N I_n(x, y) \cos \delta'_n(x, y)}. \tag{5}$$

Therefore, the motion-induced error can be expressed as [25]:

$$\begin{aligned} \Delta\phi &= \phi' - \phi \\ &= \arctan \frac{\cos 2\phi \cdot \sum_{n=1}^N \sin(2\delta_n + \varepsilon_n) - \sin 2\phi \cdot \sum_{n=1}^N \cos(2\delta_n + \varepsilon_n) - \sum_{n=1}^N \sin \varepsilon_n}{\cos 2\phi \cdot \sum_{n=1}^N \cos(2\delta_n + \varepsilon_n) + \sin 2\phi \cdot \sum_{n=1}^N \sin(2\delta_n + \varepsilon_n) + \sum_{n=1}^N \cos \varepsilon_n}. \end{aligned} \tag{6}$$

Equation (6) shows that the distribution of motion-induced error is related to the doubled frequency of the projected fringe. In this paper, three-step PSP is selected for analysis and measurement, since it requires minimum number of patterns with ensured accuracy. Three-step fringe patterns distorted by motion can be represent as:

$$\begin{aligned} I_1(x, y) &= A(x, y) + B(x, y) \cos[\phi(x, y) - \varepsilon_1], \\ I_2(x, y) &= A(x, y) + B(x, y) \cos[\phi(x, y) - (2\pi/3 + \varepsilon_1 + \varepsilon_2)], \\ I_3(x, y) &= A(x, y) + B(x, y) \cos[\phi(x, y) - (4\pi/3 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3)]. \end{aligned} \tag{7}$$

where $\varepsilon_1 = 0$. Considering a small phase shift error ε , $\sin(\varepsilon) \approx \varepsilon$, $\cos(\varepsilon) \approx 1$, the actual phase of three-step PSP can be derived from Equation (5):

$$\phi' \approx \arctan \frac{(6 + \sqrt{3}\varepsilon_3) \sin \phi - (6\varepsilon_2 + 3\varepsilon_3) \cos \phi}{(2\varepsilon_2 + \varepsilon_3) \sin \phi + (6 - \sqrt{3}\varepsilon_3) \cos \phi}. \tag{8}$$

The corresponding motion-induced error can be expressed as:

$$\begin{aligned} \Delta\phi &\approx \arctan \frac{2\sqrt{3}\varepsilon_3 \sin 2\phi - (4\varepsilon_2 + 2\varepsilon_3) \cos 2\phi - (8\varepsilon_2 + 4\varepsilon_3)}{-(4\varepsilon_2 + 2\varepsilon_3) \sin 2\phi - 2\sqrt{3}\varepsilon_3 \cos 2\phi + 12 - 2\sqrt{3}\varepsilon_3} \\ &\approx \frac{\sqrt{3}\varepsilon_3 \sin 2\phi - (2\varepsilon_2 + \varepsilon_3) \cos 2\phi}{6} - \frac{2\varepsilon_2 + \varepsilon_3}{3}. \end{aligned} \tag{9}$$

It can be seen that the phase error is periodically distributed and is a function of 2ϕ . A spectral-filter-based or iteration-based method can be adopted to eliminate the motion-induced error. However, these methods usually decrease the ability of handling complex surfaces, or have prior assumptions on the movement forms, which limits their application in different types of objects.

2.2. Three-Stream Neural Networks-Based Motion-Induced Error Compensation Method

As discussed before, it is still problematic to accurately retrieve the phase from phase-shift fringe patterns of moving objects, especially for those with complex surfaces and different types of motion. To address this issue, deep learning technique is introduced to suppress the periodical phase error caused by motion. The diagram of the proposed temporal three-stream neural networks-based method is shown in Figure 1. Firstly, image sequences of three-step fringe patterns, which contain three different kinds of phase shift orders (order1 = (0, 2π/3, 4π/3), order2 = (2π/3, 4π/3, 0), and order3 = (4π/3, 0, 2π/3)) are cyclically projected onto the object over time. Secondly, every three adjacent images of the captured fringe patterns modulated by the object are passed into CNN 1~3 according to their orders. Thirdly, the numerator $M(x,y)$ and denominator $D(x,y)$ of the arctangent function in Equation (3) are predicted for accurate phase calculation and phase unwrapping. Lastly, the 3D shape is reconstructed using the traditional triangulation method.

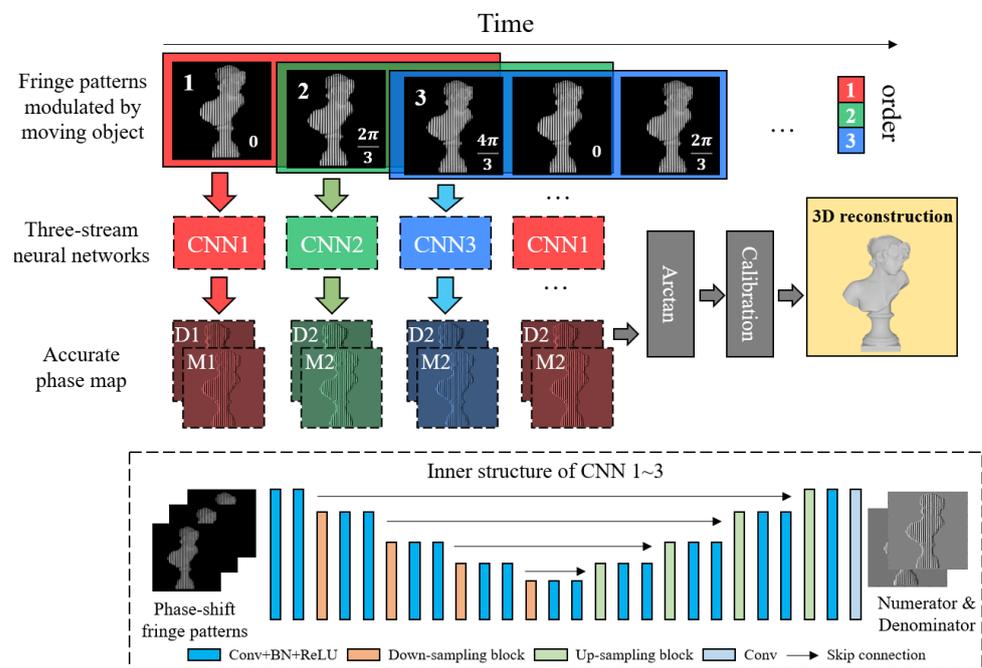


Figure 1. Proposed temporal three-stream neural networks-based method for motion-induced error compensation (top) and the inner structure of CNN1~3 (bottom).

The design principles of the proposed framework are follows: during the projection of fringe patterns, the phase information and the position of the object change continuously from frame to frame, which make it difficult to directly learn the features from image sequence using a single neural network. Therefore, these fringe patterns are artificially divided into three groups according to their orders of phase shift values and passed to different subnets of three-stream neural networks. Since the inputs of each subnet have similar distribution in spatial dimension after grouping operation, it will be easier for networks to focus on the elimination of motion-induced error.

All the CNN 1~3 adopt the architecture of U-Net [23], whose effectivity in phase prediction of non-sinusoidal fringe patterns has been verified [22]. The detail of the network is shown in bottom of Figure 1. The whole structure consists of an encoder, decoder, and skip connection. For the encoder, the input images are firstly processed by two convolutional blocks, which is implemented by a combination of a convolutional layer (Conv), batch normalization layer (BN) [26], and linear rectification function (ReLU) [27]. The max pooling followed by two convolution blocks are applied to down-sample the tensors by 1/2 in width and height, and the channel dimension is doubled. The same operations are used four times to increase the receptive field gradually. For the symmetrical

decoder, the down-sampling operations are replaced by up-sampling. The multi-level features from skip connection are concatenated with tensors of doubled resolution to enhance the detail information. The output layer only has one single convolution operation without ReLU, since both positive and negative terms existed in $M(x,y)$ and $D(x,y)$. The kernel of all convolutional layer is 3×3 with padding 1. The start features in the input layer is 64. Taking the fringe patterns modulated by moving object as input, the CNN 1~3 in three-stream neural networks will output the $M(x,y)$ and $D(x,y)$, whose motion-induced error has been suppressed.

A simple L2-norm function is used to optimize the neural network and can be written as:

$$Loss(\theta) = [M_p(\theta) - M_g]^2 + [D_p(\theta) - D_g]^2. \quad (10)$$

where θ is the parameter of the network trained to minimize the loss function. The subscript p and g of M and D represent the prediction and ground truth, respectively.

2.3. Dataset Establishment in Virtual FPP System

Deep learning-based FPP methods usually require a dataset that consists of numerous fringe patterns modulated by different objects for high accuracy network prediction. However, the dataset collection in the real world will be limited in the hardware, number of objects, and other uncontrolled factors. To solve this problem, a dataset establishment flow in a virtual FPP system for dynamic scenes is proposed to produce synthetic data conveniently and flexibly.

The availability of large image datasets has become a major bottleneck in deep learning-based techniques. Recently, synthetic dataset generation methods using computer graphics show their potential applications in various industrial use-cases, such as viewpoint estimation [28], image classification [29], single-shot FPP [30,31], etc. In this section, our goal is to establish one-to-one mapping between the actual and virtual FPP system, and the method mentioned in [30] is extended from static scene to dynamic scene. A free and open source 3D creation suite named Blender [32] is introduced to build virtual FPP system and training dataset. Meanwhile, various 3D models including sculpture, toy, and industrial component in Thingi 10K dataset [33] are selected as the objects to be measured.

The first step is to determine the positional relationship between camera and projector and their own inherent parameters, i.e., intrinsic and extrinsic matrixes. The experimental setup of FPP system and the schematic diagram of virtual FPP system are shown in Figure 2a,b, respectively. The actual experimental setup shown in Figure 2a provides the referenced structural parameters for the virtual one shown in Figure 2b and will be used in the performance evaluation of this work in Section 3. According to the traditional pinhole camera model, the image capture process from a point in the world coordinate to a pixel of the image plane can be described as:

$$\lambda^c \begin{bmatrix} u^c \\ v^c \\ 1 \end{bmatrix} = K^c [R^c | T^c] \begin{bmatrix} X^w \\ Y^w \\ Z^w \\ 1 \end{bmatrix}. \quad (11)$$

where the superscript c represents the camera imaging system. λ is a scaling factor. (u,v) is the pixel coordinate on the image plane. K is the intrinsic matrix. R and T denote the rotation matrix and translation vector from the world coordinate (X_w, Y_w, Z_w) to camera coordinate in extrinsic matrix, respectively. K , R , and T can be further rewritten as:

$$K^c = \begin{bmatrix} f_v^c & s^c & u_0^c \\ 0 & f_v^c & v_0^c \\ 0 & 0 & 1 \end{bmatrix}, \quad (12)$$

$$R^c = \begin{bmatrix} r_{11}^c & r_{12}^c & r_{13}^c \\ r_{21}^c & r_{22}^c & r_{23}^c \\ r_{31}^c & r_{32}^c & r_{33}^c \end{bmatrix}, T^c = \begin{bmatrix} t_1^c \\ t_2^c \\ t_3^c \end{bmatrix}. \tag{13}$$

f_u, f_v are the focal length along u and v directions. s represents the axis skew. (u_0, v_0) is the location of the principle point. R_{ij} and t_i are the parameters in the corresponding location of R and T , respectively. Since the projector can be regarded as an inverse camera, the image projection process can be written as:

$$\lambda^p \begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} = K^p [R^p | T^p] \begin{bmatrix} X^w \\ Y^w \\ Z^w \\ 1 \end{bmatrix}. \tag{14}$$

where the superscript p represents the projector imaging system. Assume that the projector coordinate coincides with world coordinate, and its origin is $(0,0,0)^T$, the mapping from the origin of the camera coordinate to that of the projector can be obtained:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = R^c \begin{bmatrix} X_0^c \\ Y_0^c \\ Z_0^c \end{bmatrix} + T^c. \tag{15}$$

The camera location can be determined by a rearranging of Equation (15):

$$\begin{bmatrix} X_0^c \\ Y_0^c \\ Z_0^c \end{bmatrix} = -(R^c)^T T^c. \tag{16}$$

All parameters in the intrinsic and extrinsic matrixes of the camera and projector can be calibrated by Zhang’s [34] and Li’s method [35]. Since the Euler angles along the x, y, z axes are required in Blender to describe the object’s rotation, the calibrated rotation matrix should be considered as the multiplication of three parts:

$$R = R_z(\alpha)R_y(\beta)R_x(\gamma). \tag{17}$$

where the Euler angles α, β, γ can be solved by Slabaugh’s method [36]. So far, the one-to-one relationship between the virtual and actual FPP system has been built.

The next step is to determine the location and motion type of objects and perform image rendering. As shown in Figure 2c, the most common rigid motion type of a measured object can be categorized into translation, rotation, and mixture of them, while the object remains static without any external force. To simulate a dynamic scene realistically, a dataset generation pipeline is proposed:

- (a) to load a 3D model as the measured object and randomly rescale it to 1/2~2/3 of the calibration volume. The coordinate of the object’s centroid represents its position.
- (b) to randomly select the initial position of the object at first frame and final position at F th frame. F is the number of frames the object motion. For the stationary and rotation around the centroid, the initial position equals the final one. The selected position should make the bounding box of the object inside the calibration volume.
- (c) to choose the corresponding interframe interpolation methods according to the desired motion trajectory. For example, linear interpolation corresponds to uniform linear motion, and the nodes of Bezier interpolation can be controlled to implement different accelerative motion.
- (d) to cyclically project three-step phase-shift fringe patterns in F frames and synchronously render the corresponding deformed patterns. The object poses (location and rotation) are simultaneously recorded at every frame. One phase map affected by motion is

calculated from three adjacent images, so $F-2$ phase maps can be obtained from the total F frames.

- (e) for each pose recorded in (d), to project standard 12-step phase-shift patterns and calculate the phase maps of $F-2$ poses as the corresponding ground truth for the phase maps obtained in (d).
- (f) to remove the object and return to (a) for the next object.

After processing all objects with different motion types, the obtained synthetic dataset can be used to train the model mentioned in Section 2.2. Instead of complex manually marking, the data collection using virtual FPP system and computer graphics can be performed automatically. In addition to data generation, the proposed method can also be used for the analysis of non-rigid movement or other unstable measurement scenes.

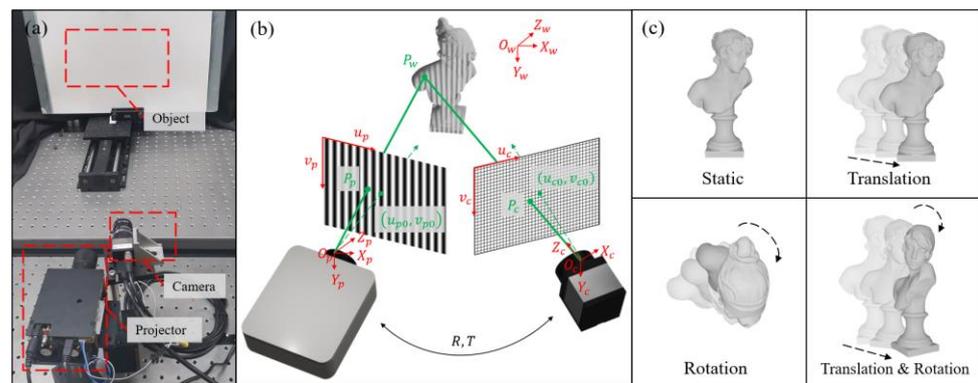


Figure 2. Dataset establishment in virtual FPP system. (a) Experimental setup of actual FPP system. (b) Schematic diagram of virtual FPP system. (c) Different motion types used in synthetic data generation.

3. Experiment

3.1. Data Acquisition

As shown in Figure 2a, the actual experimental setup of FPP system contained a camera (model: UI-3250CP-M-GL R2) with a resolution of 1600×1200 pixels and a digital light processing (DLP) projector with a resolution of 1280×800 pixels and throw ratio of 1.6. The objective lens of the camera had a focal length of 12 mm. The period number of the projected fringe patterns is 64.

Using the data establishment method mentioned in Section 2.3, we have built a virtual FPP system as a one-to-one mapping from the aforementioned actual system. The parameters of camera and projector in the virtual FPP system were identified by calibration. The mean reprojection error of both the calibrated camera and projector are 0.07 pixels, which are accurate enough for dataset establishment. It is worth noting that the spot light (projector) in the Blender has no concept of focal length, and the projected image is always in focus when the shadow soft size is 0 mm (ideal point light source). For simplicity, the intrinsic matrix of the projector was replaced by its throw ratio to adjust the size of projected images. Seven types of motion were selected to generate synthetic data: stationary, uniform translation, accelerated translation, uniform rotation, accelerated rotation, a mixture of uniform translation and rotation, and a mixture of accelerated translation and rotation. One hundred objects and 24 (F) postures for each have been collected. To reduce memory space and speed up network optimization, the resolution of the rendered images was set to 0.4; that is, the camera resolution was 640×480 pixels in this virtual FPP system. In total, $100 \times 7 \times (24 + 22 \times 12) = 201,600$ fringe patterns were rendered and $100 \times 7 \times 22 = 15,400$ pairs of input and the ground truth phase maps were obtained to make up the dataset, where 22 represents the number of phase maps calculated from 24 fringe patterns, and 12 represents the 12-step phase shift. All the steps in the dataset generation pipeline were implemented by Python script and performed automatically. It

takes about one day to generate the dataset by processing different motion type in parallel. The dataset was split into two parts, 90% training set and 10% validation set. The collection of the training set and validation set are shown in Figure 3. Several image preprocessing operations, including normalization, randomly rotating to a degree $\pm 15^\circ$, cropping into a size of 480×480 pixels, were adopted to further increase the diversity of data during training stage.

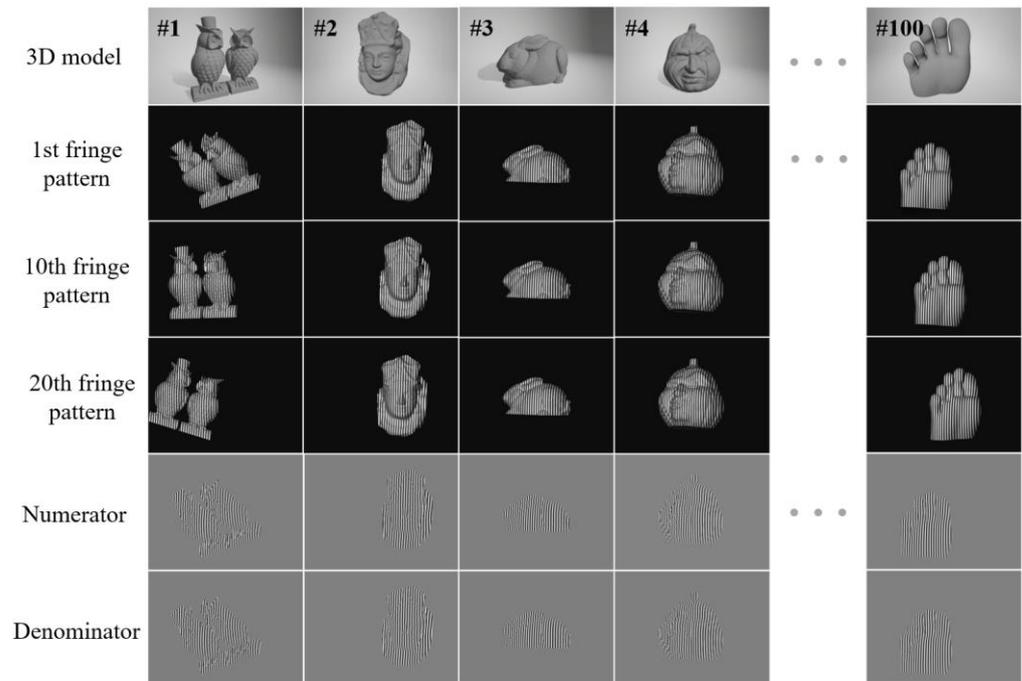


Figure 3. Collection of the training data and validation data. The first row shows the different 3D models used in the virtual FPP system. Each 3D model with different motion type was measured by 24 frames fringe patterns of the three-step PSP, in which the 1st, 10th, and 20th images are shown in the second to fourth row. The ground truth numerator and denominator of the first frame are shown in the fifth and sixth row.

3.2. Network Training

The network was implemented by Pytorch [37] 1.8.2 and the training process was completed in the hardware environment with Intel Gold 5120 CPU and the 16 GB NVIDIA Tesla P100 GPU. We used an Adam [38] optimizer with initial learning rate of 10^{-4} to minimize the loss function in Equation (10). The batch size is set to 4. A scheduler was applied to reduce the learning rate with a factor of 0.1 when there is no improvement on validation set after 10 epochs. The temporal three-streams neural networks are trained separately, and their loss curves are shown in Figure 4. Both curves show the convergence occur after around 30 epochs, then the losses steadily reduce to about 5 in the training set and 8 in validation set, respectively, which indicates the validity of the model.

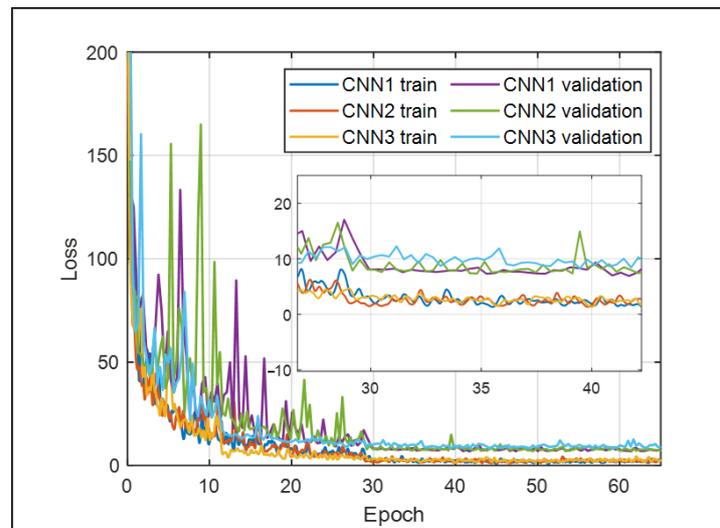


Figure 4. Loss curves of the training stage.

3.3. Quantitative Evaluation

To verify the performance of the proposed method, we first evaluated the trained model in the validation set, which is generated by a virtual FPP system. In Figure 5, the first shaped object in the first row fronted towards the projector and translated from right to left. One of the captured fringe patterns is shown in Figure 5a. Since the poses in every frame were recorded and fixed for generating the ground truth, the phase error compared with 12-step phase-shift can be calculated conveniently. The phase errors of the traditional three-step phase-shifting algorithm and the proposed method are listed in the left and right of Figure 5b. The error distributions along the marked red dotted line are shown in Figure 5c. Figure 5d–f show the corresponding results of two owl sculptures with more complex details and faster movement speed. Figure 5g–i show the results of a hat-shaped object rotated along x axis. Figure 5j–l show this for two industrial components rotated along the y axis. Figure 5m–r are the results of a Buddhist sculpture and two owl sculptures both translated and rotated around their centroids.

It can be seen from Figure 5 that the phase errors of the three-step PSP show a double frequency distribution and have an offset compared with their own ground truth, which is consistent with the Equation (9). The RMS of phase errors of the proposed method are reduced by 75%–93% compared with that of the traditional three-step PSP. The results show that the proposed method can successfully suppress the motion-induced error caused by different motion types.

Then, we tested the trained model in an actual FPP system shown in Figure 2a to evaluate the performance of the presented model in a real scene. In Figure 6, a standard ceramic plate shown in Figure 6a placed on a motorized linear stage translated in-depth direction. The movement speed of the translation stage was 1 mm/step; one fringe pattern was projected for each step. The camera was triggered synchronously by the projector and captured a deformed fringe pattern for each step. To quantitatively analyze the influence of motion-induced error, the ground truth was obtained by projecting 12-step fringe patterns at each position of the stage and shown in Figure 6b. The three-step fringe patterns modulated by moving plate were generated by selecting the first, fifth, and ninth images of the 12-step PSP. The 3D shape was reconstructed using the multiple-frequency phase unwrapping method [39] and triangular stereo calibration model [35]. As shown in Figure 6c,d, the result of the three-step PSP has obvious periodical distribution in surface, while the proposed method significantly reduces the fluctuations and the bias (shown in Figure 6e) caused by direct-current (DC) component in Equation (9). The root mean square (RMS) error in height (Z) dimension of the three-step PSP and the proposed method are 1.07 mm and 0.12 mm, respectively. The motion-induced error has been decreased about 90%.

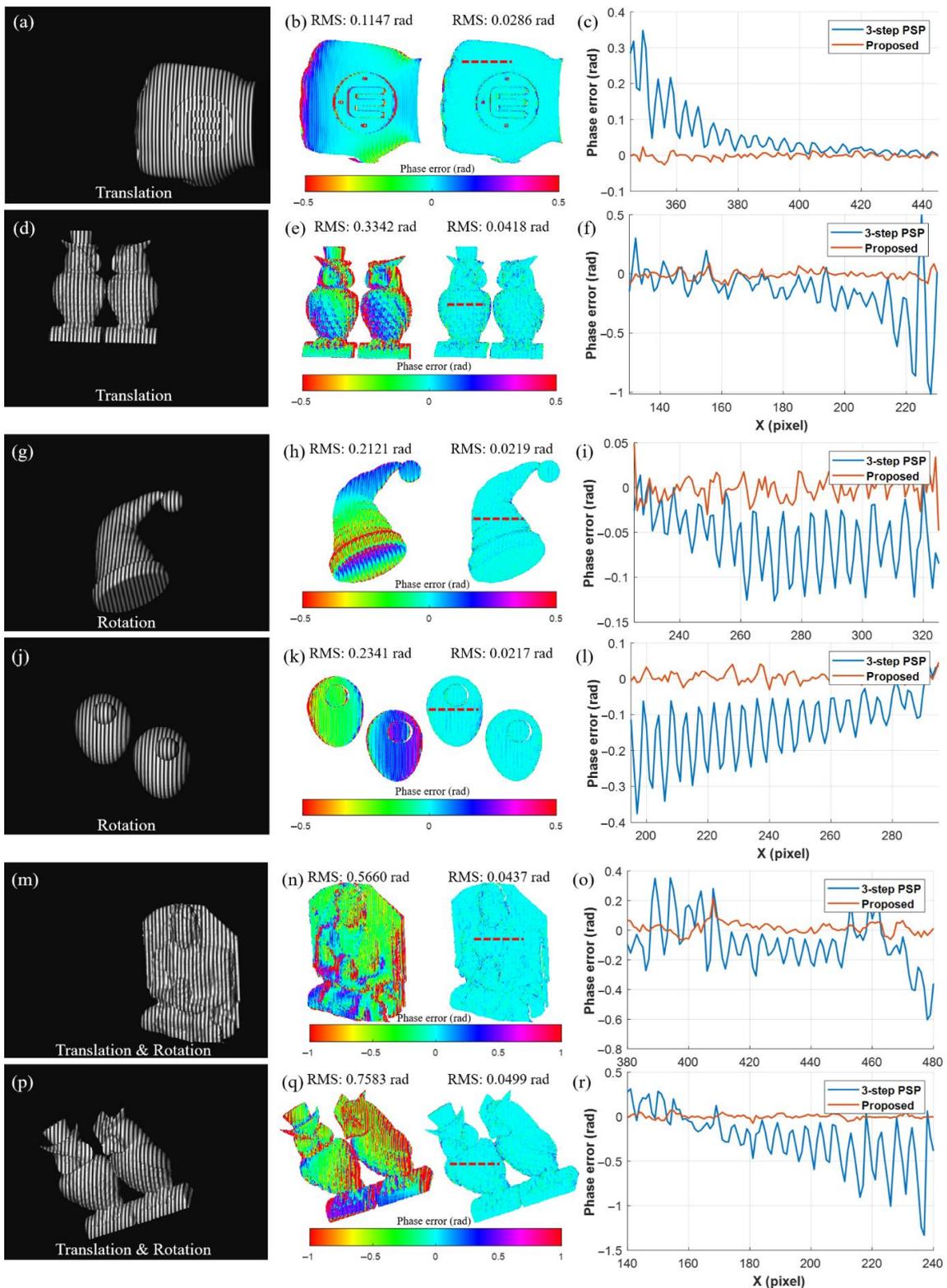


Figure 5. Measurement results of the scenes of different motion types in the validation set. (a) One fringe pattern of a translational fist-shaped object. (b) Phase error of three-step PSP (left) and proposed method (right), and their RMS errors (top). (c) Section line of phase errors along the red dotted line marked in (b). (d–r) Corresponding results of translational two owl sculptures, one rotating hat-shaped object, two rotating components, one translational and rotating Buddhist sculpture, and owl sculptures.

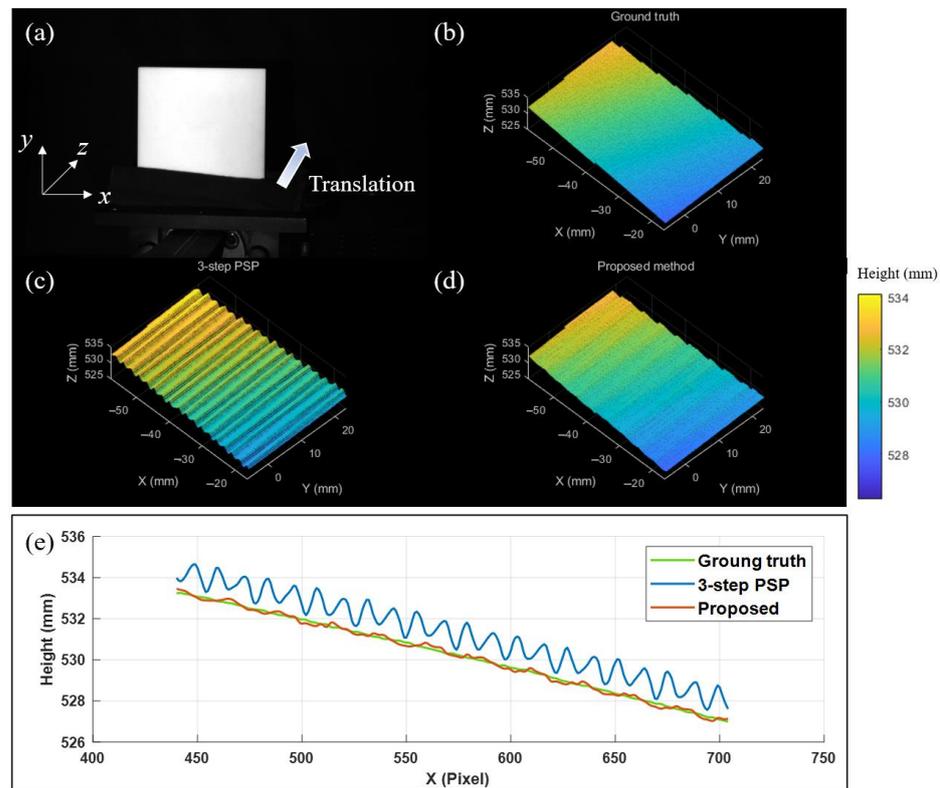


Figure 6. Measurement results of a translational scene in actual FPP system. (a) Standard ceramic plate translated in depth direction to be tested. (b–d) Reconstructed results of 12-step PSP, three-step PSP and proposed method, respectively. (e) Section line in 600th row of the height (Z) distributions of three methods.

Two standard spheres were measured under the same condition with ceramic plate. The diameters of two standard spheres are 50.7991 mm and 50.7970 mm, respectively, and the center distance is 100.2537 mm, as shown in Figure 7a. One of the deformed fringe patterns is shown in Figure 7b. Figure 7c,d illustrate the reconstructed results of traditional three-step PSP and our method, respectively. It is shown that the largest STD error of the two reconstructed spheres by two methods are 0.4001 mm and 0.1254 mm and the largest diameter error of two methods are 2.6746 mm and 0.5486 mm. The results demonstrate that the proposed method can efficiently reduce the motion-induced error.

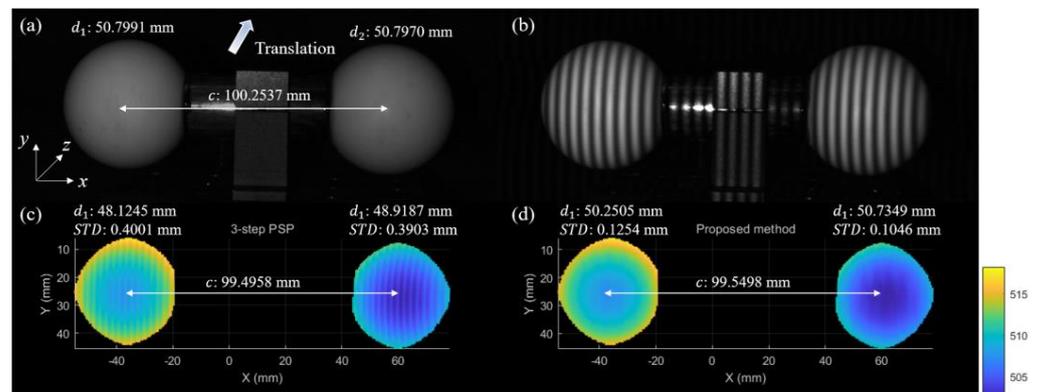


Figure 7. Measurement results of two standard spheres in actual FPP system. Symbol d represents diameter, c represents center distance. (a) Standard spheres translated in-depth direction to be tested. (b) One of deformed fringe patterns. (c,d) Reconstructed results of three-step PSP and proposed method, respectively.

In Figure 8, we measured a rotating sculpture with a speed of about 0.1 rad per frame. The experimental conditions are the same with the ceramic plate, except that the motorized linear stage was replaced with a rotation stage. The rotating sculpture and one of the deformed fringe patterns are shown in Figure 8a,c, respectively. It can be seen from Figure 8d,g that the three-step method has obvious periodical waves at nose area. The results in Figure 8e,h demonstrate that the proposed method can also deal with the motion-induced error caused by such rotational motion. Figure 8i exhibits the profile distribution of three methods along the same line.

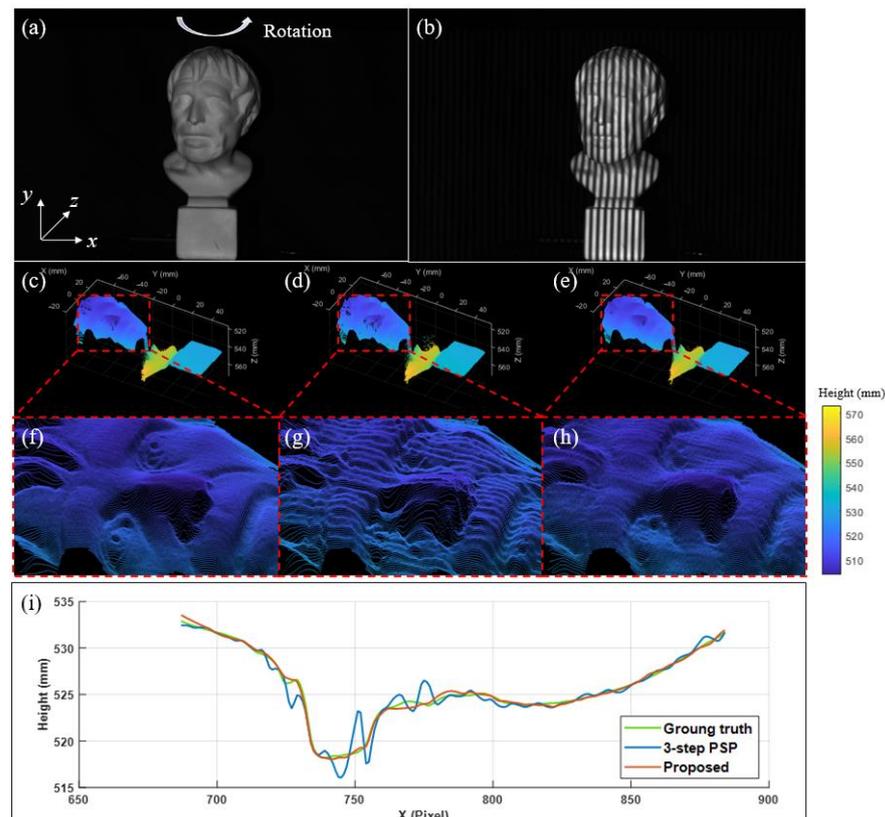


Figure 8. Measurement results of a rotating scene in an actual FPP system. (a) Rotating sculpture. (b) One of the deformed fringe patterns. (c–e) Reconstructed results of 12-step PSP, 3-step PSP, and proposed method, respectively. (f–h) Local enlarged results corresponding to (c–e). (i) Section line in 400th row of the height (Z) distributions of three methods.

4. Conclusions and Discussion

In this paper, we present a motion-induced error compensation method based on deep learning and computer graphics. By building the virtual system referring to an actual FPP system and performing image rendering, we can simulate the measurement process of a dynamic scene realistically and provide sufficient data for network training. Then, the proposed three-stream neural networks are trained using synthetic data and process three different orders of three-step fringe patterns in time series. The experimental results demonstrate that the motion-induced error introduced by various motion types can be reduced effectively compared with the traditional three-step PSP.

Compared with existing methods, the proposed method has several improvements in accuracy, ability of dealing with non-uniform motion and efficiency. Compared with FTP assisted methods, the well-trained neural networks can avoid artificial filter design, which enable our method to provide more detailed and lossless results in dealing with complex objects. Compared with motion prediction methods, the proposed dataset generation method can realize the accurate simulation of non-uniform motion and has the advantage

of dealing this type of motion. Moreover, there are no additional images except basic three-step PSP fringe patterns needed during the experiment, and each high frequency deformed fringe patterns can reconstruct a new 3D result, which improves the efficiency of 3D reconstruction.

However, there are several aspects that need to be further improved in the future investigation. Firstly, the simulated environment is confined in a laboratory condition, that is, the background behind the measured object is black, and rendered images should not be overexposed, which may cause a limited performance of the model in practical application. Moreover, there are three main noise sources that remain to be researched, including capturing noise of the camera and light of the source, motion noise that related to the camera and projector, and rendering noise in a virtual FPP system. Therefore, a more general dataset establishment method should be considered to adapt different situation. Secondly, due to the difficulty of modeling the complex 3D rigid motion (for example, irregular shake, random walk) and non-rigid motion, and the limited generalization capabilities of designed three-streams neural networks, our method might not perform an effective error compensation in these cases. Thirdly, with the decrease of projection–acquisition rate relative to motion, the phase error of the proposed method will inevitably increase, as shown in Figure 9. According to our simulation, to achieve a high accuracy measurement (phase error $< 0.01\pi$), the projection–acquisition rate should be at least about 30 fps when the movement speed of object is 50 mm/s. Lastly, the binary defocus techniques [40] are usually used to improve the measurement speed, and it will lead to a hybrid phase error caused by defocusing and motion, which are not considered in our work. So, how to design a model to balance both of mentioned factors is the key point of high-speed measurement.

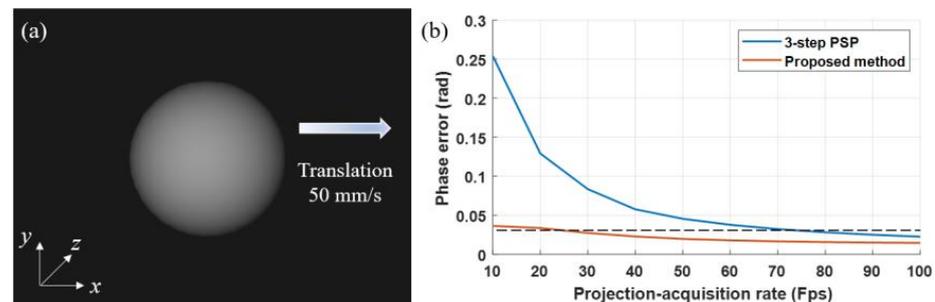


Figure 9. (a) Moving sphere with a speed of 50 mm/s in simulation scene. (b) The relationship between the projection–acquisition rate and phase error of the sphere. The dotted line represents the phase error of 0.01π .

Author Contributions: Conceptualization, Y.L., W.G., J.S., Z.W. and Q.Z.; methodology, Y.L.; software, Y.L.; validation, Y.L.; formal analysis, Y.L. and W.G.; investigation, Y.L. and W.G.; resources, Y.L.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L., W.G., J.S., Z.W. and Q.Z.; visualization, Y.L.; supervision, J.S. and Q.Z.; project administration, Q.Z.; funding acquisition, J.S. and Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 62075143 and 62105227).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gorthi, S.S.; Rastogi, P. Fringe Projection Techniques: Whither We Are? *Opt. Lasers Eng.* **2010**, *48*, 133–140. [[CrossRef](#)]
2. Wu, Z.; Guo, W.; Zhang, Q. Two-Frequency Phase-Shifting Method vs. Gray-Coded-Based Method in Dynamic Fringe Projection Profilometry: A Comparative Review. *Opt. Lasers Eng.* **2022**, *153*, 106995. [[CrossRef](#)]
3. Wu, Z.; Guo, W.; Li, Y.; Liu, Y.; Zhang, Q. High-Speed and High-Efficiency Three-Dimensional Shape Measurement Based on Gray-Coded Light. *Photon. Res.* **2020**, *8*, 819. [[CrossRef](#)]
4. Xu, J.; Zhang, S. Status, Challenges, and Future Perspectives of Fringe Projection Profilometry. *Opt. Lasers Eng.* **2020**, *135*, 106193. [[CrossRef](#)]
5. Harding, K. Engineering Precision. *Nat. Photon.* **2008**, *2*, 667–669. [[CrossRef](#)]
6. Geng, J. Structured-Light 3D Surface Imaging: A Tutorial. *Adv. Opt. Photon.* **2011**, *3*, 128–160. [[CrossRef](#)]
7. Zhang, Z.H. Review of Single-Shot 3D Shape Measurement by Phase Calculation-Based Fringe Projection Techniques. *Opt. Lasers Eng.* **2012**, *50*, 1097–1106. [[CrossRef](#)]
8. Zuo, C.; Tao, T.; Feng, S.; Huang, L.; Asundi, A.; Chen, Q. Micro Fourier Transform Profilometry (MFTP): 3D Shape Measurement at 10,000 Frames per Second. *Opt. Lasers Eng.* **2018**, *102*, 70–91. [[CrossRef](#)]
9. Zuo, C.; Feng, S.; Huang, L.; Tao, T.; Yin, W.; Chen, Q. Phase Shifting Algorithms for Fringe Projection Profilometry: A Review. *Opt. Lasers Eng.* **2018**, *109*, 23–59. [[CrossRef](#)]
10. Van der Jeught, S.; Dirckx, J.J.J. Real-Time Structured Light Profilometry: A Review. *Opt. Lasers Eng.* **2016**, *87*, 18–31. [[CrossRef](#)]
11. Lu, L.; Suresh, V.; Zheng, Y.; Wang, Y.; Xi, J.; Li, B. Motion Induced Error Reduction Methods for Phase Shifting Profilometry: A Review. *Opt. Lasers Eng.* **2021**, *141*, 106573. [[CrossRef](#)]
12. Pengyu Cong; Zhiwei Xiong; Yueyi Zhang; Shenghui Zhao; Feng Wu Accurate Dynamic 3D Sensing With Fourier-Assisted Phase Shifting. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 396–408. [[CrossRef](#)]
13. Qian, J.; Tao, T.; Feng, S.; Chen, Q.; Zuo, C. Motion-Artifact-Free Dynamic 3D Shape Measurement with Hybrid Fourier-Transform Phase-Shifting Profilometry. *Opt. Express* **2019**, *27*, 2713. [[CrossRef](#)] [[PubMed](#)]
14. Guo, W.; Wu, Z.; Li, Y.; Liu, Y.; Zhang, Q. Real-Time 3D Shape Measurement with Dual-Frequency Composite Grating and Motion-Induced Error Reduction. *Opt. Express* **2020**, *28*, 26882. [[CrossRef](#)] [[PubMed](#)]
15. Lu, L.; Xi, J.; Yu, Y.; Guo, Q. New Approach to Improve the Accuracy of 3-D Shape Measurement of Moving Object Using Phase Shifting Profilometry. *Opt. Express* **2013**, *21*, 30610. [[CrossRef](#)] [[PubMed](#)]
16. Lu, L.; Ding, Y.; Luan, Y.; Yin, Y.; Liu, Q.; Xi, J. Automated Approach for the Surface Profile Measurement of Moving Objects Based on PSP. *Opt. Express* **2017**, *25*, 32120. [[CrossRef](#)] [[PubMed](#)]
17. Liu, X.; Tao, T.; Wan, Y.; Kofman, J. Real-Time Motion-Induced-Error Compensation in 3D Surface-Shape Measurement. *Opt. Express* **2019**, *27*, 25265. [[CrossRef](#)] [[PubMed](#)]
18. Wang, Y.; Liu, Z.; Jiang, C.; Zhang, S. Motion Induced Phase Error Reduction Using a Hilbert Transform. *Opt. Express* **2018**, *26*, 34224. [[CrossRef](#)] [[PubMed](#)]
19. Guo, W.; Wu, Z.; Zhang, Q.; Wang, Y. Real-Time Motion-Induced Error Compensation for 4-Step Phase-Shifting Profilometry. *Opt. Express* **2021**, *29*, 23822. [[CrossRef](#)] [[PubMed](#)]
20. Yu, H.; Zheng, D.; Fu, J.; Zhang, Y.; Zuo, C.; Han, J. Deep Learning-Based Fringe Modulation-Enhancing Method for Accurate Fringe Projection Profilometry. *Opt. Express* **2020**, *28*, 21692. [[CrossRef](#)] [[PubMed](#)]
21. Zhang, L.; Chen, Q.; Zuo, C.; Feng, S. High-Speed High Dynamic Range 3D Shape Measurement Based on Deep Learning. *Opt. Lasers Eng.* **2020**, *134*, 106245. [[CrossRef](#)]
22. Feng, S.; Zuo, C.; Zhang, L.; Yin, W.; Chen, Q. Generalized Framework for Non-Sinusoidal Fringe Analysis Using Deep Learning. *Photon. Res.* **2021**, *9*, 1084. [[CrossRef](#)]
23. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241; ISBN 978-3-319-24573-7.
24. Nguyen, H.; Tran, T.; Wang, Y.; Wang, Z. Three-Dimensional Shape Reconstruction from Single-Shot Speckle Image Using Deep Convolutional Neural Networks. *Opt. Lasers Eng.* **2021**, *143*, 106639. [[CrossRef](#)]
25. Feng, S.; Zuo, C.; Tao, T.; Hu, Y.; Zhang, M.; Chen, Q.; Gu, G. Robust Dynamic 3-D Measurements with Motion-Compensated Phase-Shifting Profilometry. *Opt. Lasers Eng.* **2018**, *103*, 127–138. [[CrossRef](#)]
26. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the International Conference On Machine Learning, Lille, France, 6–11 July 2015*; pp. 448–456.
27. Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011*; pp. 315–323.
28. Su, H.; Qi, C.R.; Li, Y.; Guibas, L.J. Render for Cnn: Viewpoint Estimation in Images Using Cnns Trained with Rendered 3d Model Views. In *Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015*; pp. 2686–2694.
29. Peng, X.; Sun, B.; Ali, K.; Saenko, K. Learning Deep Object Detectors from 3d Models. In *Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015*; pp. 1278–1286.

30. Zheng, Y.; Wang, S.; Li, Q.; Li, B. Fringe Projection Profilometry by Conducting Deep Learning from Its Digital Twin. *Opt. Express* **2020**, *28*, 36568. [[CrossRef](#)]
31. Wang, F.; Wang, C.; Guan, Q. Single-Shot Fringe Projection Profilometry Based on Deep Learning and Computer Graphics. *Opt. Express* **2021**, *29*, 8024. [[CrossRef](#)]
32. Foundation, B. Blender.Org—Home of the Blender Project—Free and Open 3D Creation Software. Available online: <https://www.blender.org/> (accessed on 18 May 2022).
33. Zhou, Q.; Jacobson, A. Thingi10K: A Dataset of 10,000 3D-Printing Models. *arXiv* **2016**, arXiv:1605.04797.
34. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
35. Li, B.; Karpinsky, N.; Zhang, S. Novel Calibration Method for Structured-Light System with an out-of-Focus Projector. *Appl. Opt.* **2014**, *53*, 3415. [[CrossRef](#)] [[PubMed](#)]
36. Slabaugh, G.G. *Computing Euler Angles from a Rotation Matrix*; Technical Report; University of London: London, UK, 1999. Available online: <http://eecs.qmul.ac.uk/~jgslabaugh/publications/euler.pdf> (accessed on 18 May 2022).
37. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An Imperative Style, High-Performance Deep Learning Library. *Adv. Neural Inf. Processing Syst.* **2019**, *32*, 8024–8035.
38. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
39. Zuo, C.; Huang, L.; Zhang, M.; Chen, Q.; Asundi, A. Temporal Phase Unwrapping Algorithms for Fringe Projection Profilometry: A Comparative Review. *Opt. Lasers Eng.* **2016**, *85*, 84–103. [[CrossRef](#)]
40. Li, B.; Wang, Y.; Dai, J.; Lohry, W.; Zhang, S. Some Recent Advances on Superfast 3D Shape Measurement with Digital Binary Defocusing Techniques. *Opt. Lasers Eng.* **2014**, *54*, 236–246. [[CrossRef](#)]