Measurement Science
and Technology

PURPOSE-LED
PUBLISHING™

PAPER

Deep diffusion model-based dual structured light for large-format complex surface imaging

To cite this article: Lei Lu *et al* 2026 *Meas. Sci. Technol.* **37** 065009

View the article online for updates and enhancements.

You may also like

- Handheld structured light system for panoramic 3D measurement in mesoscale
Wenqing Su, Ji Tan, Zhaoshui He et al.

- Accurate and dynamic 3D shape measurement with digital image correlation-assisted phase shifting
Pengyu Hu, Shuming Yang, Fenghe Zheng et al.

- High-precision real-time 3D shape measurement based on a quad-camera system
Tianyang Tao, Qian Chen, Shijie Feng et al.

IOP Publishing Meas. Sci. Technol. 37 (2026) 065009 https://doi.org/10.1088/1361-6501/ae4253

Measurement Science and Technology

Check for updates

RECEIVED
31 October 2025

REVISED
20 January 2026

ACCEPTED FOR PUBLICATION
5 February 2026

PUBLISHED
18 February 2026

PAPER

Deep diffusion model-based dual structured light for large-format complex surface imaging

Lei Lu[1], Yingju Wang[1], Yuejiao Guo[1], Zhilong Su[2,*], Wei Pan[3], Deyang Zhang[4], Qinghui Zhang[1] and Peng Li[1,*]

[1] Institute for Complexity Science, Henan University of Technology, Zhengzhou 450001, Henan, People's Republic of China
[2] Shanghai Institute of Applied Mathematics and Mechanics, School of Mechanics and Engineering Science, Shanghai Key Laboratory of Mechanics in Energy Engineering, Shanghai University, Shanghai 200444, People's Republic of China
[3] Department of R & D, OPT Machine Vision Tech Co., Ltd, Dongguan 523860, Guangdong, People's Republic of China
[4] National Key Laboratory of Ship Structural Safety, China Ship Scientific Research Center, Wuxi 214082 Jiangsu, People's Republic of China
[*] Authors to whom any correspondence should be addressed.

E-mail: szloong@shu.edu.cn and lipeng360@bit.edu.cn

Keywords: phase shifting profilometry, dual-projector system, generative diffusion model

Abstract

Phase shifting-based structured light is widely used in three-dimensional (3D) topographic measurements due to its high accuracy. However, the typical single projector–camera configuration faces challenges such as blind spots, limited measurement range, and slow speed when reconstructing large-format complex objects. To address these limitations, we propose a dual-head structured light prototype system based on generative diffusion models. The system comprises two synchronized projector–camera units with minimal overlap to fuse 3D data. The diffusion model framework resolves fringe pattern aliasing in the overlap region, enabling the dual-head system to capture surfaces with high accuracy. We validate the proposed method through two experiments: one focusing on fringe pattern recovery and the other on 3D reconstruction. The results demonstrate that the method achieves high-quality fringe pattern recovery with a peak signal-to-noise ratio of up to 45.80 dB and a 3D reconstruction accuracy of up to 8 μm.

1. Introduction

Advances in industrial intelligence and modern manufacturing have driven a significant demand for high-fidelity three-dimensional (3D) geometric perception and measurement, especially for structural surfaces with complex curvatures [1–5]. To address this, various techniques such as stereo vision, laser scanning, and structured light fringe projection have been extensively reported in the literature, contributing to the development of diverse 3D imaging systems and devices [6–8]. While stereo vision has made significant progress, its passive nature faces challenges due to texture dependence, resulting in relatively low accuracy. In contrast, laser scanning and structured light approaches use active illumination to capture surface patterns [9–11]. However, laser scanning techniques that utilize dot and line patterns struggle to meet high-density takt time requirements, particularly for in-situ sensing of complex surfaces, and often suffer from low accuracy in 3D information fusion [12, 13].

In contrast, fringe projection-based structured light (FPSL) technique is widely used owing to the high-precision and fast full-field measurement capability. Specifically, recent studies on orthogonal spatial or spatial-temporal binary coding method [5, 14], fringe order correction [15], and complementary phase interleaving-based fringe order recognition approach [8] have paved the way for automated static and dynamic measurement applications with high accuracy. Despite significant progress in 3D measurement, traditional phase-shifting profilometry (PSP)-based systems still face two major challenges. First, the single-camera single-projector system is limited by its projection and shooting angles, hindering multi-view, wide-range 3D reconstruction. Second, complex surface contours or irregularities may obstruct the projected light, creating shadow regions that lack phase information and lead to incomplete

© 2026 IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

or inaccurate reconstruction. Additionally, these obstructions can cause blind regions that the camera cannot capture, further reducing measurement accuracy.

To overcome these issues, the approach proposed in [16] involves acquiring multiple measurements from different viewpoints or rotating the object to enable complete 3D reconstruction. Nevertheless, this approach is time-consuming and has limitations, as some objects cannot be placed on a turntable due to their fixed position or difficulty in movement. Literature [17] proposes that in stereo vision measurement, the 3D shape of an object is initially captured using the shadow profile method, and surface details such as pits, edges, and corners are measured through stereo vision matching techniques. This method involves complex algorithms and experimental procedures. To address the issues of measurement blind spots and limited measurement range, other methods often rely on additional hardware equipment. Nevertheless, when a dual or multiple structured light system simultaneously projects fringes onto the object, aliased fringes are formed on the object's surface, leading to incorrect reconstruction of the object.

Over the past few years, several approaches have emerged to address the problem of aliased fringes recovery. Wu *et al* [18] proposed a method using a semi-transparent and semi-reflective mirror to split the projected beam into two paths. These paths were then color-separated by red and blue filters, respectively. Subsequently, an all-reflective mirror was employed to project the two monochromatic beams onto the object's surface, forming an aliased fringe pattern. This aliased projection is captured by a color CCD camera, and the aliased separation is performed using a mechanism of color perception and processing. This separation method requires a high level of optical path design, precise system hardware placement, and optimized system performance. Maimon *et al* [19] mitigated the interference from aliased projections by adding a hardware structure—a motion sensor—and using a high spatial frequency mode to enhance the contrast between the sensor's projection pattern and that of other patterns. This method suffers from instability due to its sensitivity to variations in environmental conditions and hardware alignment, which limits its reliability in certain applications. Je *et al* [20] eliminated the aliased fringe patterns by multiplexing the orientations of projector patterns and the color gamut characteristics of colored fringe patterns, enabling the separation of aliased fringe patterns through explicit derivative calculations. This method requires demanding lighting conditions and poses practical difficulties in experiments. On the other hand, Bai *et al* [21] proposed a dual-projector single-camera measurement system to separate superimposed signals using the grayscale complementary relationship, but this approach did not improve the measurement range. Li *et al* [22] addressed the issue of aliased fringe patterns by controlling the order of the projected fringe patterns in a dual-projector system. Their method combines the four-step phase-shift technique with the dual-frequency method, but this approach has strict requirements regarding the order of the projected fringe patterns.

FPSL is capable of obtaining high-accuracy and high-density 3D geometries [23–25]. However, it still faces several challenges in multiple FPSL imaging scenarios, such as system calibration and global point-cloud fusion [12]. A key reason for these challenges is that adjacent FPSL systems cannot perform synchronous imaging with overlapping fields of view (FOV). Specifically, as shown in figure 1, different light patterns may be projected by the two projectors in the overlap area, leading to aliasing in the observed fringe images. Consequently, significant errors are introduced into the reconstruction results. Although several studies have been reported to address this issue, to the best of our knowledge, none of them has been able to solve the aliasing problem in simultaneous 3D imaging tasks with local overlapping multi-head FPSL systems.

Therefore, this paper presents a dual-head FPSL imaging approach based on generative deep learning, serving as a portable and scalable baseline to overcome the aforementioned issue. As illustrated in figure 1 , the dual-head FPSL is configured with a pair of synchronous projector–camera components, each measuring a portion of the target surface, but with their FOVs locally overlapping. The overlap region provides a strong posterior constraint for fusing the two parts of the measured 3D geometry. However, when the two projectors project fringe patterns simultaneously, both cameras observe aliasing in the overlap region, leading to unstable reconstruction. To address this issue, we introduce a generative deep learning method based on diffusion models to resolve fringe pattern aliasing, enabling successful 3D geometry reconstruction of the overlap region using the two projector–camera components. Unlike existing methods that eliminate aliasing by sacrificing simultaneous measurement capability, our approach directly separates the observed aliased images by training a denoising diffusion model, thus facilitating simultaneous reconstruction and 3D point cloud fusion. Specifically, the phase estimation and 3D reconstruction of each imaging head of the system follow the principles of conventional fringe projection profilometry; and the 3D point cloud fusion is performed by the iterative closest point (ICP) algorithm. Both techniques can be efficiently implemented.

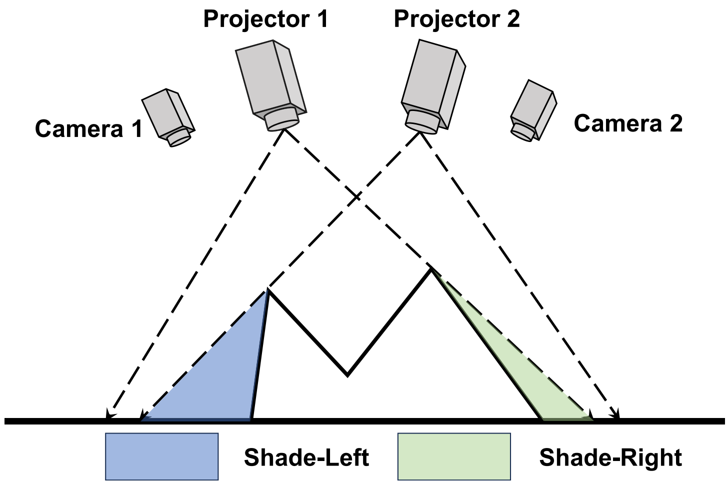

Figure 1. Schematic diagram of the dual-head structured-light system, consisting of a left head (camera–projector 1) and a right head (camera–projector 2) for large-format complex surface imaging, where the shaded regions observed by the left and right heads cannot be observed by the right and left heads, respectively.

The rest of this paper is organized as follows: section 2 introduces the principle of traditional PSP and the aliasing image formation in dual-head FPSL, followed by a description of how to generate fringe patterns without aliasing using the diffusion model framework. In section 3, we experimentally validate the performance of the proposed method. Section 4 concludes the paper.

2. Methodology

Figure 2 presents the overall workflow of the proposed dual-head FPSL system integrated with a diffusion-based dealiasing network. The pipeline consists of four main stages: (1) synchronous acquisition of phase-shifted fringe patterns using a dual-projector and dual-camera configuration, (2) diffusion-based removal of fringe aliasing, (3) phase recovery through wrapped and unwrapped phase computation, and (4) 3D reconstruction based on calibrated system parameters. Owing to the dual-head configuration, two partial point cloud slices are obtained and subsequently fused into a complete 3D model using the ICP algorithm. In the following sections, we begin with a revisit of the principle of traditional PSP (section 2.1) to introduce the core methods of the dual-head FPSL including its configuration and image formation in step 1 (section 2.2), dealiasing with diffusion model in step 2 (section 2.3), and phase estimation and point cloud reconstruction in steps 3 and 4 (section 2.4).

2.1. Principle of traditional PSP

The sinusoidal fringe patterns captured by the camera can be represented as follows:

$$I_n(x,y) = a + b\cos\left[\varphi(x,y) + \frac{2\pi}{N}\right] \tag{1}$$

where a, b, φ(x,y) are the background light intensity of the image, the modulation of the sinusoidal fringe pattern, and the phase value on the surface of the object to be sought, respectively, and N is the step size of the projected fringe pattern [26], and the existence of the three unknowns requires that at least three phase-shifted images are needed, and a four-step phase-shifting method is usually used, so that the distribution of light intensity of four fringe patterns is:

$$\begin{cases} I_1(x,y) = a + b\cos[\varphi(x,y)] \\ I_2(x,y) = a + b\cos\left[\varphi(x,y) + \frac{\pi}{2}\right] \\ I_3(x,y) = a + b\cos[\varphi(x,y) + \pi] \\ I_4(x,y) = a + b\cos\left[\varphi(x,y) + \frac{3\pi}{2}\right]. \end{cases} \tag{2}$$

The phase distribution is calculated from equation (3):

$$\varphi = \arctan\left(\frac{I_1 - I_3}{I_0 - I_2}\right). \tag{3}$$

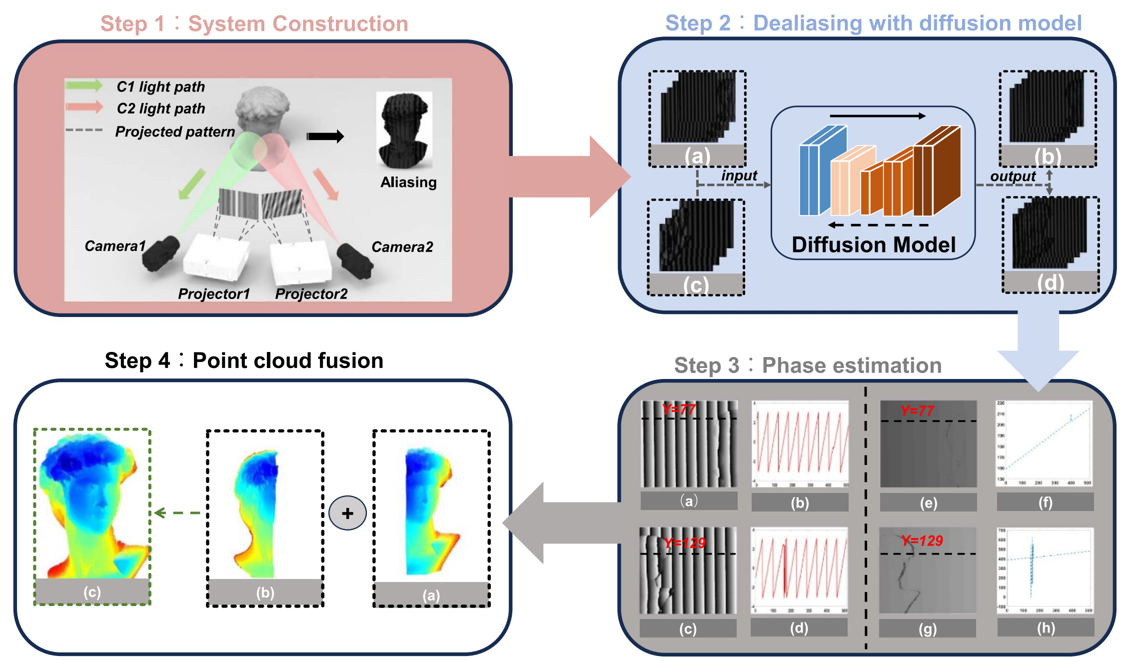

Figure 2. Overview of the work pipeline for the proposed dual-head FPSL method.

The phase obtained by the four-step phase shift algorithm has a jagged distribution between [0,2π], and phase unwrapping is required to obtain a continuous absolute phase [27]. In this paper, we use the multi-frequency heterodyne principle for phase unwrapping to obtain the absolute phase information of the measured object [28]. To obtain the 3D information of an object, it is necessary to calibrate both the camera and the projector. After acquiring the calibration data, the world coordinates of the object can be determined by combining the transformation relationships between the coordinate systems.

2.2. Image formation of due-head FPSL

According to the imaging principle of CMOS cameras, the observed light pattern results from the camera's integration of all incoming light during the exposure period. Consider a pixel point $\mathbf{u} \in \mathbb{R}^2$ in the image sensor. Its intensity $I(\mathbf{u})$, which is the photoresponse that reflected from some points on the target surface within exposure time t_e, can be formulated as an integral process starting from the shutter time t_i:

$$\mathcal{I}(\mathbf{u}) = \int_{t_i}^{t_i+t_e} K_l K_r \sum_{\mathbf{v}} R(\mathbf{u},\mathbf{v})\, l(\mathbf{v},t)\,\mathrm{d}t + \int_{t_i}^{t_i+t_e} c(t) + K_r c'(t)\,\mathrm{d}t \tag{4}$$

where $l(\mathbf{v})$ represents the lighting intensity of pixels at $\mathbf{v}$ in the projector contributing to $I(\mathbf{u})$, $R(\mathbf{u},\mathbf{v}) \in [0,1]$ weights the contribution of the projected light corresponding to the aperture effect, K_l and K_r are the factors of projected light conversion and object surface reflection, respectively, c and c' indicate the direct and reflected ambient light components. For a dual-head FPSL as shown in figure 3, when the pixel in $\mathbf{u}$ viewing the overlapped FOV formed by the intersection of captured regions A and B, the first term in equation (4) will be:

$$\int_{t_i}^{t_i+t_e} K_l K_r \sum_{n=1}^{2} \sum_{\mathbf{v}_n} R_n(\mathbf{u},\mathbf{v}_n)\, l_n(\mathbf{v}_n,t)\,\mathrm{d}t \tag{5}$$

where $n = 1,2$ represents the two projectors (projector 1 and 2) and $\mathbf{u} \in \Omega_{\mathrm{FOV}}$. We observe that the observed light pattern in the overlapping FOV by both cameras is formed through a complex process involving weighting and summation in the local image domain Ω_{FOV}, leading to long-standing challenges that have hindered the development of seamless structured light arrays. This motivates us to solve the problem by proposing the method based on generative deep learning approach. We aim to learn a neural separator $\mathcal{P}: \mathcal{I} \mapsto \mathcal{I}(l_n)$, which maps a locally aliased light pattern to a clear fringe pattern corresponding to the projected pattern l_n, within the diffusion denoising learning framework.

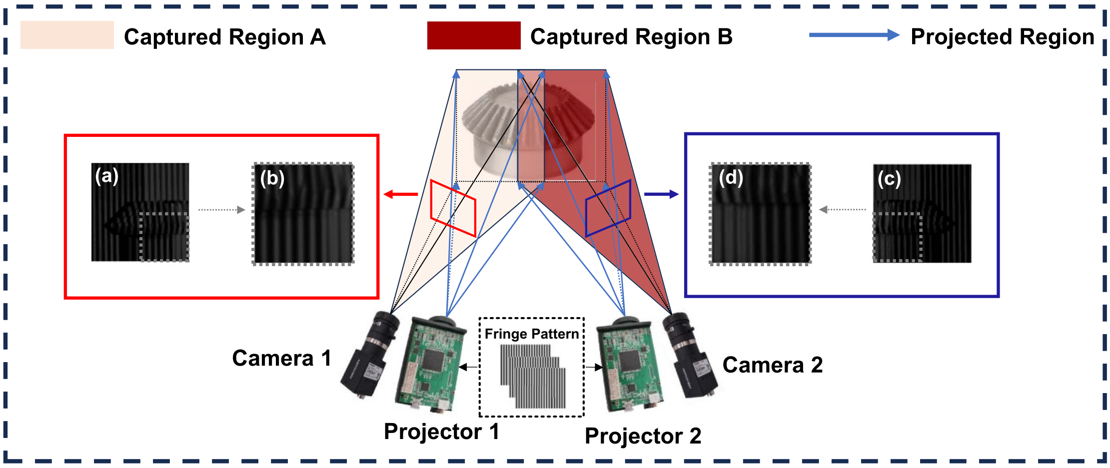

Figure 3. Dual-head structured-light system with an overlapping field of view between regions A and B. The fringe images in (a) and (b) are captured by camera 1 and camera 2, respectively. The corresponding aliased fringe patterns observed in these images are shown in (c) and (d), respectively.

To this end, we reformulate equation (4) to align with the line of denoising diffusion pipeline. As the background light contained in the fringe patterns is assumed not to affect the phase estimation, we consider only the first aliasing term in equation (4) and expand the sum on the projector index n, resulting in:

$$\mathcal{I}(\mathbf{u}) = \mathcal{I}(\mathbf{u};l_1) + \mathcal{I}(\mathbf{u};l_2), \mathbf{u} \in \Omega_{\mathrm{FOV}} \tag{6}$$

where $\mathcal{I}(\mathbf{u};l_1) = \int_{t_\mathrm{i}}^{t_\mathrm{i}+t_\mathrm{e}} K_\mathrm{l}K_\mathrm{r}\sum_{\mathbf{v}_1} R_1(\mathbf{u},\mathbf{v}_1)l_1(\mathbf{v}_1,t)\mathrm{d}t$ and $\mathcal{I}(\mathbf{u};l_2) = \int_{t_\mathrm{i}}^{t_\mathrm{i}+t_\mathrm{e}} K_\mathrm{l}K_\mathrm{r}\sum_{\mathbf{v}_2} R_2(\mathbf{u},\mathbf{v}_2)l_2(\mathbf{v}_2,t)\mathrm{d}t$. With this formula, the intensity of the pixel $\mathbf{u}$ in Ω_{FOV} can be given by

$$\mathcal{I}(\mathbf{u};l_n) = \mathcal{I}(\mathbf{u}) - \mathcal{I}(\mathbf{u};l_{3-n}) \tag{7}$$

for the projected pattern l_n. Equation (7) rightly describes a denoising process if we view $\mathcal{I}(\mathbf{u};l_{3-n})$ as a noisy pattern, allowing us to train the neural separator $\mathcal{P}$ to identify $\mathcal{I}(\mathbf{u};l_n)$ within the denoising diffusion framework [29], given that $\mathcal{I}(\mathbf{u})$. Details are given in the following content of this paper.

2.3. Learning within the conditional denoising diffusion framework

The overall framework of training the neural network $\mathcal{P}$ within the conditional diffusion framework is shown in figure 4. Though most of the current networks are capable of performing the task of image noise prediction, we here propose to use the U-Net to predict the sample noise because it shows excellent performance in neural fringe pattern processing [30].

The network $\mathcal{P}_\theta$ is dedicated to learning about the conditional distribution of the Gaussian noise added to the original patterns $\{\mathbf{x}_0\}$ corresponding to $\mathcal{I}(\mathbf{u};l_n)$ in equation (7). This means that during the aliased pattern recovery process, the model will particularly emphasize the impact of the conditional signal y (corresponding to the observed fringe patterns $\mathcal{I}(\mathbf{u})$ in equation (7)) on fringe pattern generation. The learning of the conditional diffusion model is divided into two processes: forward diffusion and reverse diffusion, as shown in the upper part of figure 4. In the forward diffusion stage, Gaussian noise $z \sim \mathcal{N}(0,I)$ is continuously added to the sample data $\{\mathbf{x}_0\}$ until the patterns are completely corrupted.

This process can be represented by a fixed Markov chain in equation (8) [31, 32], where the next state of the sample data is obtained by sampling a Gaussian distribution, which can also be interpreted as a process in which the data is progressively corrupted over time step $t \in [0,T]$,

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t|x_{t-1}\sqrt{1-\beta_t}, \beta_t I; y\right). \tag{8}$$

In equation (8), $\beta_t = \sigma_t^2$ determines the amount of noise added by diffusion at each step. The noisy state x_t can also be derived from the initial state x_0 by following:

$$q(x_t|x_0) = N\left(x_t; \sqrt{\bar{\alpha}_t}x_0, \left(1-\bar{\alpha}_t\right)I; y\right) \tag{9}$$

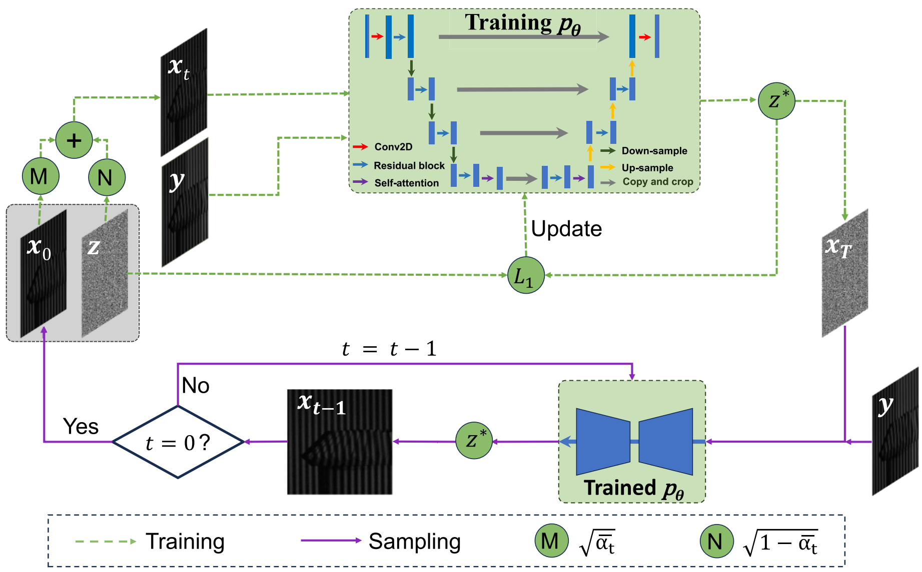

Figure 4. Training and sampling framework to generate non-aliasing fringe pattern with diffusion model.

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, $\alpha_t = 1 - \beta_t$. The reverse diffusion process aims to learn the network $\mathcal{P}_\theta$, given that the condition y, to predict the noise added to the initial pattern x_0. However, the diffusion process is complex and recovering x_0 directly from the noisy state x_t is bound to have a large error. Therefore, in this process, the network is trained to predict the noise distribution in the time step t to obtain x_{t-1} iteratively [33]. Based on Bayesian posterior probability, the backward diffusion process can be described as:

$$p_\theta(x_{0:T}) = p(x_T)\,\Pi_{t=1}^{T} p_\theta(x_{t-1}|x_t) \tag{10}$$

where $p(x_T) = \mathcal{N}(x_T; 0, I)$ and $p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta, \sigma_t^2 I, y)$. The mean μ_θ is estimated by the trainable network as $\mu_\theta = \mathcal{P}_\theta(x_t, t, y)$. Once the network is trained, the dealiased fringe pattern is given by a sampling procedure, as shown in the lower part of figure 4. The iteration kernel in the sampling pass can be expressed as follows:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\mu_\theta\right) + \sigma_t z. \tag{11}$$

For training the network $\mathcal{P}_\theta$, we trigger one head of the dual-head FPSL system to obtain the fringe pattern $\{x_0\}$ without aliasing firstly and then, trigger both heads to capture the corresponding aliased fringe patterns $\{y\}$, resulting in the dataset $\{x_0, y\}$ where the aliased observations are used as conditions in the above training pipeline. Given an initial pattern x_0 associated with a condition y, the key step is to generate time step t to regulate the diffusion processes as illustrated in equations (8) and (10) since the model does not have the capability to recognize the specific time step to which the input image belongs. We here adopt a mechanism similar to positional encoding in transformers [34, 35]. The time step is input as an additional information together with x_t into the network to produce a noise prediction $z^* \sim \mathcal{N}(\mu_\theta, \sigma^2 I)$. Followed by this, we measure the difference between z and z^* using L_1 loss to train the model using backpropagation algorithm. The loss function is given by: $L_1(z, z^*) = \frac{1}{n}\sum_{i=1}^{n} |z_\text{i} - z^*{}_\text{i}|$ and the training and test loss dynamics are shown in figure 5.

After the network training is completed, the model is able to complete high-quality image reconstruction based on the conditioned signals. The detailed steps are as follows: Starting from time step T, first randomly sample from the standard Gaussian distribution to make a noisy sample x_t. Subsequently, at each time step t, the noisy image x_t, the time step encoding, and the conditioned signal are used as inputs to the network to obtain the estimated noise z^*. With the help of the estimation noise and the inverse diffusion formula, the data distribution x_{t-1} at the previous step can be obtained, and the non-aliased fringe pattern x_0 (i.e. $\mathcal{I}(\mathbf{u}; l_n)$) can be obtained through T-step iterations.

It is worth noting that the conditional denoising diffusion framework above is a probabilistic generative modeling approach that learns the conditional distribution of clean, non-aliased fringe patterns given aliased observations by coupling the forward Markovian diffusion process in equation (8) with

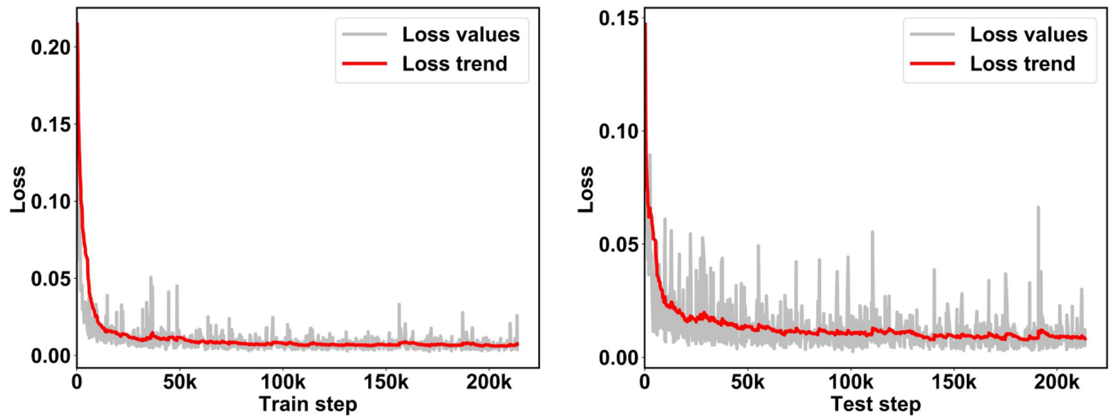

Figure 5. Loss curves of training (left) and testing (right) the model $\mathcal{P}_\theta$.

the learned, condition-guided reverse denoising process. In this framework, clean fringes are progressively perturbed into Gaussian noise, and the neural network is trained to iteratively invert this process while being constrained by the observed aliased fringe patterns, as shown in figure 4. This formulation is highly effective for fringe aliasing removal because the conditional reverse process enforces consistency with the input measurements, while the multi-step denoising trajectory provides stable convergence toward physically plausible solutions, enabling robust recovery of non-aliased fringes under complex and highly variable aliasing conditions. Aliasing separation relies on two key mechanisms: progressive noise refinement over multiple steps and condition-guided reconstruction using aliased fringes as input with U-Net skip connections to fuse low- and high-level features. This approach provides more stable and accurate reconstruction than GANs, which may suffer from mode collapse, or autoencoders, which can lose high-frequency details. By implicitly capturing the underlying properties of fringe patterns, the diffusion model achieves high-fidelity recovery with a relatively small training set, as confirmed by high peak signal-to-noise ratio (PSNR), low mean squared error (MSE), and low reconstruction error in section 3.

2.4. 3D shape reconstruction

Followed by obtaining the non-aliased fringe images for each head of the proposed dual-head FPSL system with the trained diffusion model $\mathcal{P}_\theta$, we can estimate the phase maps to recover the 3D shape of the object surface being measured. In our implementation, we first estimate the phase map for each camera-projector head to reconstruct the object surface part that observed by the head and then fuse them with point cloud registration to obtain the complete 3D geometry in the FOV of the dual-head FPSL system.

For phase estimation, we use the well-established four-step phase-shifting algorithm introduced in section 2.1. As each camera-projector head is a typical structured light unit when the fringe aliasing in the overlapping region is removed, the four-step phase-shifting algorithm can be used to estimate the wrapped phase map for each head and then to retrieve the absolute phase map via multi-frequency heterodyne unwrapping. Once the absolute phase maps are obtained for both heads, the 3D shape of the part of the object observed by each head can be reconstructed immediately via triangulation by accessing the horizontal (or vertical) phase values Φ_x from the estimated absolute phase map and the pixel coordinates $\mathbf{u} = (u, v)$ as follows:

$$\mathbf{X} = \mathrm{tri}\left(\mathbf{P}_\mathrm{C}, \mathbf{P}_\mathrm{P}, \mathbf{u}, f(\Phi_x)\right), \tag{12}$$

where $\mathbf{P}$ the reconstructed 3D coordinates of the surface points, $\mathbf{K}_\mathrm{C}$ and $\mathbf{K}_\mathrm{P}$ are the calibrated projection matrices of the projector and camera, respectively, which can be determined by Zhang's calibration algorithm [36] or the improved variant [37], and $f(\Phi_x) = \frac{\Phi_x(\mathbf{u})}{2\pi N} W_\mathrm{P}$ is the coordinate function in the projector plane with width of W_P and fringe period N.

Finally, we fuse the 3D shapes reconstructed by both camera–projector heads using the ICP algorithm to obtain the full geometric shape of the object surface in the FOV of the dual-head FPSL system. Because of the overlapping region as shown in figure 3, the ICP algorithm can be implemented

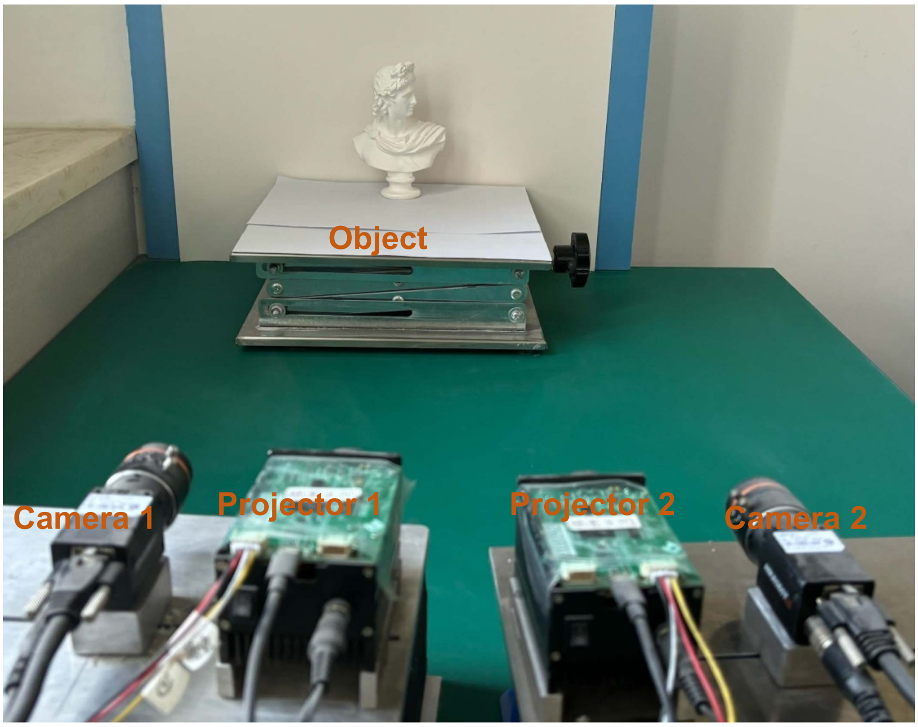

Figure 6. Dual-head projection hardware system.

by finding the point cloud correspondences between the overlapped parts and estimating the rigid transformation by minimizing

$$\min_{\mathbf{R},\mathbf{t}} \sum_i \|\mathbf{R}\mathbf{X}_i + \mathbf{t} - \mathbf{Y}_i\|^2, \tag{13}$$

where $\mathbf{X}_i$ and $\mathbf{Y}_i$ denote corresponding 3D points in the overlapping regions, and $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and translation vector, respectively.

3. Experiments

This experiment aims to address the issue of measurement limitations (including shooting and projection limitations) that arise when measuring larger objects or objects with complex morphology. A dual-camera dual-projector imaging system is proposed to overcome these challenges. Capturing the complete 3D information of an object requires two cameras. Therefore, the reconstruction process involves creating two datasets for model training and employing point cloud fusion to obtain the complete 3D surface geometry. A dual-projection structured light system, comprising two TJ20U2 projectors and two CMOS cameras with a resolution of 3072 × 2048, was utilized to acquire the training and test datasets. Figure 6 illustrates the constructed dual-projection hardware system. Synchronization between the projectors and cameras was achieved using synchronous control lines, with custom software enabling simultaneous projection and triggering of the corresponding cameras for synchronized data acquisition.

3.1. Dataset and training

As previously mentioned, the proposed method requires aliased fringe patterns and their corresponding non-aliased versions to train the diffusion network $\mathcal{P}_\theta$. During the training process of the diffusion model, the aliased pattern serves as a conditional input, guiding the data distribution of the generated image with high accuracy. By varying the frequency and initial phase values, we generated 4043 samples for model training and 1583 samples for model testing, each with a size of 512 × 512 pixels. The dataset creation process involves the following steps: first, sinusoidal fringe patterns are designed based on the phase-shift method and loaded into the left and right projectors. The trigger signals for both projectors are synchronized to ensure simultaneous projection onto the object's surface, generating aliased fringe

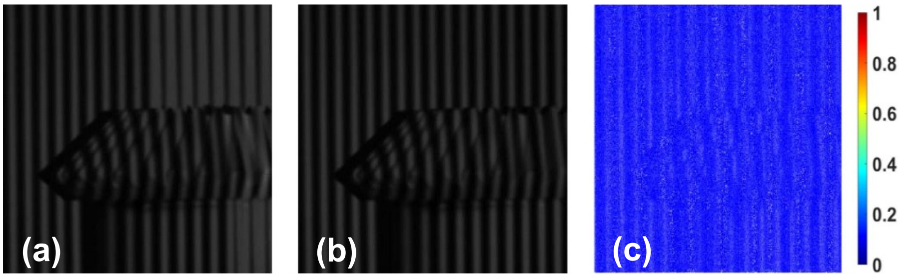

Figure 7. Demonstration of the dealiasing quality of the proposed method. (a) shows an aliased fringe pattern captured by one head of the dual-head FPSL system; (b) shows the corresponding dealiased fringe pattern recovered by the network model; and (c) shows the difference between (b) and the non-aliased fringe pattern.

patterns. The camera captures these aliased fringe patterns, with data acquisition ending when the trigger signal ceases. Finally, the corresponding non-aliased fringe patterns are also captured to complete the dataset.

Due to the similar working modes on both sides, we will take the left area as an example for introduction. When capturing the left non-aliased image, the right projector needs to be turned off. The non-aliased pattern serves as a condition during the training process of the diffusion model, ensuring that it can accurately guide the data distribution of the generated image. For the generation of projection fringes, we have chosen to apply the PSP theory by selecting four sinusoidal fringe patterns with periods of 100, 110, 120, and 130. Each period includes four phase-shifted patterns with a phase shift of 0.5π. By varying the initial phase, a large number of aliased fringe patterns are obtained. Since the objects do not need to maintain the same position during dataset generation, the network must not only learn the data distribution and key features of aliased fringe patterns at different object attitudes but also effectively reconstruct these patterns based on the corresponding labeled data distribution to restore the true morphology of the experimental object. Given the complexity of the model's training task, a sufficient amount of sample data is required to meet the network's needs. Specifically, 4043 samples are used for model training to enhance the network's learning capabilities, while 1583 samples are reserved for testing to ensure a comprehensive and objective evaluation of the model's performance.

In the training process, the number of diffusion steps is set to $T = 4000$. The noise variance added by the forward process increases linearly from 1×10^{-5} to 0.0002, while the noise variance of the inverse process is maintained consistent with that of the forward process. The total number of training epochs is 800.

3.2. Validation of fringe pattern separation performance

The fringe pattern separation performance is evaluated first, and the result is presented in figure 7. Figure 7(a) shows the captured aliased fringe patterns, while figure 7(b) displays the estimated patterns obtained by the proposed method. The estimated fringe patterns are then compared with the non-aliased ones, and the result is shown in figure 7(c). The performance is quantitatively evaluated using the PSNR, MSE, and structural similarity index (SSIM). The PSNR reaches 45.80 dB, the MSE is 1.85, and the SSIM is 0.98.

To better demonstrate the potential of the model in image generation, we analyze its performance from the perspectives of phase computation and object reconstruction. The experiment is based on the theory of PSP, where the model needs to recover a set of aliased fringe patterns using a four-step phase shift for phase computation. A sample dataset captured by a structured light system is used in the experiment. In figures 8, (a)–(c) show the wrapped phase obtained using the four-step phase-shifting method. Figure 8(a) displays the wrapped phase image obtained from a single projection, (b) shows the wrapped phase image obtained from the fringe pattern recovered by the network, and (c) presents the heatmap generated by subtracting (b) from (a). To better illustrate the phase differences, the unwrapped phases from figures 8(b) and (c) are extracted along the same row at $y = 350$ pixels, and the phase curves are plotted in figure 8(e) with a local magnification. The results indicate that the two phase curves exhibit a high degree of consistency, with a maximum phase difference of 0.05 rad.

3.3. Validation on 3D reconstruction

Finally, the object is reconstructed using the dual-head FPSL system, and the results obtained from the left and right sides are shown in figure 9(c). The object is successfully reconstructed. To evaluate the reconstruction performance, the results obtained by the traditional method using the same aliased fringe

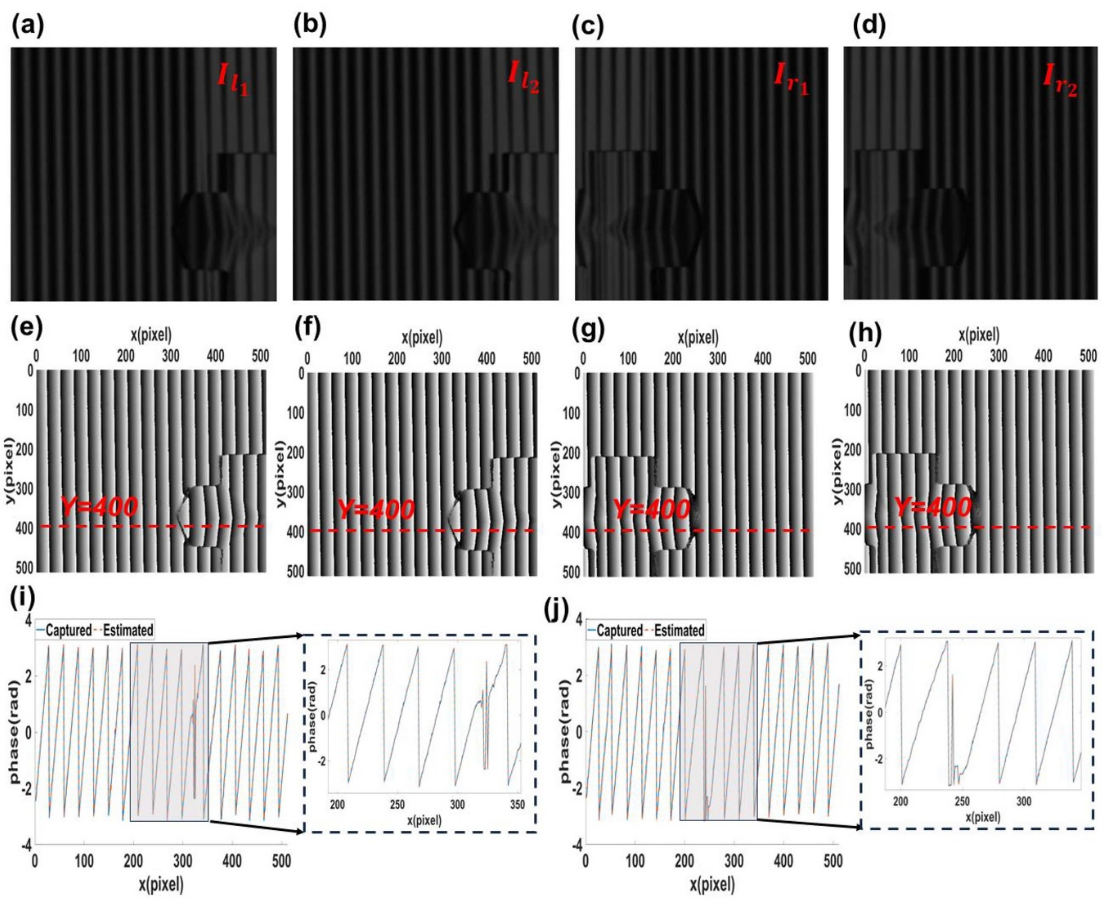

Figure 8. Results of phase comparison. (a)–(c) show the wrapped phase maps obtained from the aliased, inferred, and non-aliased fringe patterns, respectively. (d) shows the difference between (b) and (c). (e) presents the phase values along the cross section of (b) and (c) at $y = 350$.

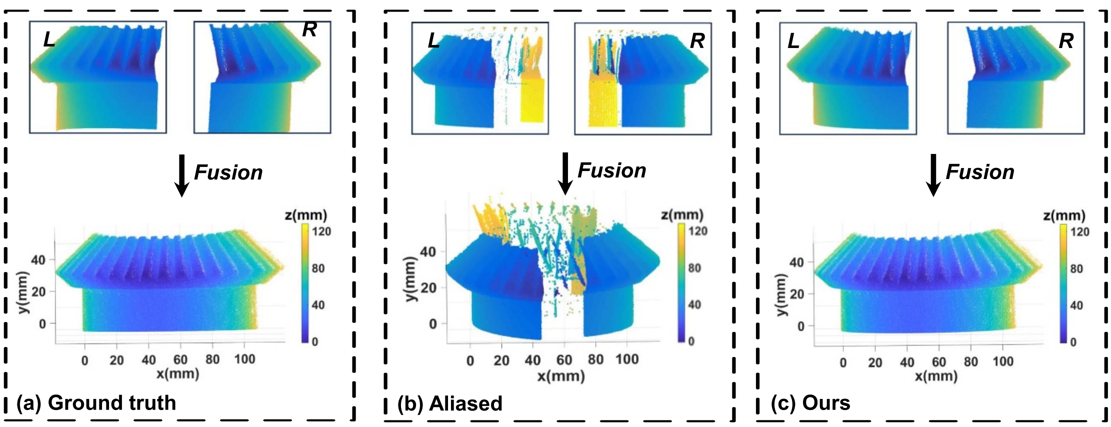

Figure 9. Comparison of reconstruction results. (a) shows the point clouds reconstructed separately by the two structured light systems and their fused result, which is used as the ground truth. (b) shows the reconstructed point clouds and fused result when both systems operate simultaneously, where fringe aliasing occurs. (c) shows the reconstructed point clouds after aliasing recovery using the proposed diffusion model, along with the fused complete point cloud.

pattern and the non-aliased fringe pattern are shown in figures 9(b) and (a), respectively. Significant errors are introduced by fringe pattern aliasing in (b). To quantitatively analyze the reconstruction performance, the root mean square is calculated, and the result is 0.057 mm. The results demonstrate that the proposed method can successfully reconstruct the object and effectively eliminate errors caused by fringe aliasing.

To further validate the effectiveness and superiority of the proposed deep diffusion model–based dual structured light method for 3D reconstruction, a comparative experiment using a derivative-based

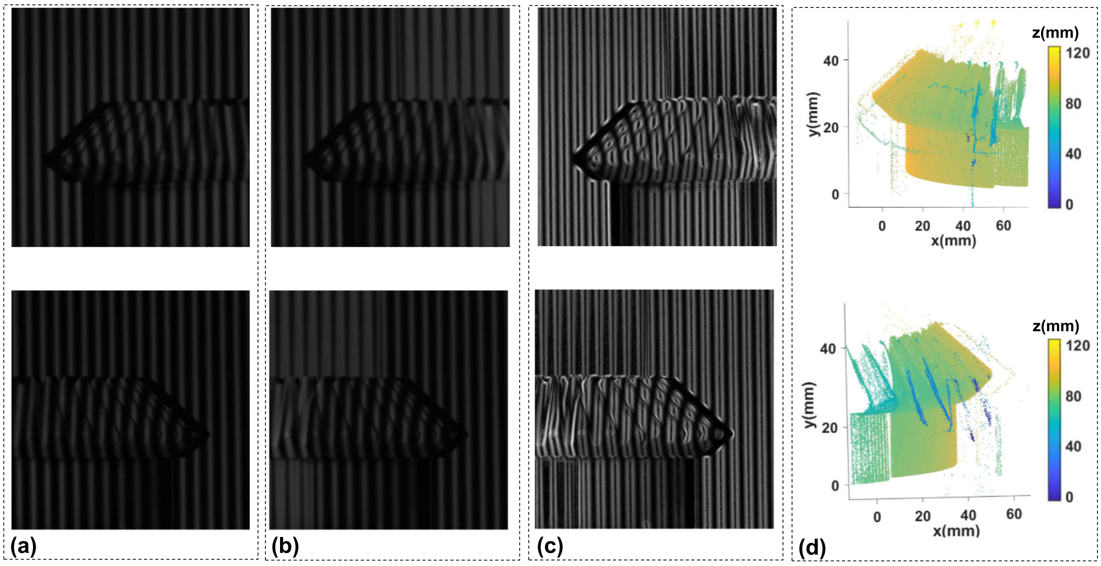

Figure 10. Comparative experiment results. (a) Shows the fringe patterns captured when the dual structured light systems operate independently. (b) Presents the aliased fringe patterns acquired when the two systems operate simultaneously. (c) Illustrates the fringe patterns recovered from (b) using the derivative-based aliasing recovery method. (d) Shows the 3D point cloud reconstructed from the recovered fringe patterns in (c).

method was conducted [22]. As illustrated in figure 10, (a) shows the non-aliased fringe patterns captured when the two structured light systems operate independently. These images serve as the reference for evaluating the performance of subsequent aliased fringe recovery methods. (b) Presents the aliased fringe patterns acquired when both systems operate simultaneously, (c) depicts the non-aliased fringe patterns recovered using the derivative-based method. As a traditional aliased fringe recovery approach, the derivative-based method attempts to restore non-aliased fringes by computing the spatial derivatives of the fringe patterns, (d) shows the 3D point cloud reconstruction results obtained using the fringe patterns recovered by the derivative-based method. From the 2D perspective, a comparison between (a) and (c) reveals evident limitations of the derivative-based method when dealing with locally aliased fringe patterns. Compared with the original non-aliased fringes in (a), the recovered fringes in (c) become noticeably denser, accompanied by significant local intensity amplification. This indicates that, while suppressing aliased components, the derivative-based method inevitably disturbs non-aliased regions, leading to a degradation in fringe quality. From the 3D perspective, the reconstructed point cloud in (d) contains a large amount of noise and exhibits pronounced layering artifacts. This is mainly because the errors introduced during derivative-based aliased fringe recovery are further amplified in the subsequent 3D reconstruction process, severely compromising both reconstruction accuracy and completeness. In contrast, the proposed method demonstrates clear advantages. By leveraging a diffusion model, the system directly restores fringe information in aliased regions, thereby effectively avoiding the error propagation commonly introduced by traditional recovery methods and enabling more accurate and robust 3D reconstruction.

To more intuitively demonstrate the phenomenon of sparse or missing point clouds in blind regions during the 3D reconstruction of objects with complex contours, a prism with prominent surface protrusions was selected as the test object. This strategy amplifies the blind spot phenomenon caused by occlusion. It also highlights the significant role of the dual-camera and dual-projector system in expanding the measurement range. Additionally, it demonstrates the strong generalization capability of the network model. Figure 11 primarily presents the aliased fringes captured by the two systems, along with the results of analyzing the recovery of these fringes from a phase perspective. In addition, figure 12 illustrates the 3D reconstruction results of the test object under various conditions. In figures 11(a)–(d) depict aliased fringes captured by two systems, with noticeably thicker fringes in the aliased regions. (e) shows the wrapped phase calculated from the fringe image captured by the left single-camera single-projector system, which serves as the baseline for phase-level evaluation of aliased fringe recovery. (f) displays the wrapped phase calculated from the fringe image estimated by the network. (g) presents the wrapped phase obtained from the right single-camera single-projector system, while (h) shows the wrapped phase after the network recovers the aliased fringes on the right side. (i) illustrates the data distribution of (e) and (f) at $y = 400$ pixels in the same image, with local magnification indicating consistent distributions.

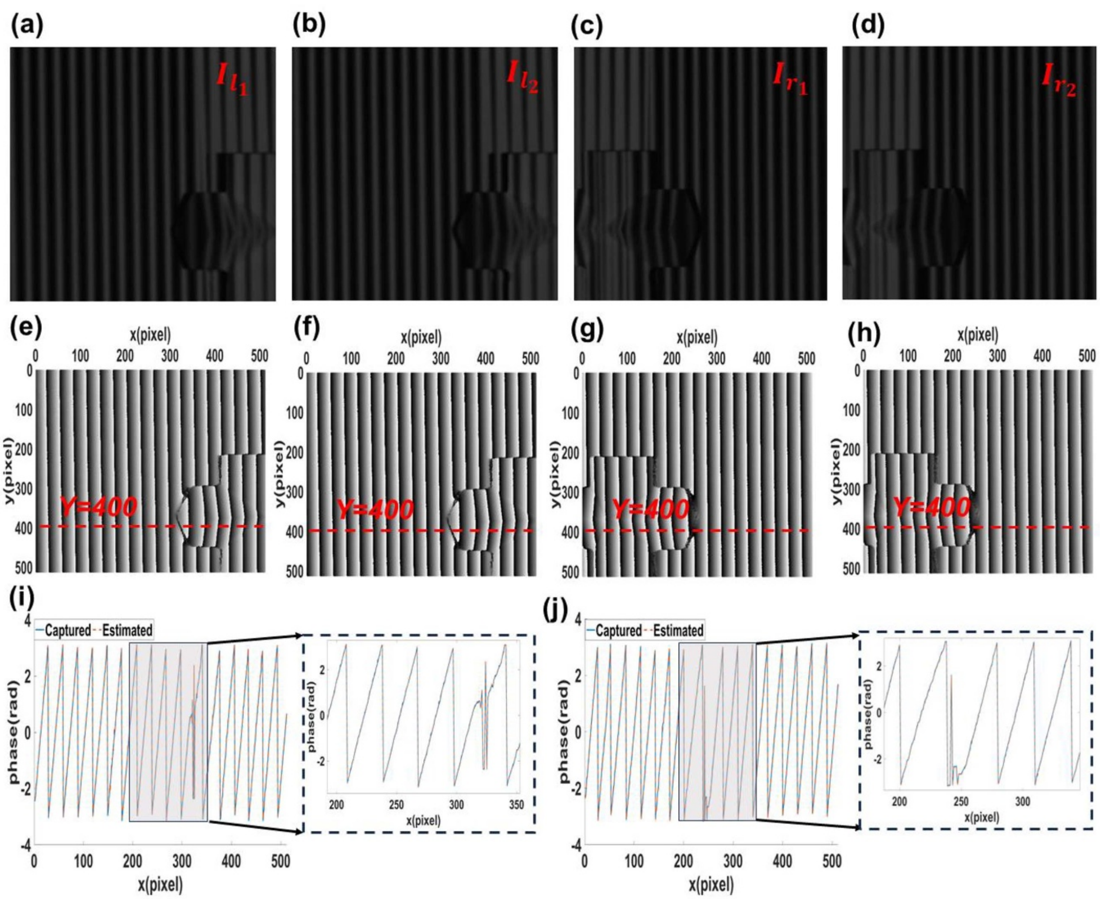

Figure 11. Demonstration of phase estimation. (a)–(d) Show the aliased fringe patterns captured by the dual structured-light system. (e) and (g) Present the wrapped phase maps computed from fringe patterns acquired when the two systems operate independently, whereas (f) and (h) show the wrapped phase maps obtained after recovering the aliased fringe patterns using the proposed diffusion model. (i) Provides a magnified view of (e) and (f) along the line $y = 400$ pixels, and (j) shows the corresponding magnified results of (g) and (h) along the same line.

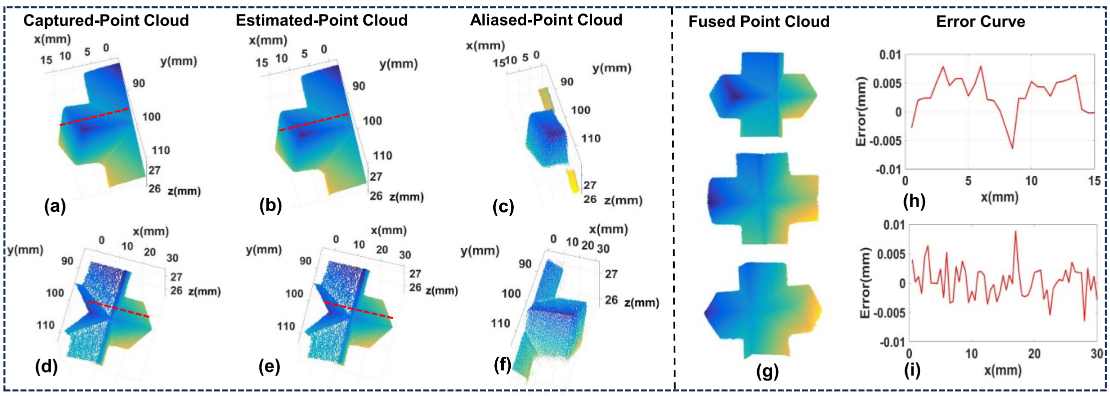

Figure 12. Demonstration of reconstruction results. (a)–(c) Present the reconstructions obtained from the left structured-light system, while (d)–(f) show those from the right structured-light system. (g) Depicts the three recovered poses after point-cloud fusion based on fringe patterns restored by the diffusion model. (h) and (i) Show the corresponding error analyses along the dashed lines indicated in (a) and (b), and in (d) and (e), respectively.

Similarly, (j) demonstrates the data distribution of (g) and (h) at $y = 400$ pixels in the same image, with local magnification also showing consistency.

In figures 12(a)–(c) show the reconstruction results of the left system, indicating that the reconstruction using the network-recovered fringes is comparable to that of the single-camera-single-projector system. (d)–(f) Present the reconstruction results of the right system, which demonstrate excellent quality. (g) Illustrates the distributions of the fused point clouds from different perspectives, clearly showing the expanded reconstruction range and the filling of sparse areas after point cloud fusion. (h) and (i) display the reconstruction error analysis after sampling the depth profiles along the dashed lines in (a)–(b) and (d)–(e), respectively. The results show that the reconstruction error remains within 8 μm, with an MAE of 0.046 mm. Our results clearly demonstrate that the diffusion model effectively handles the recovery of image aliasing. By integrating an attention mechanism into the U-Net structure for image denoising, the model enhances the processing of feature information, thereby improving overall performance [38, 39].

It is worth noting that, although an explicit outlier ratio is not separately reported, the effects of outliers are implicitly reflected in the error statistics and depth profile analyses, where no abrupt error spikes or heavy-tailed distributions are observed (as shown in figures 12(h) and (i)), indicating effective suppression of large reconstruction deviations. Experiments on objects with different curvatures and local protrusions show consistent reconstruction accuracy, with MAEs remaining within 0.057 mm, demonstrating that the proposed method is robust to variations in surface geometry and material-induced fringe distortions. This stability is attributed to the complementary dual-head configuration and the diffusion-based fringe recovery, which preserves phase continuity while mitigating aliasing-related error amplification.

3.4. Limitations and scope

It should be noted that the experimental scope of this work is defined by the objective of achieving true simultaneous dual-head structured light measurement with locally overlapping FOV. While comparative experiments with existing methods are often desirable, most reported approaches for mitigating fringe aliasing in multi-projector systems rely on specific hardware configurations or projection strategies, such as color or orientation multiplexing, optical-path separation, or strictly controlled temporal projection sequences. These methods are not directly compatible with the proposed system configuration and would require fundamental modifications to the hardware architecture or acquisition protocol. Moreover, approaches that suppress aliasing by disabling simultaneous projection through time-division schemes fall outside the scope of this study. Therefore, the reconstruction results obtained from non-aliased single-projector measurements are used as a reference, while reconstructions directly derived from aliased fringe patterns using conventional PSP serve as the baseline. This evaluation strategy enables a clear and consistent assessment of the impact of fringe aliasing and the effectiveness of the proposed diffusion-based suppression method under identical measurement conditions.

4. Conclusion

In conclusion, this research presents a novel approach for enhancing 3D reconstruction using a dual-head structured light system combined with conditional diffusion-based deep learning framework. By addressing key challenges such as occlusion, aliasing, and limited measurement range, the proposed system significantly improves the accuracy, speed, and scalability of 3D reconstruction. The dual-camera and projector configuration effectively minimizes occlusion, while the deep learning-based method for separating aliased fringe patterns allows for faster and more precise measurements. Compared to existing methods that eliminate aliasing via hardware separation or sequential projection, which are sensitive to environmental variations and sacrifice synchronous measurement, the proposed dual-head FPSL combined with the conditional denoising diffusion model achieves efficient and high-precision 3D reconstruction through synchronous dual-camera and dual-projector acquisition. The diffusion model learns the mapping from aliased to non-aliased fringes directly from data, providing robustness against harsh lighting and complex surfaces. The integration of conditional diffusion modeling with structured light further enhances the system's capability to deliver high-precision reconstructions by leveraging diverse training data. Experimental results validate the effectiveness of this approach, demonstrating its potential for a wide range of applications that require fast, precise, and comprehensive 3D measurements. Overall, the proposed method offers a robust solution for overcoming the limitations of traditional single-camera systems and sets the stage for future advancements in high-quality 3D reconstruction techniques.

The dual-head structured light system demonstrates strong practical robustness, effectively accommodating moderate variations in ambient illumination and diverse object surface textures through conditional diffusion model-based adaptive fringe pre-processing, phase-shifting estimation, and multi-frequency heterodyne techniques. The overlapping regions between the two heads provide stable correspondences for accurate ICP-based alignment, enabling automatic correction of external parameters and thereby reducing the need for repeated full-system calibration. In addition, the proposed dual-head FPSL system is designed for straightforward setup and operation by users familiar with conventional structured light or PSP systems. System deployment involves mechanically mounting the two projector–camera units, synchronizing projection and acquisition via hardware triggers, and independently calibrating each projector–camera pair using standard calibration procedures identical to those of single-head systems. In practice, a one-time calibration is sufficient unless the hardware configuration changes. Together, these design features ensure reliable and accurate 3D reconstruction in realistic deployment scenarios, highlighting the system's suitability for complex measurement tasks.

Data availability statement

The data cannot be made publicly available upon publication because they contain commercially sensitive information. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (12572209, 62375078, 12002197), the National Key Laboratory of Ship Structural Safety (NAKLAS2024KF016-S), and Shanghai Municipal Science and Technology Program Project (24ZR1423100).

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

Lei Lu 0000-0002-3050-6542
Conceptualization (equal), Funding acquisition (equal), Methodology (equal), Project administration (equal), Validation (equal), Writing – original draft (equal)

Yingju Wang
Conceptualization (equal), Methodology (equal), Validation (equal), Visualization (equal), Writing – original draft (equal)

Yuejiao Guo 0009-0004-1551-8563
Methodology (equal), Validation (equal), Visualization (equal)

Zhilong Su 0000-0002-7389-6333
Conceptualization (equal), Funding acquisition (equal), Methodology (equal), Project administration (equal), Writing – review & editing (equal)

Wei Pan 0000-0002-0933-2453
Resources (equal), Software (equal), Validation (equal)

Deyang Zhang
Investigation (equal), Resources (equal)

Qinghui Zhang
Resources (equal)

Peng Li
Supervision (equal)

References

[1] Zhang W, Zhu S, Ran C, Zhang Q and Wang Y 2024 Large-depth-range 3D measurement based on optimized multi-focal projection system *Opt. Express* **32** 29573–86
[2] Xia Z, Fan Z, Mei X, Sun X, Yan J, Zhang J and Huang W 2025 Accurate reconstruction of turbine blade point cloud and multiple point cloud registration based on structured light *Measurement* **242** 115897
[3] Yang G and Wang Y 2022 Three-dimensional measurement of precise shaft parts based on line structured light and deep learning *Measurement* **191** 110837
[4] Shang H, Liu C and Wang R 2022 Measurement methods of 3D shape of large-scale complex surfaces based on computer vision: a review *Measurement* **197** 111302
[5] Wu H, Cao Y, Dai Y and Wei Z 2024 Orthogonal spatial binary coding method for high-speed 3D measurement *IEEE Trans. Image Process.* **33** 2703–13

[6] Marinov E, Martins R J, Youssef M A B, Kyrou C, Coulon P-M and Genevet P 2023 Overcoming the limitations of 3D sensors with wide field of view metasurface-enhanced scanning lidar *Adv. Photon.* **5** 046005
[7] Hu X, Xu W, Fan Q, Yue T, Yan F, Lu Y and Xu T 2024 Metasurface-based computational imaging: a review *Adv. Photon.* **6** 014002
[8] Wu H and Cao Y 2025 Complementary phase interleaving-based fringe order recognition for temporal phase unwrapping *Pattern Recognit.* **157** 110937
[9] Chen Y, Jiang W, Luo Z and Yang L 2024 A novel 3D reconstruction method with a binocular-line laser system *Measurement* **227** 114238
[10] Lu L, Bu C, Su Z, Guan B, Yu Q, Pan W and Zhang Q 2024 Generative deep-learning-embedded asynchronous structured light for three-dimensional imaging *Adv. Photon.* **6** 046004
[11] Geng J 2011 Structured-light 3D surface imaging: a tutorial *Adv. Opt. Photon.* **3** 128–60
[12] Zhang X, Miao Y, Tang Q, Cai Z, Chen Z, Liang A, Yin Y, Peng X and Liu X 2024 Large-scale-adaptive fringe projection 3D measurement *Opt. Lett.* **49** 4485–8
[13] Zhang S 2018 High-speed 3D shape measurement with structured light methods: a review *Opt. Lasers Eng.* **106** 119–31
[14] Wu H, Cao Y, Dai Y and Qin J 2025 Spatial–temporal 3-d directional binary coding method for fringe projection profilometry *IEEE Trans. Instrum. Meas.* **74** 1–11
[15] Qin J, Jiang Y and Cao Y 2025 Generalized multi-feature-guided progressive fringe order correction for fringe projection profilometry *IEEE Trans. Image Process.* **34** 5792–802
[16] Man D, Jones M W, Li D, Zhang H and Song Z 2019 Calibration of turntable based 3D scanning systems *IEICE Trans. Inf. Syst.* **E102.D** 1833–41
[17] Cheung G, Baker S and Kanade T 2003 Visual hull alignment and refinement across time: a 3D reconstruction algorithm combining shape-from-silhouette with stereo *2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2003. Proc.* vol 2 p II–375
[18] Wu J, Sang B and Zhao H 2001 Double projecting measurement method of 3D object based on color CCD *Laser Infrared* **31** 187–9
[19] Maimone A and Fuchs H 2012 Reducing interference between multiple structured light depth sensors using motion *2012 IEEE Virtual Reality Workshops (VRW)* pp 51–54
[20] Je C, Lee K H and Lee S W 2013 Multi-projector color structured-light vision *Signal Process. Image Commun.* **28** 1046–58
[21] Bai J, Qu X, Feng W and Zhang F 2018 Separation method of overlapping phase-shift gratings in three-dimensional measurement of double projection structured light *Acta Opt. Sin.* **38** 1112002
[22] Li Y, Qu X, Zhang F and Zhang Y 2020 Separation method of superimposed gratings in double-projector structured-light vision 3D measurement system *Opt. Commun.* **456** 124676
[23] Gorthi S S and Rastogi P 2010 Fringe projection techniques: whither we are? *Opt. Lasers Eng.* **48** 133–40
[24] Zhu S, Wu Z, Zhang J, Zhang Q and Wang Y 2022 Superfast and large-depth-range sinusoidal fringe generation for multi-dimensional information sensing *Photon. Res.* **10** 2590–8
[25] Zhang Z, Chang C, Liu X, Li Z, Shi Y, Gao N and Meng Z 2021 Phase measuring deflectometry for obtaining 3D shape of specular surface: a review of the state-of-the-art *Opt. Eng.*, **60** 020903
[26] Wu K, Xie Y, Lu L, Yin Y and Xi J 2023 Three-dimensional reconstruction of moving hdr object based on psp *Opt. Lasers Eng.* **163** 107451
[27] Cai B, Zi A, Tong C, Wu Q, Zhao B and Chen X 2025 Spatial coding strategy for dual-frequency phase-shifting profilometry *Measurement* **239** 115437
[28] Yao Y and Guo J 2012 The 3D optical measurement system based on Multi-frequency Heterodyne Principle *2012 7th Int. Conf. on Computing and Convergence Technology (ICCCT)* (*Seoul, Korea (South)*, *03–05 December 2012*) pp 1130–4
[29] Cao H, Tan C, Gao Z, Xu Y, Chen G, Heng P-A and Li S Z 2024 A survey on generative diffusion models *IEEE Trans. Knowl. Data Eng.* **36** 2814–30
[30] Feng S, Zuo C, Hu Y, Li Y and Chen Q 2021 Deep-learning-based fringe-pattern analysis with uncertainty estimation *Optica* **8** 1507–10
[31] Zhang Q, Liu F, Lu L, Su Z, Pan W and Dai X 2024 Reconstruction of transparent objects using phase shifting profilometry based on diffusion models *Opt. Express* **32** 13342–56
[32] Lu L, Guo Y, Su Z, Zhang Q, Zhang D and Li P 2024 Deep diffusion learning of mutual-reflective structured light patterns for multi-body three-dimensional imaging *Opt. Express* **32** 36171–84
[33] Mao S, He Y, Chen H, Zheng H, Liu J, Yuan Y, Le M, Li B, Chen J and Xu Z 2023 High-quality and high-diversity conditionally generative ghost imaging based on denoising diffusion probabilistic model *Opt. Express* **31** 25104–16
[34] Chu X, Tian Z, Zhang B, Wang X and Shen C 2021 Conditional positional encodings for vision transformers *Int. Conf. on Learning Representations* (available at: https://api.semanticscholar.org/CorpusID:256827775)
[35] Kazemnejad A, Padhi I, Ramamurthy K N, Das P and Reddy S 2023 The impact of positional encoding on length generalization in transformers *Neural Information Processing Systems* (available at: https://api.semanticscholar.org/CorpusID:258987259)
[36] Zhang Z 2000 A flexible new technique for camera calibration *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 1330–4
[37] Su Z, Xu Y, Qin X, Zhang D, Wang H and Lu L 2025 Automatic crack tracking with consistent stereo imaging and wavelet analysis *Int. J. Mech. Sci.* **304** 110662
[38] Shifei Ding L G J Z, Wang Q and Ding L 2024 A novel image denoising algorithm combining attention mechanism and residual unet network *Knowl. Inf. Syst.* **66** 581–611
[39] Gurrola-Ramos J, Dalmau O and Alarcón T E 2021 A residual dense u-net neural network for image denoising *IEEE Access* **9** 31742–54