Laser & Photonics Reviews

LASER & PHOTONICS REVIEWS

www.lpr-journal.org

RESEARCH ARTICLE

Temporal Super-Resolution for High-Speed 3D Imaging Using Triple-Frequency Color-Multiplexed Fringe Projection

Yifan Liu[1,2,3,4] | Wenwu Chen[1,2,3,4] | Shijie Feng[1,2,3,4] | Yutong Xiao[1,2,3,4] | Jinyang Jiang[1,2,3,4] | Shengqi Yu[1,2,3,4] | Yiheng Liu[1,2,3,4] | Wei Yin[1,2,3,4] | Qian Chen[1,2,3,4] | Chao Zuo[1,2,3,4]

[1]Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, China | [2]Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, Nanjing, Jiangsu, China | [3]Jiangsu Key Laboratory of Visual Sensing and Intelligent Perception, Nanjing, Jiangsu, China | [4]State Key Laboratory of Extreme Environment Optoelectronic Dynamic Measurement Technology and Instrument, Taiyuan, Shanxi, China

Correspondence: Shijie Feng (shijiefeng@njust.edu.cn) | Qian Chen (chenqian@njust.edu.cn) | Chao Zuo (zuochao@njust.edu.cn)

Received: 29 October 2025 | **Revised:** 28 January 2026 | **Accepted:** 15 April 2026

Keywords: 3D imaging | color multiplexing | deep learning | fringe projection profilometry | temporal super-resolution

ABSTRACT

Recent advances in artificial intelligence have enabled temporal super-resolution fringe projection profilometry (FPP) to overcome the inherent frame-rate limitations of image sensors through frequency-domain multiplexing, thereby achieving high-speed 3D imaging with low-frame-rate cameras. In recent studies, researchers have significantly enhanced frequency-domain multiplexing efficiency by compressing multiple frequencies of fringe patterns into a grayscale image, enabling 3D imaging with higher temporal resolution. However, this approach relies on frequency-domain compression, which limits further improvements in compression ratio and restricts the flexibility in selecting fringe pattern frequencies. To address this limitation, we introduce, for the first time, the integration of color-channel multiplexing with frequency-domain multiplexing, proposing a triple-frequency color-multiplexed FPP (TFCMFPP). In this approach, multi-directional fringe patterns at three distinct frequencies are encoded into the projector's RGB channels. Under high-speed projection, a single long-exposure image can be acquired by a low-speed color camera to encapsulate the multiplexed fringes, thus enabling one-shot dense information packing. To decode the multiplexed patterns, we developed a three-stage deep neural network comprising zero-order removal, spatial-frequency decoupling, and direction-aware expert modules. By combining this architecture with projection distance minimization (PDM) based phase unwrapping, we can robustly recover multi-directional absolute phases without relying on stereo or temporal priors. This allows 3D reconstruction at up to 24× the native device speed. We validate the proposed method on a variety of transient scenes, including collapsing pillars, rotating rotor-engine models, and multi-object rotational scenarios. Experiments demonstrate that TFCMFPP achieves an effective 3D reconstruction rate approaching 5,000 Hz using only a 208 Hz color camera, while maintaining high spatial fidelity, low depth error, and excellent robustness to motion. With its low cost, compactness, and scalability, TFCMFPP integrates color, frequency, direction, and temporal dimensions into a unified coding framework, opening new avenues for high-throughput 3D imaging.

Yifan Liu and Wenwu Chen contributed equally to this work.

1 | Introduction

The genesis of three-dimensional vision can be traced to the 19th century, when stereoscopes first harnessed binocular disparity to evoke compelling illusions of depth [1]. This early fascination with perceptual dimensionality gradually evolved into rigorous optical metrology through successive breakthroughs in interferometry, holography, and structured illumination–transforming visual artifice into quantifiable, traceable measurement [2–6]. Today, advanced 3D measurement stands as a cornerstone technology, enabling non-contact geometric capture with exceptional density and accuracy in demanding applications such as precision manufacturing, autonomous robotics, biomechanics, and structural health monitoring [7, 8]. These domains increasingly demand the ability to reconstruct the surface morphology on ultrashort time scales without compromising resolution [9, 10].

Among various 3D measurement techniques, fringe projection profilometry (FPP) has become a powerful tool for dynamic 3D imaging due to its advantages of high accuracy, non-contact measurement, and robustness to surface texture [11]. Nevertheless, when applied to dynamic or transient events, FPP faces fundamental speed limitations imposed by: (1) hardware, namely the frame rates of projectors and cameras, and (2) algorithm, specifically the number of fringe patterns needed for reliable 3D reconstruction [12].

On the hardware side, binary defocusing has been introduced to accelerate projection. By slightly defocusing the projection lens, one-bit binary patterns can generate quasi-sinusoidal fringes, leveraging the native binary mode of digital micromirror devices (DMDs) to reach projection rates of tens of kHz [13, 14]. On the algorithmic side, various fringe encoding and decoding schemes, including dual-frequency phase shifting [15], 2+2 phase shifting [16], bi-frequency phase shifting [17], composite phase shifting with geometric constraints [18], speckle-embedded Fourier transform methods [19], and micro Fourier transform profilometry (μFTP) [20], have been developed to reduce coding redundancy and minimize the number of required patterns. Despite these advances, recovering high-quality absolute phase information from a single fringe image remains an enduring challenge.

More recently, deep learning (DL) has brought a paradigm shift to optical metrology [21–25]. By automatically extracting features from large datasets, DL provides data-driven solutions for complex measurement tasks, and has been successfully applied to fringe denoising, phase retrieval, and unwrapping [26, 27]. DL-based methods have even demonstrated the ability to reconstruct the absolute phase directly from a single fringe image, when combined with color encoding [28], geometric priors [29], or frequency multiplexing [30]. Nevertheless, traditional FPP still requires a strict one-to-one synchronization between fringe projection and image acquisition, which fundamentally constrains the imaging temporal resolution to the native sensor frame rate, thereby limiting its applicability to high-speed dynamic scenarios. Most current solutions to improve imaging speed rely heavily on high-refresh-rate hardware, which inevitably increases system complexity and cost [32].

To circumvent hardware bottlenecks, compressive imaging approaches such as compressed ultrafast photography (CUP) [32], coded aperture compressive temporal imaging (CACTI) [32], and the programmable pixel compressive camera (P2C2) [33] have been proposed, building on advances in compressive sensing [34–36] and computational imaging [37, 38]. While effective for 2D imaging, these systems rely on complex optical modulation and inevitably suffer photon loss, reducing signal-to-noise ratios and limiting their practicality for 3D metrology. In the context of FPP, researchers recently introduced deep-learning-enabled multiplexed FPP (DLMFPP) [39], which overlaps multiple directional fringe patterns in a single long exposure and employs neural networks for demultiplexing and phase recovery. DLMFPP achieved up to 9× temporal super-resolution, enabling kHz-rate 3D reconstruction with cameras operating at only 111 Hz. Building on this concept, dual-frequency angular-multiplexed FPP (DFAMFPP) [40] incorporated dual-frequency fringes at different orientations alongside a physics-guided neural network. This approach achieved 16× temporal super-resolution, reaching 10 kHz 3D imaging with a single 625 Hz camera. However, these methods confine all directional fringe patterns to a single image channel. Consequently, as the compression ratio increases, spectral aliasing and error accumulation become severe [39, 40]. This fundamentally limits their applicability to complex or rapidly changing scenes.

To break through this bottleneck, we extend multiplexing into multiple spectral channels and propose triple-frequency color-multiplexed FPP (TFCMFPP). In this method, multi-directional fringes at three distinct frequencies are encoded into the projector's RGB channels [41]. Under high-speed projection, a low-speed color camera records a single long-exposure image containing all multiplexed fringes, achieving joint channel–direction–frequency encoding. The method exploits projection-distance-minimization (PDM) frequency combinations and channel orthogonality to suppress aliasing. For robust demultiplexing and high-accuracy phase reconstruction, we design a multi-stage deep neural network pipeline: zero-order and color-crosstalk removal network (ZOCR-Net) for zero-order and crosstalk removal, spatial–frequency decomposition network (SFD-Net) for direction/frequency decomposition, and fringe direction decoding network (FDD-Net) with a gating mechanism for accurate phase prediction and unwrapping. Without requiring stereo or temporal priors, TFCMFPP robustly reconstructs 24 absolute phase maps from a single image, achieving 24× temporal super-resolution. Experiments on dynamic scenes–including collapsing pillars, rotating rotor-engine models, and multi-object rotational scenarios–demonstrate that TFCMFPP achieves an effective 3D reconstruction rate of 5000 Hz using only a 208 Hz color camera. By advancing from spatial-domain multiplexing to joint color–direction–frequency encoding, TFCMFPP establishes a new paradigm for high-speed, low-cost, and scalable 3D vision systems.

2 | Methods

2.1 | Color-Multiplexed Fringe Pattern Design and Projection

Fringe projection profilometry (FPP) is a representative structured-light 3D imaging technique [42]. Its basic principle is to project sinusoidal fringe patterns onto the surface of the

measured object using a projector [43], while a camera captures the deformed fringe images from different viewpoints after modulation by the object's 3D geometry. The height variations on the object surface induce spatial phase changes in the fringes, which manifest as distortions in the captured images. Ideally, the sinusoidal fringe pattern projected onto the object surface can be expressed as:

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

Here, (x^p, y^p) are the pixel coordinates of the projected image, a^p denotes the background intensity, b^p represents the fringe modulation, λ denotes the fringe period. After interaction with the surface, geometry is encoded in the fringe phase; phase demodulation yields a wrapped phase, which is unwrapped to the absolute phase and mapped via a calibrated phase-to-height model to metric depth for full-field 3D reconstruction.

In this work, to overcome the compression bottleneck of single-channel spatial-domain multiplexing, we develop a TFCMFPP system. It employs multiple groups of directional fringe patterns, with each group consisting of three RGB fringes at different spatial frequencies. A high-speed projector projects these patterns onto the scene at a rate exceeding the native frame rate of the camera. The camera, operating in long-exposure mode, records the multiplexed image, thereby achieving temporal multiplexing. Figure 1 illustrates the basic workflow of the TFCMFPP method. In this approach, the projector sequentially projects M groups of RGB fringes along different directions, where the red, green, and blue channels are respectively encoded with fringes of different spatial frequencies. The specifically designed fringe pattern I_m^p can be expressed as:

$$I_m^p(x^p, y^p) = \begin{cases} a + b\cos\left(\frac{2\pi x^p\cos\theta_n}{\lambda_x^R} + \frac{2\pi y^p\sin\theta_n}{\lambda_y^R}\right), \\ \qquad m = 3n-2, \text{ for red channel} \\ a + b\cos\left(\frac{2\pi x^p\cos\theta_n}{\lambda_x^G} + \frac{2\pi y^p\sin\theta_n}{\lambda_y^G}\right), \\ \qquad m = 3n-1, \text{ for green channel} \\ a + b\cos\left(\frac{2\pi x^p\cos\theta_n}{\lambda_x^B} + \frac{2\pi y^p\sin\theta_n}{\lambda_y^B}\right), \\ \qquad m = 3n, \text{ for blue channel} \end{cases} \tag{2}$$

Here, $a = b = 0.5$ denote the average intensity and modulation. The subscript m represents the index of the projected fringe pattern, where $m = 1, 2, \ldots, 3M$. λ_x^R and λ_y^R are the periods of the red low-frequency fringes in the x and y directions, λ_x^B and λ_y^B are those of the blue medium-frequency fringes, and λ_x^G and λ_y^G are those of the green high-frequency fringes. The variable n denotes the index of the projection angle ($n = 1, 2, \ldots, M$), and θ_n represents the orientation of the fringes, defined as:

$$\theta_n = (-1)^{n+1}\left(\frac{n}{2} - \frac{(-1)^n + 1}{4}\right)\theta_0 \tag{3}$$

where θ_0 is a scalar characterizing the specific angular orientation of the fringe, constrained by the carrier frequency and the maximum surface gradient to ensure effective spectral separation (see Section S6 for the design principles of θ_0). To ensure unambiguous phase unwrapping, we adopt the design principle of the projection distance minimization (PDM) [44] in designing the fringe frequencies. Specifically, the fringe periods in the RGB channels should satisfy:

$$\mathrm{LCM}(\lambda_x^R, \lambda_x^G, \lambda_x^B) \geq W, \quad \mathrm{LCM}(\lambda_y^R, \lambda_y^G, \lambda_y^B) \geq H \tag{4}$$

where $W \times H$ denotes the resolution of the projected pattern, and $\mathrm{LCM}(\cdot)$ represents the least common multiple. When the fringe pattern I_m^p is projected onto the object surface, the deformed fringe intensity captured by the camera can be expressed as:

$$I_m^c(x, y) = A_m(x, y) + B_m(x, y)\cos[\phi_m(x, y)] \tag{5}$$

where $c \in \{R, G, B\}$. Specifically, $c = R$ corresponds to $m = 3n - 2$, $c = G$ corresponds to $m = 3n - 1$, and $c = B$ corresponds to $m = 3n$, with $n = 1, 2, \ldots, M$. Here, (x, y) are the pixel coordinates in the captured image, $A_m(x, y)$ and $B_m(x, y)$ denote the average intensity and modulation of the captured fringes, and $\phi_m(x, y)$ is the modulation phase.

To encode temporal information into a single frame, we employ a long-exposure acquisition strategy in which a total of $3M$ RGB fringe patterns $I_m(x, y)$ are temporally integrated into a single image. At each pixel location (x, y), the long-exposure color fringe image I_{LE} is constructed by summing the corresponding time-series frames across the R/G/B channels and stacking them into a single RGB vector. The equation for I_{LE} is shown in Equation (6).

$$I_{LE}(x, y) = \left(\sum_{n=1}^{M} I_{3n-2}^R(x, y),\ \sum_{n=1}^{M} I_{3n-1}^G(x, y),\ \sum_{n=1}^{M} I_{3n}^B(x, y)\right)^\top \tag{6}$$

The resulting color-multiplexed image $I_{LE}(x, y)$ encapsulates the 3D scene information over $3M$ time instants. Subsequently, a 2D Fourier transform is applied independently to each color channel of the long-exposure image, revealing their respective frequency-domain representations. Since the R/G/B channels are encoded with distinct spatial frequencies, the resulting frequency spectra manifest as concentric ring-shaped bands when visualized together. These bands are spatially distributed across orientations and frequency radii, enabling effective separation of directional components, as illustrated by F_{LE} in Figure 1.

It is worth noting that although the three frequencies share the same orientation, the difference in their frequencies ensures that the corresponding fringe patterns satisfy the projection distance minimization principle, providing sufficient ambiguity-free support for phase unwrapping and thus ensuring accurate phase reconstruction. At the same time, this approach avoids the frequency range conflict encountered in conventional stereo phase unwrapping methods. Consequently, a specifically trained three-stage deep neural network can be employed to decode the color-multiplexed image and recover $3M$ high-precision absolute phase maps, enabling dense and robust 3D reconstruction.

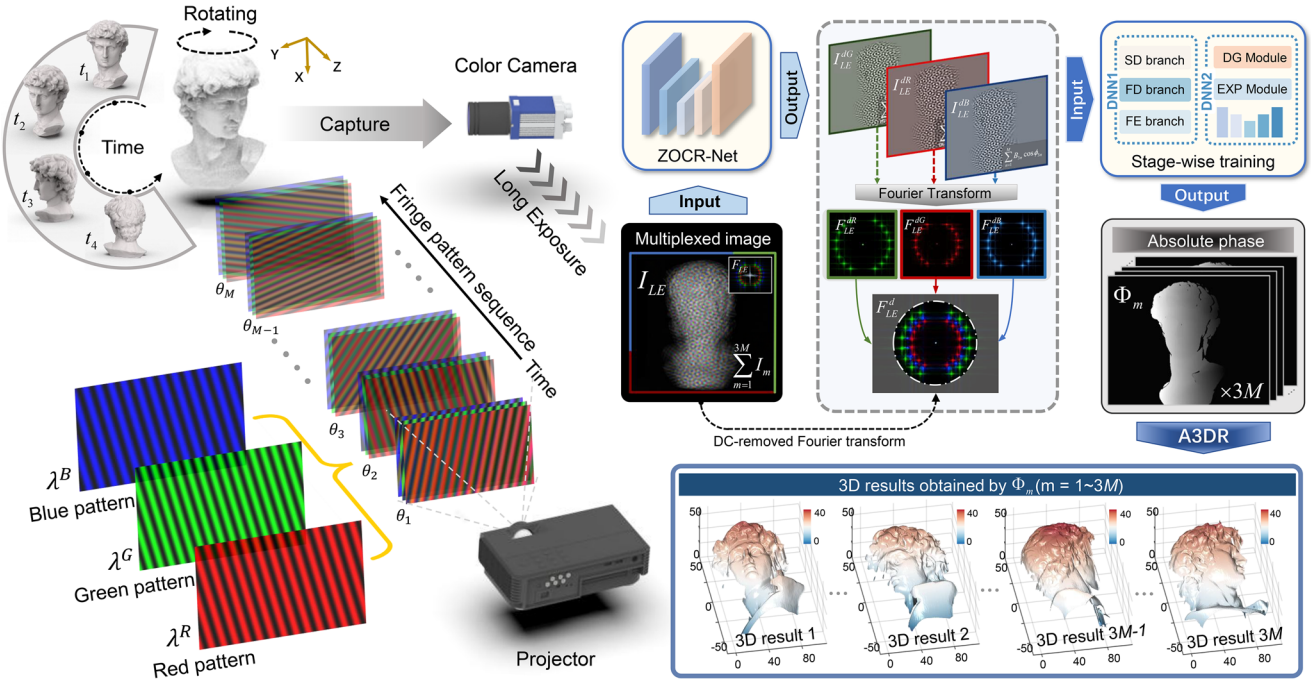

FIGURE 1 | Overall workflow of the proposed TFCMFPP method. Over a single acquisition window, the moving object is illuminated by a high-speed sequence of triple-frequency directional fringes, where red, green, and blue patterns of different frequencies are sequentially projected and encoded into the RGB channels. A color camera operating in long-exposure mode captures a single color-multiplexed image that fuses information from multiple time segments. This image is then fed into a decoding framework consisting of zero-order crosstalk removal, ensemble learning, and expert network modules. Through spectral analysis, directional fringe patterns are extracted, and multiple absolute phase maps Φ_m are obtained, which are subsequently used by the Augmented 3D Reconstruction (A3DR) algorithm to accomplish multi-frame dynamic 3D reconstruction.

2.2 | Analysis of Color-Multiplexed Fringe Image

To recover temporally resolved 3D information from the color-multiplexed image I_{LE}, the proposed TFCMFPP employs a multi-stage deep neural network architecture, as illustrated in Figure 2. This framework consists of three major steps. Step 1 involves a zero-order and crosstalk removal module, which is designed to suppress zero-order components and remove color crosstalk. Step 2 employs a spatial–frequency demultiplexing network that accurately decodes directional fringe patterns with high fidelity. Step 3 incorporates a set of direction-aware expert networks that robustly recovers absolute phase maps.

Step 1: To extract clean and directional fringe signals from the color-multiplexed image, the input image is first separated into three color channels, each containing directional fringes. However, due to the overlap of zero-order components in the frequency domain, severe spectral crosstalk occurs, particularly when dealing with complex surfaces. In addition, the three channels of a color camera may introduce inter-channel interference, causing mixing of originally independent directional fringes and further degrading the accuracy of the subsequent reconstruction.

To suppress the zero-order background components and inter-channel crosstalk in the fringe images, we introduce ZOCR-Net, the data without these factors are used as labels to train this network. ZOCR-Net is built upon a TriPath-CNN [45] architecture that leverages three parallel perceptual branches to extract multi-scale spatial–frequency features. Furthermore, a squeeze-and-excitation (SE) module is incorporated to adaptively recalibrate channel information and enhance feature representation. More details about the TriPath-CNN architecture can be found in Supplementary Information Section 1. The network is trained in a supervised manner, and the outputs are the three zero-order and crosstalk removed channel images $\{I_{LE}^{dR}, I_{LE}^{dG}, I_{LE}^{dB}\}$, expressed as:

$$I_{LE}^{dc} = \sum_{n=1}^{M} B_{3n-\delta_c}(x,y)\cos\phi_{3n-\delta_c}(x,y), \quad \delta_c \in \{\delta_R, \delta_G, \delta_B\} \tag{7}$$

where $\delta_R = 2$, $\delta_G = 1$, $\delta_B = 0$, $c \in \{R, G, B\}$, and $n = 1, 2, \ldots, M$. The combined image of these three channels is denoted as I_{LE}^{d}. Subsequently, the zero-order and crosstalk removed images are transformed into the frequency domain by Fourier transform [46]:

$$F_{LE}^{dc} = \mathcal{F}(I_{LE}^{dc}) \tag{8}$$

where $\mathcal{F}(\cdot)$ denotes the two-dimensional Fourier Transform of the input image. The resulting spectra $\{F_{LE}^{dR}, F_{LE}^{dG}, F_{LE}^{dB}\}$ exhibit concentric circular distributions, in which the zero-order components are effectively suppressed. This provides the foundation for the subsequent stable demultiplexing of the composite fringe images.

Step 2: To further extract directional fringes from the multiplexed images, we design SFD-Net. This network takes as input the pre-processed images and their spectra, and employs spatial, frequency, and ensemble branches to achieve complementary feature learning and enhancement [39, 47]. The spatial branch processes image features and aggregates edge

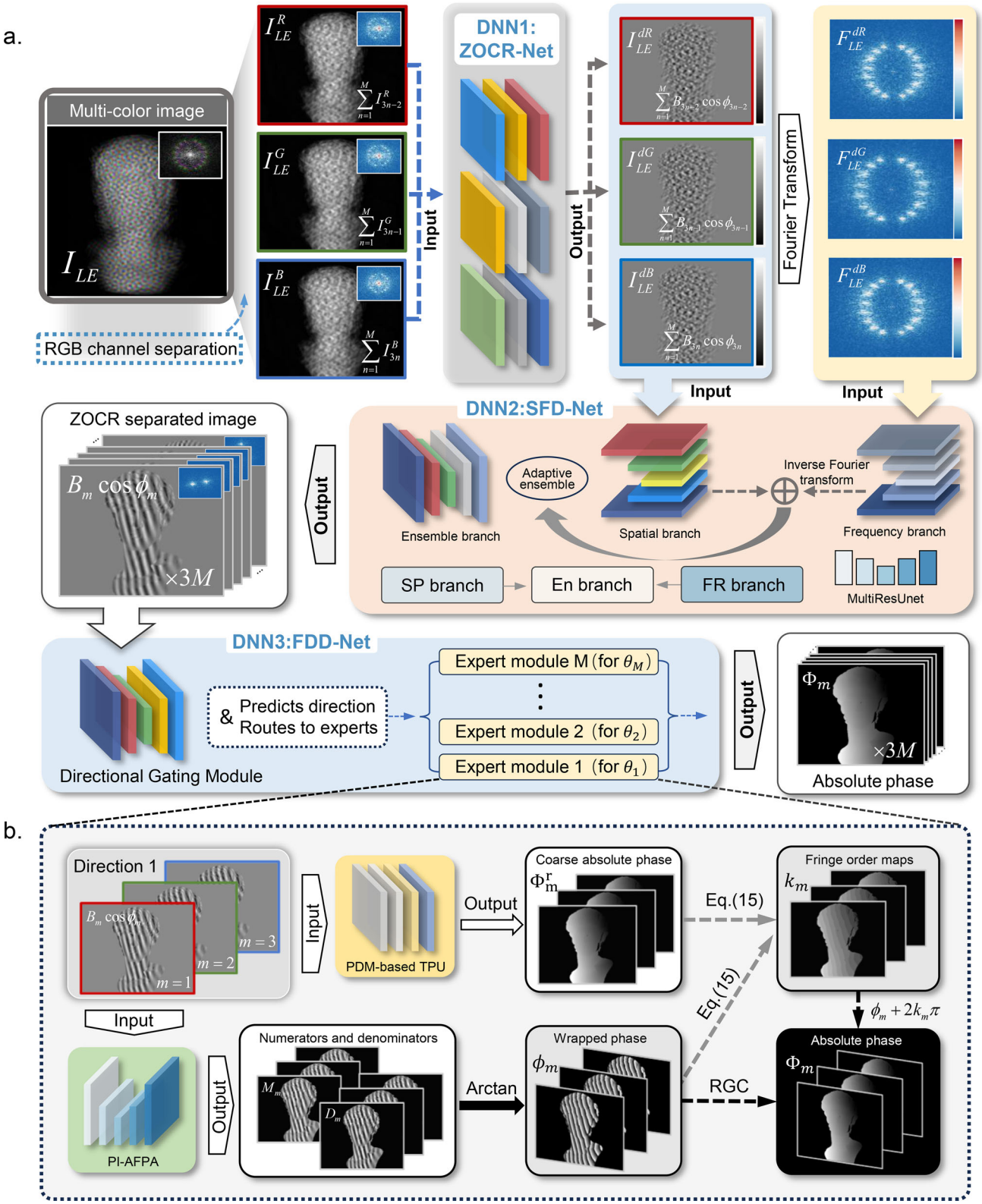

FIGURE 2 | Deep learning architecture for phase recovery in the proposed TFCMFPP method. (a) The input color-multiplexed fringe image is separated into three RGB channels and pre-processed by ZOCR-Net to remove zero-order components and color crosstalk. The pre-processed images and their spectra are then simultaneously fed into a neural network composed of spatial, frequency, and ensemble branches, which reconstructs 3*M* clear directional fringe images. These images are subsequently assigned to orientation-specific expert sub-networks for phase recovery, yielding 3*M* absolute phase maps. (b) Architecture of an expert sub-network. For three directional fringe images with the same orientation, two sub-branches are used: PI-AFPA for wrapped phase estimation and PDM-based TPU for coarse absolute phase estimation. The final absolute phase map is obtained by combining the two results, while the introduced RGC algorithm further improves accuracy in edge regions.

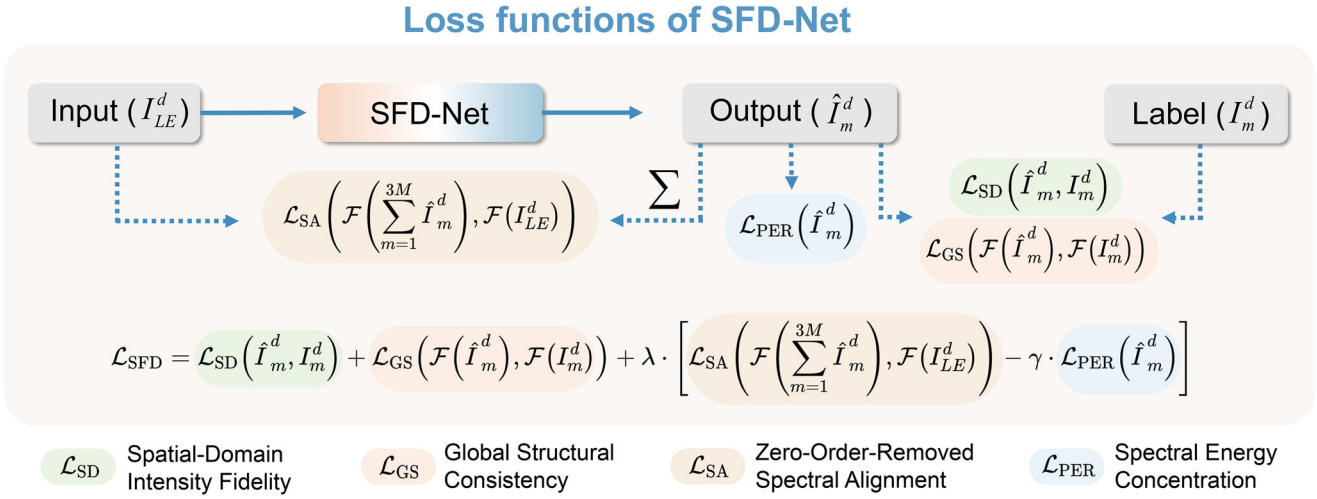

FIGURE 3 | Loss functions of SFD-Net. As defined in Equation (9), the data-based terms ($\mathcal{L}_{SD}$ and $\mathcal{L}_{GS}$) are employed to ensure fidelity to the ground truth labels. The physics-based terms ($\mathcal{L}_{SA}$ and $\mathcal{L}_{PER}$) are introduced to enforce spectral self-consistency and energy concentration constraints.

information of directional fringes; the frequency branch processes spectral inputs and performs multi-directional frequency decomposition, capturing bandwidth and harmonic structures; the ensemble branch employs a learnable attention mechanism to integrate spatial and frequency-domain features, thereby improving demodulation accuracy. The entire network adopts MultiResUNet [48] as its backbone (see Section S1 for the detailed MultiResUnet structure), with the three branches sharing the encoder and decoder. In each branch, multi-scale residual blocks are introduced to balance global and local features. Finally, SFD-Net outputs a sequence of directional fringes, which are fed into the subsequent phase estimation module. To ensure physical consistency in directional selection, we define the following combined loss function $\mathcal{L}_{SFD}$. Figure 3 illustrates the overall design of the joint spatial–spectral loss functions in SFD-Net, which is formulated as:

$$\begin{aligned}\mathcal{L}_{SFD} &= \mathcal{L}_{SD}(\hat{I}^d_m, I^d_m) + \mathcal{L}_{GS}(\mathcal{F}(\hat{I}^d_m), \mathcal{F}(I^d_m)) \\ &+ \lambda\cdot\left[\mathcal{L}_{SA}(\mathcal{F}(\sum_{m=1}^{3M}\hat{I}^d_m), \mathcal{F}(I^d_{LE})) - \gamma\cdot\mathcal{L}_{PER}(\hat{I}^d_m)\right] \quad (9)\end{aligned}$$

Specifically, the proposed objective function comprises four key components: Spatial-Domain Intensity Fidelity ($\mathcal{L}_{SD}$), Global Structural Consistency ($\mathcal{L}_{GS}$), Zero-Order-Removed Spectral Alignment ($\mathcal{L}_{SA}$), and Spectral Energy Concentration ($\mathcal{L}_{PER}$). The mathematical definitions and detailed physical significance of these constituent terms are elaborated below:

Spatial-Domain Intensity Fidelity ($\mathcal{L}_{SD}(\hat{I}^d_m, I^d_m)$):

$$\mathcal{L}_{SD}(\hat{I}^d_m, I^d_m) = \frac{1}{3M}\sum_{m=1}^{3M}\left[\frac{1}{N}\sum_{x,y}\|\hat{I}^d_m(x,y) - I^d_m(x,y)\|^2\right] \quad (10)$$

where N denotes the total number of pixels, and (x, y) represents the spatial coordinates. This term serves as the foundational constraint for network training. By minimizing the L_2 Euclidean distance (Mean Squared Error) between the predicted fringe patterns $\hat{I}^d_m$ and the ground truth I^d_m, it ensures that the generated images achieve pixel-level numerical convergence.

Global Structural Consistency ($\mathcal{L}_{GS}(\mathcal{F}(\hat{I}^d_m), \mathcal{F}(I^d_m))$):

$$\begin{aligned}&\mathcal{L}_{GS}(\mathcal{F}(\hat{I}^d_m), \mathcal{F}(I^d_m)) \\ &= \frac{1}{3M}\sum_{m=1}^{3M}\left[\frac{1}{N}\sum_{u,v}|\mathcal{F}(\hat{I}^d_m)(u,v) - \mathcal{F}(I^d_m)(u,v)|\right] \quad (11)\end{aligned}$$

While spatial loss ensures intensity matching, it often fails to capture high-frequency structural details, leading to over-smoothed fringes. To address this, we introduce the Global Structural Consistency loss in the frequency domain. $\mathcal{F}(\cdot)$ denotes the Discrete Fourier Transform (DFT). By regularizing the L_1 distance between the complex-valued Fourier spectra of the prediction and the ground truth, this term explicitly penalizes spectral deviations. Since fringe patterns are fundamentally periodic signals defined by specific spatial frequencies, this loss preserves the integrity of the carrier frequency and orientation, preventing structural distortions and blurring artifacts.

Zero-Order–Removed Spectral Alignment ($\mathcal{L}_{SA}(\mathcal{F}(\sum_{m=1}^{3M}\hat{I}^d_m), \mathcal{F}(I^d_{LE}))$):

$$\begin{aligned}&\mathcal{L}_{SA}\left(\mathcal{F}\left(\sum_{m=1}^{3M}\hat{I}^d_m\right), \mathcal{F}(I^d_{LE})\right) \\ &= \frac{1}{N}\sum_{u,v}\left|\mathcal{F}\left(\sum_{m=1}^{3M}\hat{I}^d_m\right)(u,v) - \mathcal{F}(I^d_{LE})(u,v)\right| \quad (12)\end{aligned}$$

where (u, v) denote the spatial frequency coordinates. Ideally, the sum of the demodulated dynamic fringes should mathematically correspond to the captured raw image after background removal. This "self-consistency" constraint enforces spectral alignment between the aggregated predicted fringes and the zero-order-removed multiplexed image I^d_{LE}. It acts as a physical regularization that prevents the network from generating artifacts

that might satisfy individual channel constraints but violate the overall energy conservation of the imaging system.

Spectral Energy Concentration ($\mathcal{L}_{\mathrm{PER}}(\hat{I}_m^d)$):

$$\mathcal{L}_{\mathrm{PER}}(\hat{I}_m^d) = \frac{1}{3M}\sum_{m=1}^{3M}\left[\frac{\iint_{\Omega_{peak}} |\mathcal{F}(\hat{I}_m^d)(u,v)|^2 du dv}{\iint_{\Omega_{all}} |\mathcal{F}(\hat{I}_m^d)(u,v)|^2 du dv}\right] \tag{13}$$

Here, Ω_{peak} denotes the frequency neighborhood centered on the fundamental carrier, and Ω_{all} represents the entire frequency spectrum. By optimizing this ratio, the loss function imposes a constraint on the spectral energy distribution, penalizing energy dispersion outside the carrier frequency band. This mechanism serves to mitigate spectral leakage and harmonic interference, thereby preserving the fundamental sinusoidal components necessary for phase demodulation.

Step 3: To avoid parameter redundancy and unstable predictions caused by directly decoding all directional fringes simultaneously, we introduce FDD-Net. This framework enables accurate and robust phase recovery for fringe patterns with different directions and frequencies by employing a set of direction-aware expert networks.

Specifically, the system first applies a directional gating module to the fringe images decoded in Step 2, producing a direction-aware weight vector $\mathbf{w} \in \mathbb{R}^{M\times 1}$. The directional gating module is designed based on the Gating-Net [49, 50] architecture (see Supplementary information, Section 1 for the detailed Gating-Net structure), which consists of multiple 3×3 convolutional layers and max-pooling encoders to extract directional features from the input fringes. Each element of the $\mathbf{w}$ represents the likelihood that the current input fringe belongs to one of M candidate orientations. This weight vector guides the expert network assignment for precise phase decoding. Based on the maximum element in the weight vector $\mathbf{w}$, each directional fringe is routed to its corresponding expert network. Each expert network (Expert Module) consists of two functional sub-branches:

(1) **PI-AFPA branch** (Phase Inference via Attention-Fused Phase Approximation): This branch specializes in wrapped phase estimation. The PI-AFPA is built upon a U-Net [51] architecture (see Section S1 for the detailed U-Net structure), where the input is the three-frequency fringes along a specific direction. The network emulates the Fourier Transform Profilometry (FTP) and phase-shifting decoding processes, separately predicting the numerator $M_m(x,y)$ and denominator $D_m(x,y)$ of the phase term [26], and computes the wrapped phase as:

$$\phi_m(x,y) = \arctan\left(\frac{M_m(x,y)}{D_m(x,y)}\right) \tag{14}$$

By embedding physical priors into the network [27, 52, 53], PI-AFPA ensures accurate and reliable wrapped phase estimation, providing a strong foundation for robust 3D reconstruction.

(2) **PDM-based TPU branch** (Projection Distance Minimization based Temporal Phase Unwrapping): To achieve coarse absolute phase recovery, we introduce a temporal unwrapping strategy based on the projection distance minimization (PDM) [44]. A ResUNet-based deep network [54] is trained to learn the phase difference maps from the three-frequency directional fringes, outputting the rough absolute phase $\Phi_m^r(x,y)$. This rough phase serves as auxiliary guidance for subsequent unwrapping. The final integer fringe order $k_m(x,y)$ is determined by combining the rough absolute phase with the wrapped phase $\phi_m(x,y)$ [28] as:

$$k_m(x,y) = \mathrm{Round}\left(\frac{\Phi_m^r(x,y) - \phi_m(x,y)}{2\pi}\right) \tag{15}$$

Finally, the accurate absolute phase map $\Phi_m(x,y)$ is computed as:

$$\Phi_m(x,y) = \phi_m(x,y) + 2\pi \cdot k_m(x,y) \tag{16}$$

With the aid of the physical prior mechanism based on the PDM principle, this module achieves precise prediction of absolute phase, further enhancing the stability and accuracy of 3D reconstruction in complex scenes.

By constructing M parallel expert networks combined with the controller, the proposed TFCMFPP method enables fine-grained decoding of directional and frequency-dependent fringe information, ultimately recovering a sequence of $3M$ absolute phase maps from the multiplexed color fringe image. To improve robustness against weak textures and noise, TFCMFPP further incorporates the reliability-guided compensation (RGC) algorithm, which refines and corrects wrapped phase boundaries, yielding structurally continuous and low-error phase maps. Together with the system calibration and the augmented 3D reconstruction (A3DR) algorithm (see Section S3 for the detailed augmented 3D reconstruction algorithm), the absolute phase maps Φ_m are converted into depth maps, achieving temporally super-resolved, high-accuracy, and large-range dynamic 3D reconstruction. The proposed method achieves an effective reconstruction speed 3M times higher than the camera's native frame rate, enabling significant temporal super-resolution.

3 | Results

3.1 | Experimental Setup

To validate the effectiveness of the proposed method, we built the experimental system illustrated in Figure 1. The system consists of a Basler acA640-750uc color camera and a DMD-based projection unit. The projector employs a LightCrafter 4500 module (Texas Instruments, USA) equipped with a 912 × 1140 resolution DMD, integrated into a custom-designed optical path. For fringe projection along a given direction, the number of groups was set to $M = 8$. Based on the theoretical derivation detailed in Section S6, we selected a universal angular interval of $\theta_0 = 18°$ for all experiments. This configuration provides a spectral guard band and ensures robustness by avoiding the singular vertical frequency axis. The fringe periods were configured as: $\lambda_x^R = 23$, $\lambda_x^B = 21$, $\lambda_x^G = 18$, $\lambda_y^R = 18$, $\lambda_y^B = 15$, $\lambda_y^G = 13$. The least common multiples are calculated as: $\mathrm{LCM}(\lambda_x^R, \lambda_x^G, \lambda_x^B) = 2898$, $\mathrm{LCM}(\lambda_y^R, \lambda_y^G, \lambda_y^B) = 1170$, leading to an effective unambiguous region of 2898×1170 pixels.

Operating in binary (1-bit) mode, the DMD can be tuned to achieve a refresh rate exceeding 5,000 Hz. During experiments, the projector sequentially projects 24 directional fringe patterns, which a single trigger signal prompts the camera to capture a color-multiplexed fringe image containing information from 24 temporal instants. The DLP hardware ensures accurate synchronization for reliable signal acquisition.

For system calibration, we adopted a projector–camera calibration strategy based on the inverse camera model, which unifies the phase coordinates of fringes from different directions into a common 3D reconstruction coordinate system, thereby establishing the quantitative relationship between absolute phase and spatial coordinates. For more details on projector–camera calibration strategy, please refer to Section S2.

For network training, all models were implemented in the PyTorch framework and trained on an NVIDIA RTX 4090 GPU. This hardware also facilitates efficient deployment, the typical inference time for reconstructing a full 3D sequence from a single color-multiplexed image is approximately 90 ms. The networks were trained in a transfer learning manner: pre-trained on the digital-twin dataset and fine-tuned on real captured data. In the pre-training stage, 1,000 groups of synthetic datasets were generated in Blender, including 800 groups for training and 200 groups for validation (see Note S4 for details on synthetic data generation). The hyperparameters λ and γ are set to 0.1 and 0.05, respectively. The Adam optimizer was used, with an initial learning rate of 1×10^{-4}, a maximum batch size of 2, and 250 training epochs. A learning rate decay schedule was applied: if the validation loss did not decrease for 10 consecutive epochs, the learning rate was automatically reduced to further optimize training. Based on the pretrained models, the networks were fine-tuned with real captured datasets. The real dataset included a wide range of objects (e.g., human faces, stone, metal, ceramics, etc.), ensuring that the trained networks achieved robust generalization and strong adaptability to real-world dynamic scenarios.

3.2 | Non-Periodic Motion Scene Test

To evaluate the temporal super-resolution capability of the proposed TFCMFPP method in non-periodic dynamic 3D reconstruction, we designed a structural collapse experiment. The experimental scene consists of a collapsing ionic column (falling at approximately 25 ms) and a static Venus statue, both characterized by complex geometry and rich surface details. The projection system sequentially projects RGB fringe patterns along eight preset orientations ($M = 8$) at 960 Hz, while the color camera operates at 40 Hz in long-exposure mode. As a result, each captured image encodes information from 24 temporal instants, enabling single-shot dynamic aggregation.

As shown in Figure 4a, the captured color-multiplexed fringe image I_{LE} is first processed by ZOCR-Net to remove zero-order background and inter-channel crosstalk, yielding clean images I_{LE}^{d}. The three RGB channels are further separated into I_{LE}^{dR}, I_{LE}^{dG}, and I_{LE}^{dB}, from which 24 directional fringes $B_m \cos \phi_m$ are extracted via directional demultiplexing, serving as inputs for subsequent phase recovery. Figure 4b presents the 24 absolute phase maps Φ_m reconstructed from a single multiplexed image, clearly revealing the temporal evolution of the collapsing structure. In each map, the background color indicates the corresponding input channel. The TFCMFPP framework adopts a gated mixture-of-experts strategy to assign each directional fringe to its corresponding expert sub-network, ensuring high accuracy and consistency with ground truth.

To assess phase recovery accuracy, three representative fringes ($m = 1, 14, 21$) were selected under two static poses for evaluation. Each fringe corresponds to a specific frequency–direction pair. As illustrated in Figure 4c, the phase recovery results of FTP, DLMFPP, and TFCMFPP were quantitatively compared against high-quality ground-truth phase maps. The results show that for $m = 1, 14, 21$, the MAE of TFCMFPP is 0.0641 rad, 0.0668 rad, and 0.0682 rad, markedly lower than FTP (average MAE: 0.8986 rad). Moreover, by incorporating a preprocessing mechanism and a gated expert network architecture, TFCMFPP further refines the phase reconstruction pipeline, leading to a substantial accuracy improvement over DLMFPP (average MAE: 0.1356 rad) and demonstrating superior adaptability and robustness in complex imaging scenarios.

Figure 4d further demonstrates the 3D reconstruction results at four representative instants ($m = 1, 9, 17, 24$; corresponding to $t = 0$, 8.3, 16.7, and 23.9 ms). Despite the complex structural geometry and rapid collapse dynamics, TFCMFPP successfully reconstructs the complete collapse process with fine structural details faithfully preserved. This experiment confirms that TFCMFPP achieves temporal super-resolution equivalent to 960 Hz 3D reconstruction using only a 40 Hz color camera.

3.3 | High-Speed Dynamic Scene Test of a Single Object

To further evaluate the performance of TFCMFPP in periodic dynamic scenarios, we conducted a high-speed measurement experiment on a rotating engine rotor model. As shown in Figure 5a, the engine model has a size of approximately 200 mm × 197 mm and a complex internal structure consisting of a rotor and a stator. To ensure stable rotation at high speed, the experimental rotor model was further reinforced. Using a 60 Hz color camera operating in long-exposure mode, the system captured images at 24× temporal super-resolution, achieving an effective 3D reconstruction frame rate equivalent to 1,440 Hz.

To reconstruct the periodic motion of the rotor, we collected six color-multiplexed images. Figure 5b shows one captured image and its processed version after ZOCR-Net, where the zero-order and crosstalk are effectively suppressed. Based on these six images, 144 directional fringes were decoded. Figure 5c illustrates six representative directional fringes (corresponding to $m = 1, 29, 60, 88, 116, 144$).

The decoded fringes were then fed into FDD-Net for phase recovery. Figure 5d presents 3D reconstruction results at six representative time instants, showing the continuous motion and deformation of the rotor during periodic rotation. To analyze internal motion characteristics, depth cross-sections along the combustion chamber (X–Z plane) were extracted, as shown in Figure 5e. At $t = 41.65$ ms and $t = 99.97$ ms, the spatial

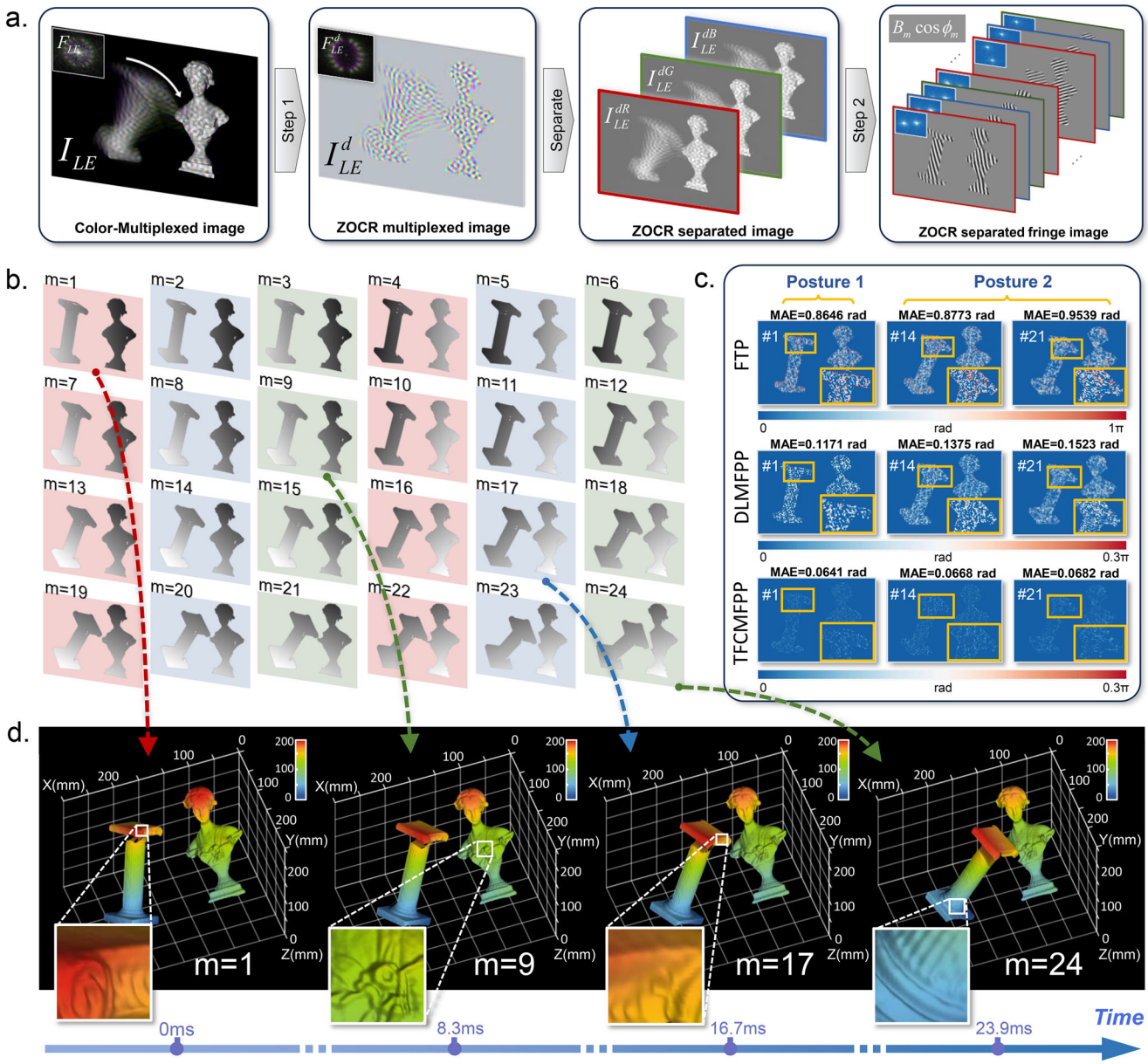

FIGURE 4 | 3D reconstruction results of dynamic scenes involving a collapsing ionic column and a static Venus statue. (a) Illustration of the image decoding workflow. From left to right: the captured color-multiplexed image I_{LE}, the pre-processed images $I_{LE}^{dR}, I_{LE}^{dG}, I_{LE}^{dB}$ after ZOCR-Net, and the 24 directional fringes $B_m \cos \phi_m$ obtained after demodulation. (b) Temporal evolution of the object surface reconstructed from 24 absolute phase maps Φ_m by TFCMFPP. The background color corresponds to the original RGB channels (red, green, or blue). (c) Comparison of three selected fringes ($m = 1, 14, 21$) under two static poses (pose 1 and pose 2), evaluated by the mean absolute error (MAE) between the reconstructed phase maps of FTP, DLMFPP, and TFCMFPP and the ground-truth high-quality reference. (d) 3D reconstruction results at four time instants (t=0, 8.3, 16.7 and 23.9 ms), with enlarged insets highlighting detailed structures and embedded temporal information.

displacement of the rotor tip is clearly observed, revealing subtle periodic variations that are otherwise difficult to measure.

By analyzing the 3D trajectory of the rotor tip, the rotation period was calculated to be 50.98 ms, corresponding to a rotational speed of approximately 1,177 RPM. The complete process of TFCMFPP and 3D results of the entire dynamic turbine engine process is provided in Visualization S1. This further verifies the reliability of the proposed method for high-speed dynamic measurements in periodic scenarios.

3.4 | Dynamic Scene Test With Multiple Moving Objects

To further demonstrate the 3D reconstruction capability of TFCMFPP in multi-object dynamic scenarios, we designed a scene consisting of three high-speed rotating targets: a high-speed turbofan in the center, a high-speed four-blade fan on the right, and an exhaust fan at the lower left. Under long-exposure acquisition, the exhaust fan rotated slowly, appearing nearly static, while the other two targets remained in non-synchronized

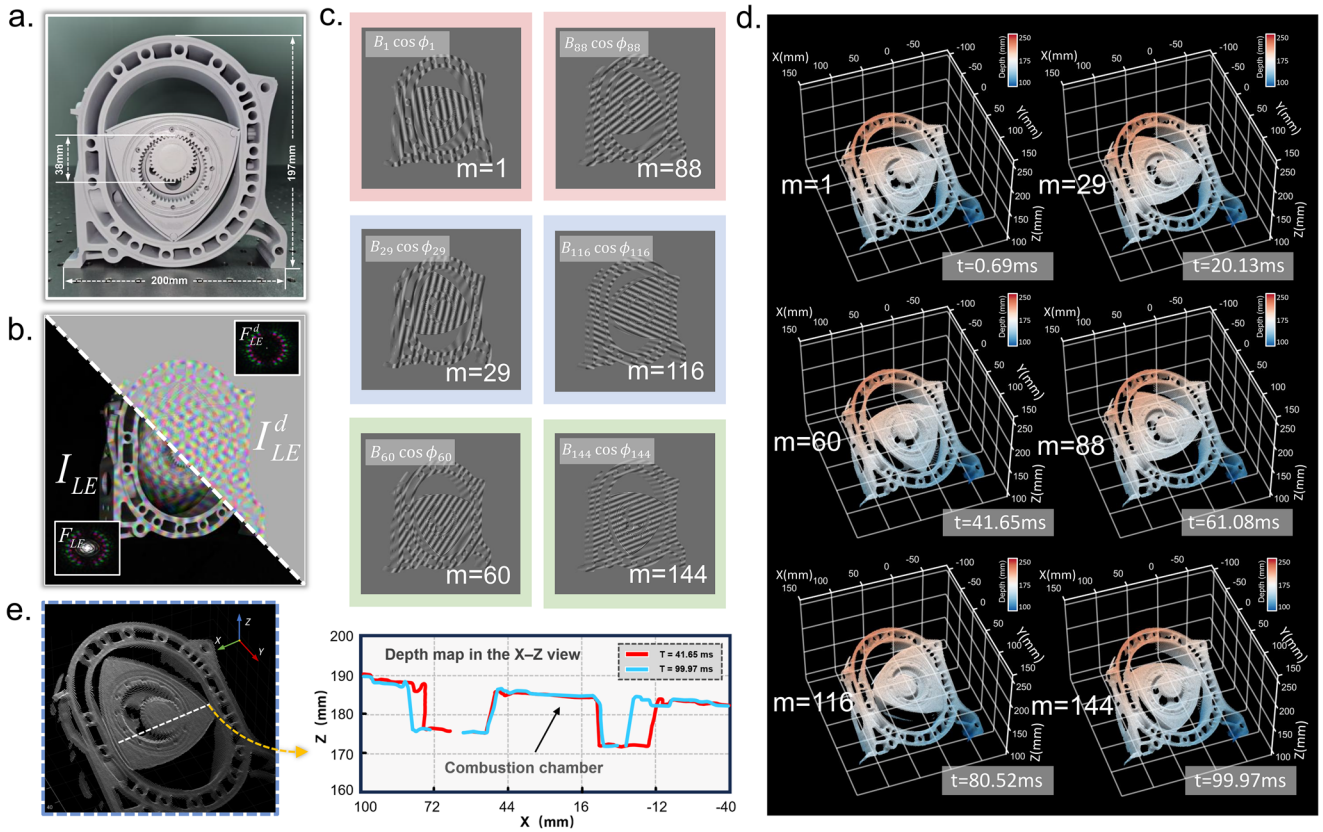

FIGURE 5 | High-speed 3D dynamic measurement results of the rotating engine model. (a) Photograph of the rotating engine model with dimensional annotations. (b) Captured long-exposure color-multiplexed fringe image I_{LE} and its processed version I^d_{LE} after ZOCR-Net, and the corresponding spectral distribution is also shown. (c) Six representative directional fringes $B_m \cos \phi_m$ decoded from I^d_{LE}, with fringe indices $m = 1, 29, 60, 88, 116, 144$. The colored boxes indicate the RGB channel origin of each fringe. (d) 3D reconstruction results at six different time instants, revealing key structural deformations during rotor rotation. (e) Cross-sectional depth maps along the X–Z plane of the combustion chamber at $t = 41.65$ ms and $t = 99.97$ ms, demonstrating the spatial displacement process of the rotor tip.

dynamic motion. The experiment was performed using a 208 Hz color camera, which captured a single long-exposure color-multiplexed image. Combined with the proposed TFCMFPP decoding framework, 24 directional fringe patterns at different frequencies were extracted, enabling effective 3D reconstruction at 5,000 Hz.

As shown in Figure 6a, the original fringe image I_{LE} and its processed version I^d_{LE} after ZOCR-Net demonstrate effective suppression of zero-order background interference while preserving high-frequency directional fringe components. The corresponding spectral distributions further confirm the presence of clear directional frequency bands. From I^d_{LE}, twenty-four representative fringes $B_m \cos \phi_m$ were decoded across eight orientations, serving as continuous high-temporal-resolution inputs for reconstruction (Figure 6b). Figure 6c illustrates the 3D reconstruction results at five representative instants, successfully capturing the motion pose of all three targets within a single multiplexed frame.

To analyze the turbofan motion in greater detail, three reference points (A, B, and C) were selected on the blade tips (Figure 6d), and their displacement trajectories along the Z axis were tracked over time. As shown in Figure 6e, the displacement curves exhibited clear periodic oscillations, with stable periods, consistent phases, and amplitudes–demonstrating reliable periodic motion of the high-speed blades. Figure 6f shows the 3D reconstruction results of the four-blade fan and exhaust fan at two representative instants ($T = 0$ ms and $T = 4.2$ ms), where structural details and surface textures are well preserved. Figure 6g further illustrates cross-sectional depth maps of the four-blade fan blades at multiple instants, clearly reflecting the smooth temporal evolution of surface morphology.

By tracking the temporal displacement trajectories of selected reference points on the fan blades across consecutive frames, we performed motion frequency analysis to estimate the rotation periods. The turbofan and four-blade fan exhibited rotation periods of 12.6 and 19.2 ms, respectively, corresponding to rotational speeds of approximately 4,760 and 3,125 RPM. This further verifies the accuracy and reliability of the proposed method in multi-object dynamic scenarios. These results demonstrate that TFCMFPP can achieve temporally super-resolved 3D reconstruction of multiple unsynchronized moving targets from a single multiplexed image.

3.5 | Accuracy Analysis

To quantitatively evaluate the measurement accuracy of the proposed TFCMFPP method, we conducted a standard sphere calibration experiment. Two ceramic spheres (A and B) with radii of 25.3967 and 25.3989 mm, respectively, and a center-to-

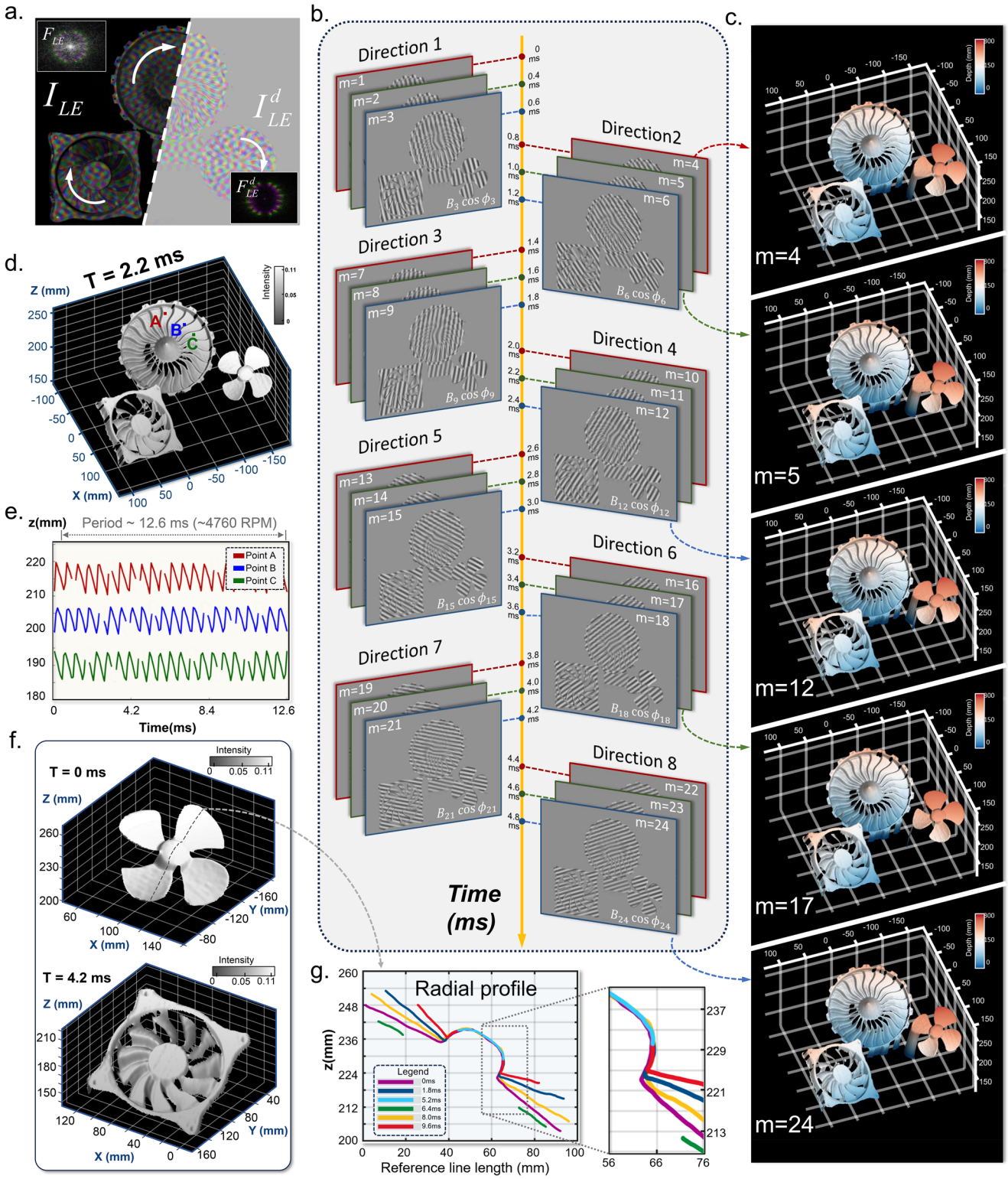

FIGURE 6 | Dynamic 3D measurement results of a multi-object rotating scene. (a) Original captured color-multiplexed fringe image I_{LE}, its processed version I^d_{LE} after ZOCR-Net, and the corresponding spatial spectra F_{LE} and F^d_{LE}. (b) Twenty-four directional fringes $B_m \cos \phi_m$ decoded from I^d_{LE}, grouped by direction within a 4.8 ms exposure window. (c) 3D reconstruction results at five representative time instants ($m = 4, 5, 12, 17, 24$), with color coding used to visualize the rotation of three groups of blades. (d) Complete 3D reconstruction at $T = 2.2$ ms, with three reference points (A, B, C) marked on the blade tip. (e) Temporal displacement curves of the three points (A–C) along the Z direction, revealing periodic rotational motion. (f) 3D reconstructions of the four-blade fan and exhaust fan at $T = 0$ ms and $T = 4.2$ ms. (g) Cross-sectional depth maps of the four fan blades at multiple time instants.

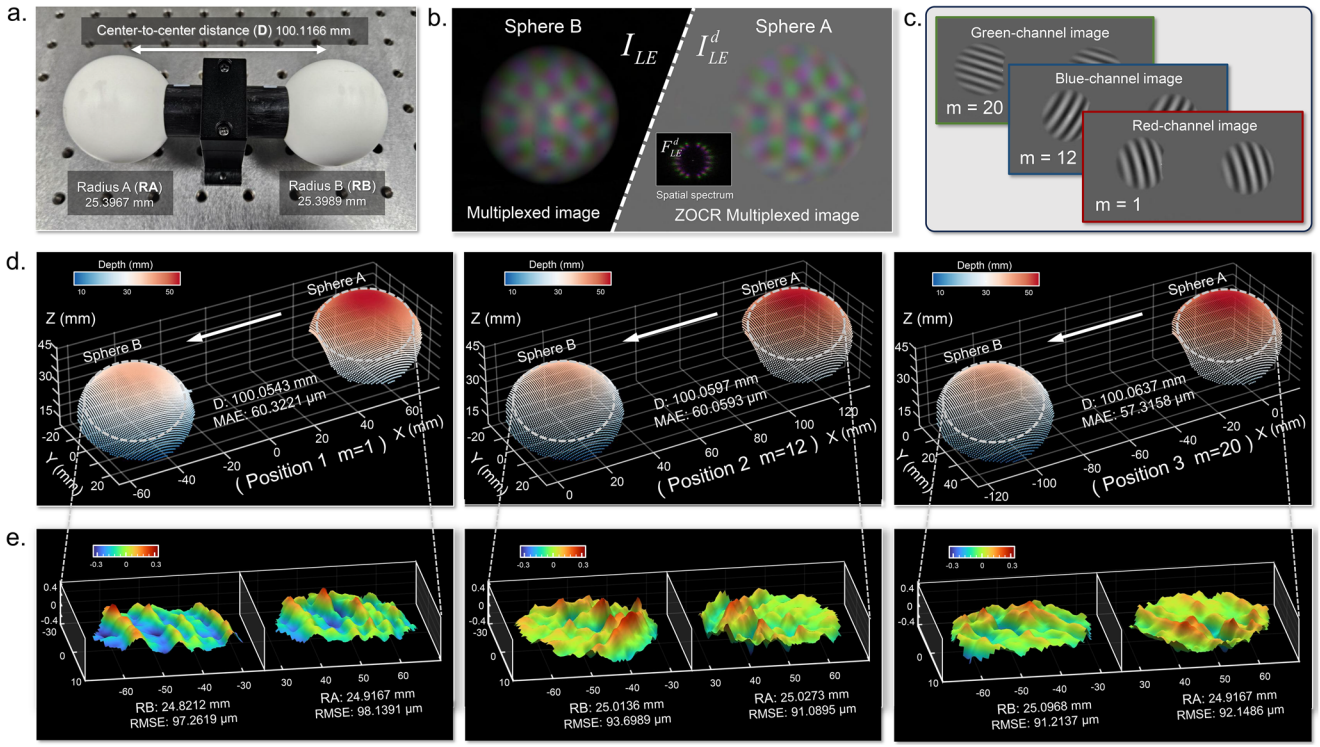

FIGURE 7 | Verification of 3D measurement accuracy using standard ceramic spheres. (a) Two calibration ceramic spheres (A and B), with radii and center-to-center distance measured by a coordinate measuring machine (CMM). The certified values are: $R_A = 25.3967$ mm, $R_B = 25.3989$ mm, and $D = 100.1166$ mm. (b) Captured color-multiplexed fringe image I_{LE}, the pre-processed result I^d_{LE} after ZOCR-Net, and the corresponding spatial spectra, with an enlarged view highlighting frequency separation. (c) Three representative directional fringes decoded from I_{LE}, corresponding to $m = 1$ (red channel), $m = 12$ (blue channel), and $m = 20$ (green channel). (d) 3D reconstruction of the two spheres from the three decoded fringes, with the measured center-to-center distance and mean absolute errors (MAEs) of the reconstructed radii annotated. (e) Residual error maps computed by best-fit sphere fitting, showing the root mean square error (RMSE) and MAEs of the sphere radii and center distance, providing an overall evaluation of the geometric accuracy of the reconstructed 3D results.

center distance of 100.1166 mm were used as calibration objects, as illustrated in Figure 7a. All dimensions were measured by a high-precision coordinate measuring machine (CMM) and served as ground-truth references for accuracy validation.

A single color-multiplexed fringe image was acquired, followed by ZOCR-Net preprocessing to remove zero-order components and crosstalk (Figure 7b). Three directional fringe components from the red, green, and blue channels (corresponding to $m = 1, 12, 20$) were extracted and used for phase demodulation and 3D reconstruction, as shown in Figure 7c. Figure 7d illustrates the reconstructed 3D surfaces of the two spheres at these three instants, with their radii and center distance annotated. The mean absolute errors (MAEs) were also computed for quantitative evaluation.

Furthermore, best-fit sphere fitting was applied to each reconstruction to calculate the residual error maps, from which the root mean square errors (RMSEs) were derived. The 3D visualization of these error maps is shown in Figure 7e. Specifically, under $m = 1$ (red channel), the reconstructed radii of spheres A and B were 24.9167 and 24.8212 mm, with RMSEs of 98.14 and 97.26 μm, and a center distance MAE of 60.32 μm. Under $m = 12$ (blue channel), the reconstructed radii were 25.0273 and 25.0136 mm, with RMSEs of 91.09 and 93.70 μm, and a center distance MAE of 60.06 μm. Under $m = 20$ (green channel), the reconstructed radii were 24.9167 mm and 25.0968 mm, with RMSEs of 92.15 and 91.21 μm, and a center distance MAE of 57.32 μm. These results demonstrate that TFCMFPP achieves surface-fitting RMSEs below 100 μm under single-frame conditions, highlighting its strong potential for high-precision metrology applications.

3.6 | Comparative Analysis

To assess the advantages of the proposed TFCMFPP framework, we conducted a comparative study against two representative state-of-the-art (SOTA) deep-learning methods: deep-learning-enabled multiplexed fringe projection profilometry (DLMFPP) [39] and deep-learning-based color fringe projection profilometry (DLCFPP) [28]. We adopted a "Divide-and-Conquer" benchmarking strategy. This approach selects the SOTA method to perform controlled comparisons, ensuring a comprehensive evaluation of both demultiplexing quality and phase retrieval accuracy.

To rigorously evaluate the demultiplexing capability, we established a benchmark against the network architecture of DLMFPP. We retrained the DLMFPP network on our color-multiplexed dataset to ensure domain adaptation. For strict fairness, both the DLMFPP network and our proposed decomposition module were fed an identical dataset. We then processed the output fringes from both methods using the same standard phase retrieval back-

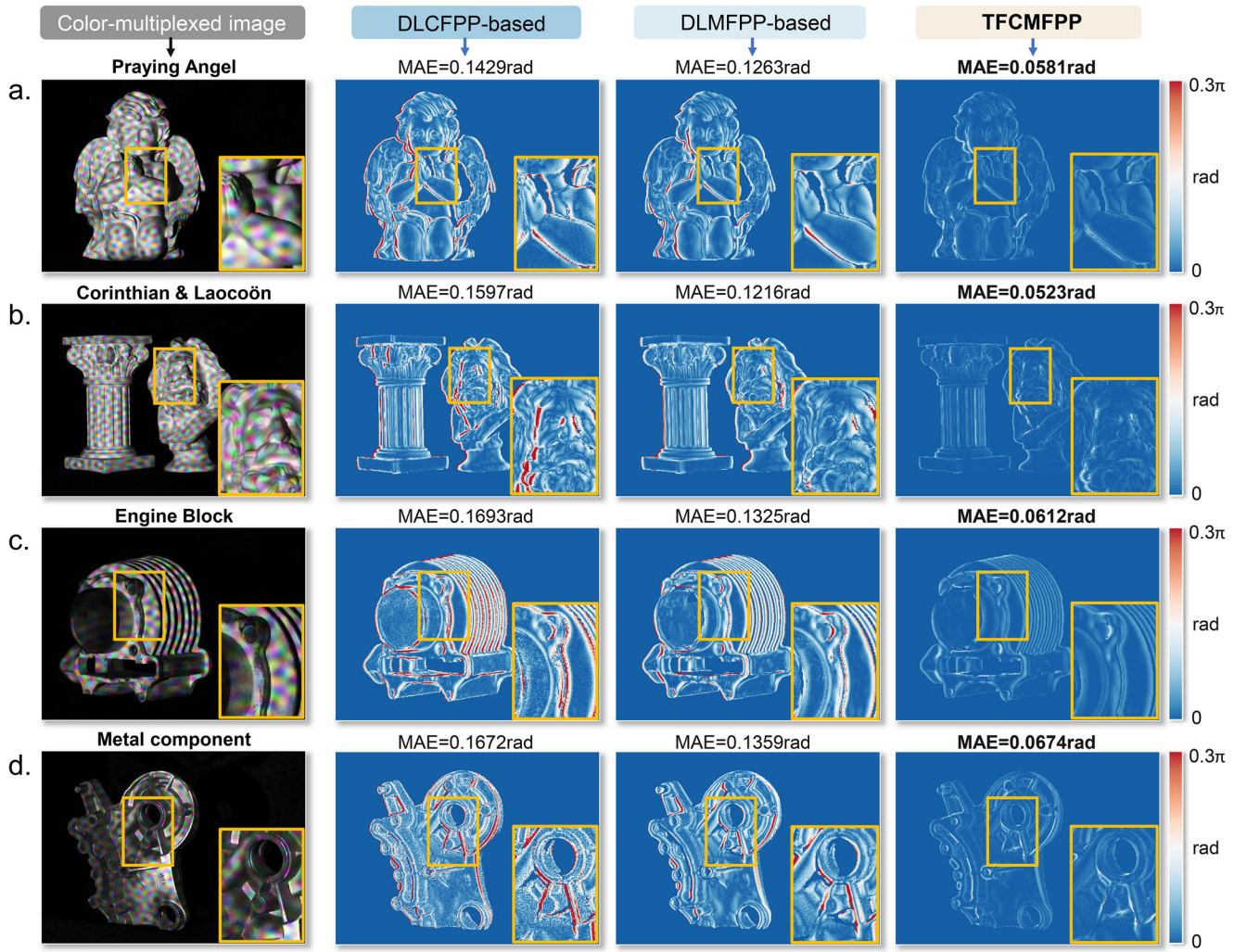

FIGURE 8 | Comparison with deep-learning baselines. (a–d), Absolute phase error maps for four test objects with different geometries and materials: Praying Angel (plaster, a), Corinthian & Laocoön (resin, b), Engine Block (aluminum alloy, c), and Metal component (industrial metal, d). Columns (left to right) show the single-shot color-multiplexed input image, results from the DLCFPP-based baseline, the DLMFPP-based baseline, and the proposed TFCMFPP. Yellow boxes indicate regions enlarged in the insets. Color scale: 0–0.3π rad; MAE is reported above each panel.

end. This fixed back-end approach isolates the demultiplexing quality. We then calculated the error between the final recovered absolute phase and the ground truth.

To evaluate the back-end phase analysis, we benchmarked our physics-embedded module against DLCFPP. As the original DLCFPP cannot decode multiplexed signals, we adapted the setup for fairness. We provided both the retrained DLCFPP network and our analysis module with identical high-quality demultiplexed fringes (generated by our front-end). This setup strictly isolates the phase retrieval stage. It allows a direct comparison between our analytical decoding and the data-driven regression approach using the same intermediate data. We then evaluated the accuracy by comparing the recovered absolute phase against the ground truth.

The comparative results, visualized in Figure 8, demonstrate the consistent performance of the proposed TFCMFPP framework across diverse test scenarios. From the absolute phase error maps in Figure 8, it can be observed that the DLCFPP-based and DLMFPP-based methods exhibit high-error bands along depth discontinuities, whereas our proposed method achieves superior performance. As shown in the specific metrics above each panel, our method consistently yields the lowest Mean Absolute Error (MAE). For the complex geometry of the *Praying Angel* (Figure 8a), TFCMFPP reduces the error to 0.0581 rad, surpassing the best baseline (DLMFPP, 0.1263 rad) by over 54%. For the *Engine Block* (Figure 8c), our method achieves an MAE of 0.0612 rad, whereas the DLCFPP and DLMFPP baselines lag behind at 0.1693 rad and 0.1325 rad, respectively. Similar trends are observed for the *Metal component* (Figure 8d), where TFCMFPP effectively halves the reconstruction error of the decomposition baseline (approximately 50% reduction), lowering the MAE from 0.1359 to 0.0674 rad. These results confirm that TFCMFPP maintains consistent, high-fidelity 3D reconstruction performance across diverse measurement scenarios.

3.7 | Ablation Study

The proposed TFCMFPP framework incorporates a Zero-Order and Crosstalk Removal Network (ZOCR-Net) to purify the input

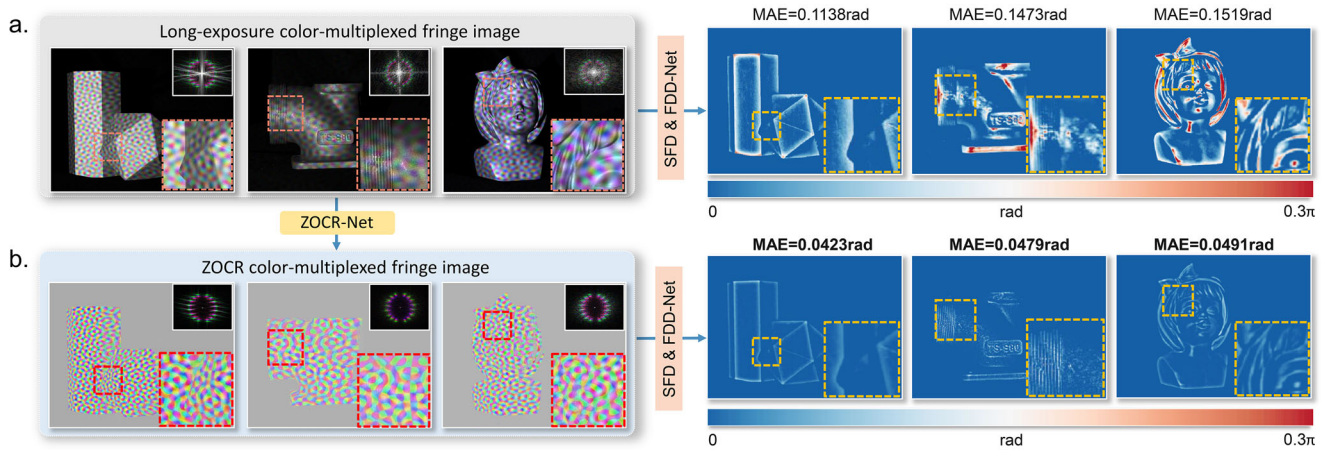

FIGURE 9 | Ablation study of ZOCR-Net. (a) Baseline without ZOCR-Net: raw long-exposure color-multiplexed images (left) and the corresponding phase error maps after reconstruction (right). (b) With ZOCR-Net: ZOCR-processed color-multiplexed images (left) and the corresponding error maps (right). From left to right: white geometric object, dark-gray mechanical part, and blue sculpture; MAE values are labeled in each panel.

single-shot images before phase demodulation. To rigorously quantify the contribution of this module to the final 3D reconstruction accuracy, we conducted an ablation study comparing the performance of the full pipeline against a baseline configuration. In the baseline setup, the ZOCR-Net module was removed. Consequently, the raw captured long-exposure color-multiplexed images, which contain high-intensity zero-order background and inter-channel crosstalk, were directly fed into the subsequent superimposed fringe decomposition network (SFD-Net) and frequency direction decoding network (FDD-Net). The evaluation metric is the Mean Absolute Error (MAE) calculated between the reconstructed absolute phase and the high-precision ground truth.

As illustrated in Figure 9, the qualitative and quantitative comparisons demonstrate the critical necessity of the ZOCR-Net. Visually, the error maps of the baseline method (Figure 9a) exhibit high-amplitude error artifacts distributed broadly across the object surfaces, indicating that the subsequent networks cannot fully decouple the phase information from the strong background and spectral interference inherent in the raw data. In contrast, the proposed method (Figure 9b) yields significantly smoother error maps with artifacts largely suppressed. Quantitatively, the inclusion of ZOCR-Net results in a consistent and substantial reduction in phase error across all test objects. Specifically, the MAE decreased from 0.1138 to 0.0423 rad for the white geometric object, from 0.1473 to 0.0479 rad for the dark-gray mechanical part, and from 0.1519 to 0.0491 rad for the blue sculpture. Overall, the average MAE across the three objects decreased from 0.1377 to 0.0464 rad, corresponding to a 66.3% reduction. These results confirm that explicit suppression of zero-order and crosstalk components is a prerequisite for achieving high-precision 3D reconstruction in the proposed single-shot framework.

4 | Discussion and Conclusion

In this work, we proposed a novel 3D reconstruction framework termed TFCMFPP (Triple-Frequency Color-Multiplexed Fringe Projection Profilometry), which integrates multi-dimensional information encoding with deep neural network decoding. By embedding "color–direction–frequency–time" four-dimensional multiplexed information into a single color-multiplexed image, the method significantly enhances the compression capability of single-shot structured light sequences while maintaining high information throughput. The proposed multi-stage neural framework employs a progressive architecture that sequentially refines feature representation from preprocessing and frequency separation to direction-aware phase recovery, ensuring high fidelity and robustness in 3D reconstruction. TFCMFPP achieves 3D reconstruction rates exceeding the native frame rate of low-speed cameras by more than an order of magnitude, reaching the kHz level, while still maintaining excellent spatial fidelity and temporal resolution in complex, high-speed dynamic multi-object scenarios. Experimental results demonstrate that the system, using only a 208 Hz color camera, achieved an effective 5,000 Hz 3D reconstruction rate, corresponding to 24× temporal super-resolution. Furthermore, validation with standard ceramic spheres confirmed surface-fitting RMSEs below 100 μm under single-frame reconstruction conditions, highlighting its potential for high-precision metrology, structural dynamics monitoring, and other advanced applications. With its low-cost hardware requirements and capability for single-camera, and temporally super-resolved 3D reconstruction, TFCMFPP establishes a robust bridge between physical encoding and deep-learning-based decoding. This provides a new technical pathway for the future development of high-speed 3D visual sensing systems.

Although TFCMFPP significantly improves the compression ratio by introducing an RGB channel–based frequency-domain multiplexing strategy, the approach has inherent physical limitations. First, the Bayer pattern of color sensors and inter-channel crosstalk may impose slight constraints on spatial resolution and signal fidelity, though these are largely mitigated by our deep-learning-based preprocessing. More importantly, there are upper limits to the temporal super-resolution factor, pushing the compression ratio significantly beyond the current 24× may introduce bottlenecks in frequency-domain pattern separability and network stability due to severe spectral aliasing, which remains a key challenge for future optimization.

Looking ahead, TFCMFPP can be further enhanced through advances in both optical hardware and computational modeling. On the hardware side, combining the proposed color multiplexing with polarization multiplexing or adopting high-speed DMD and LED-array projectors could further expand the system's acquisition speed and robustness under complex illumination. On the algorithmic side, integrating physics-informed neural networks (PINNs) or all-optical neural networks may enable faster and more interpretable phase recovery, paving the way toward real-time, high-speed, and high-fidelity 3D reconstruction. Overall, TFCMFPP establishes a unified framework that bridges physical encoding and intelligent decoding, pushing the boundaries of precision, speed, and efficiency in 3D visual sensing.

Acknowledgements

This work was supported by National Natural Science Foundation of China (U21B2033, 62205147, 62522508, 62571249), Fundamental Research Funds for the Central Universities (2023102001, 2024202002), National Key Laboratory of Shock Wave and Detonation Physics (JCKYS2024212111), China Postdoctoral Science Fund (2023T160318), Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (JSGP202105, JSGP202201, JSSJCG2026010) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX25_0695, SJCX25_0188).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. C. Wheatstone, "Contributions to the Physiology of Vision.—Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision," *Philosophical Transactions of the Royal Society of London* 128 (1838): 371–394.

2. E. N. Leith and J. Upatnieks, "Wavefront Reconstruction with Continuous-Tone Objects," *Journal of the Optical Society of America* 52, no. 10 (1962): 1123–1130.

3. J. Salvi, J. Fernandez, T. Pribanic, et al., "A State of the Art in Structured Light Patterns for Surface Profilometry," *Pattern Recognition* 43, no. 8 (2010): 2666–2680.

4. S. S. Gorthi and P. Rastogi, "Fringe Projection Techniques: Whither We Are?" *Optics and Lasers in Engineering* 48, no. 2 (2010): 133–140.

5. W. Chen, B. Zhang, L. Gu, et al., "Snapshot Compressive Imaging Based Digital Image Correlation: Temporally Super-Resolved Full-Resolution Deformation Measurement," *Optics Express* 30, no. 19 (2022): 33554–33573.

6. Z. Yin, Y. Chen, P. She, et al., "Modeling the Measurement Precision of a Multi-Camera System," *Optics Letters* 50, no. 21 (2025): 6489–6492.

7. E. N. Malamas, E. G. Petrakis, M. Zervakis, et al., "A Survey on Industrial Vision Systems, Applications and Tools," *Image and Vision Computing* 21, no. 2 (2003): 171–188.

8. C. R. Farrar and K. Worden, "An Introduction to Structural Health Monitoring," *Philosophical Transactions of the Royal Society A* 365, no. 1851 (2007): 303–315.

9. S. Heist, P. Dietrich, M. Landmann, et al., "Gobo Projection for 3D Measurements at Highest Frame Rates: A Performance Analysis," *Light: Science & Applications* 7, no. 1 (2018): 71.

10. X. Xi, Y. Liu, P. Xue, et al., "High-Speed Multi-Camera Videogrammetric Measurement of Full-Field 3D Motion and Deformation in Full-Scale Crash Testing of Typical Civil Aircraft," *Aerospace Science and Technology* 152 (2024): 109375.

11. C. Zuo, S. Feng, L. Huang, et al., "Phase Shifting Algorithms for Fringe Projection Profilometry: A Review," *Optics and Lasers in Engineering* 109 (2018): 23–59.

12. S. Kleinfelder, S. Lim, X. Liu, et al., "A 10000 Frames/s CMOS Digital Pixel Sensor," *IEEE Journal of Solid-State Circuits* 36, no. 12 (2001): 2049–2059.

13. B. Li, Y. Wang, J. Dai, et al., "Some Recent Advances on Superfast 3D Shape Measurement with Digital Binary Defocusing Techniques," *Optics and Lasers in Engineering* 54 (2014): 236–246.

14. C. Zuo, Q. Chen, S. Feng, et al., "Optimized Pulse Width Modulation Pattern Strategy for Three-Dimensional Profilometry with Projector Defocusing," *Applied Optics* 51, no. 19 (2012): 4477–4490.

15. K. Liu, Y. Wang, D. L. Lau, et al., "Dual-Frequency Pattern Scheme for High-Speed 3-D Shape Measurement," *Optics Express* 18, no. 5 (2010): 5229–5244.

16. C. Zuo, Q. Chen, G. Gu, et al., "High-Speed Three-Dimensional Profilometry for Multiple Objects with Complex Shapes," *Optics Express* 20, no. 17 (2012): 19493–19510.

17. C. Zuo, Q. Chen, G. Gu, et al., "High-Speed Three-Dimensional Shape Measurement for Dynamic Scenes Using Bi-Frequency Tripolar Pulse-Width-Modulation Fringe Projection," *Optics and Lasers in Engineering* 51, no. 8 (2013): 953–960.

18. S. Zhang, "Absolute Phase Retrieval Methods for Digital Fringe Projection Profilometry: A Review," *Optics and Lasers in Engineering* 107 (2018): 28–37.

19. S. Feng, Q. Chen, and C. Zuo, "Graphics Processing Unit–Assisted Real-Time Three-Dimensional Measurement Using Speckle-Embedded Fringe," *Applied Optics* 54, no. 22 (2015): 6865–6873.

20. C. Zuo, T. Tao, S. Feng, et al., "Micro Fourier Transform Profilometry (μFTP): 3D Shape Measurement at 10,000 Frames Per Second," *Optics and Lasers in Engineering* 102 (2018): 70–91.

21. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature* 521, no. 7553 (2015): 436–444.

22. J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Networks* 61 (2015): 85–117.

23. C. Zuo, J. Qian, S. Feng, et al., "Deep Learning in Optical Metrology: A Review," *Light: Science & Applications* 11, no. 1 (2022): 39.

24. Z. Li, J. Sun, Y. Fan, et al., "Deep Learning Assisted Variational Hilbert Quantitative Phase Imaging," *Opto-Electronic Science* 2, no. 4 (2023): 220023.

25. B. Wang, W. Chen, J. Qian, et al., "Single-Shot Super-Resolved Fringe Projection Profilometry (SSSR-FPP): 100,000 Frames-Per-Second 3D Imaging with Deep Learning," *Light: Science & Applications* 14, no. 1 (2025): 70.

26. S. Feng, Q. Chen, G. Gu, et al., "Fringe Pattern Analysis Using Deep Learning," *Advanced Photonics* 1, no. 2 (2019): 025001.

27. W. Yin, Y. Che, X. Li, et al., "Physics-Informed Deep Learning for Fringe Pattern Analysis," *Opto-Electronic Advances* 7, no. 1 (2024): 230034.

28. J. Qian, S. Feng, Y. Li, et al., "Single-Shot Absolute 3D Shape Measurement with Deep-Learning-Based Color Fringe Projection Profilometry," *Optics Letters* 45, no. 7 (2020): 1842–1845.

29. J. Qian, S. Feng, T. Tao, et al., "Deep-Learning-Enabled Geometric Constraints and Phase Unwrapping for Single-Shot Absolute 3D Shape Measurement," *APL Photonics* 5, no. 4 (2020): 046105.

30. Y. Li, J. Qian, S. Feng, et al., "Deep-Learning-Enabled Dual-Frequency Composite Fringe Projection Profilometry for Single-Shot Absolute 3D Shape Measurement," *Opto-Electronic Advances* 5, no. 5 (2022): 210021.

31. L. Gao, J. Liang, C. Li, et al., "Single-Shot Compressed Ultrafast Photography at One Hundred Billion Frames Per Second," *Nature* 516, no. 7529 (2014): 74–77.

32. P. Llull, X. Liao, X. Yuan, et al., "Coded Aperture Compressive Temporal Imaging," *Optics Express* 21, no. 9 (2013): 10526–10545.

33. D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable Pixel Compressive Camera for High Speed Imaging," *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2011): 329–336, https://doi.org/10.1109/CVPR.2011.5995542.

34. R. G. Baraniuk, "Compressive Sensing [Lecture Notes]," *IEEE Signal Processing Magazine* 24, no. 4 (2007): 118–121.

35. J. Park and L. Gao, "Cascaded Compressed-Sensing Single-Pixel Camera for High-Dimensional Optical Imaging," *PhotoniX* 5, no. 1 (2024): 37.

36. Z. Zhang, B. Zhang, X. Yuan, et al., "From Compressive Sampling to Compressive Tasking: Retrieving Semantics in Compressed Domain with Low Bandwidth," *PhotoniX* 3, no. 1 (2022): 19.

37. J. N. Mait, G. W. Euliss, and R. A. Athale, "Computational Imaging," *Advances in Optics and Photonics* 10, no. 2 (2018): 409–483.

38. Y. Luo, Y. Zhao, J. Li, et al., "Computational Imaging Without a Computer: Seeing Through Random Diffusers at the Speed of Light," *eLight* 2, no. 1 (2022): 4.

39. W. Chen, S. Feng, W. Yin, et al., "Deep-Learning-Enabled Temporally Super-Resolved Multiplexed Fringe Projection Profilometry: High-Speed kHz 3D Imaging with Low-Speed Camera," *PhotoniX* 5, no. 1 (2024): 25.

40. W. Chen, Y. Liu, S. Feng, et al., "Dual-Frequency Angular-Multiplexed Fringe Projection Profilometry with Deep Learning: Breaking Hardware Limits for Ultra-High-Speed 3D Imaging," *Opto-Electronic Advances* 8, no. 9 (2025): 250021.

41. J. Qian, S. Feng, Y. Li, et al., "Single-Shot Absolute 3D Shape Measurement with Deep-Learning-Based Color Fringe Projection Profilometry," *Optics Letters* 45, no. 7 (2020): 1842–1845.

42. J. Xu and S. Zhang, "Status, Challenges, and Future Perspectives of Fringe Projection Profilometry," *Optics and Lasers in Engineering* 135 (2020): 106193.

43. S. Lei and S. Zhang, "Digital Sinusoidal Fringe Pattern Generation: Defocusing Binary Patterns vs Focusing Sinusoidal Patterns," *Optics and Lasers in Engineering* 48, no. 5 (2010): 561–569.

44. H. Li, Y. Hu, T. Tao, et al., "Optimal Wavelength Selection Strategy in Temporal Phase Unwrapping with Projection Distance Minimization," *Applied Optics* 57, no. 10 (2018): 2352–2360.

45. J. Gu, Z. Wang, J. Kuen, et al., "Recent Advances in Convolutional Neural Networks," *Pattern Recognition* 77 (2018): 354–377.

46. R. N. Bracewell, "The Fourier Transform," *Scientific American* 260, no. 6 (1989): 86–95.

47. S. Feng, Y. Xiao, W. Yin, et al., "Fringe-Pattern Analysis with Ensemble Deep Learning," *Advanced Photonics Nexus* 2, no. 3 (2023): 036010.

48. N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation," *Neural Networks* 121 (2020): 74–87.

49. W. Hua, Y. Zhou, C. De Sa, et al., "Boosting the Performance of CNN Accelerators with Dynamic Fine-Grained Channel Gating," *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture* (2019): 139–150, https://doi.org/10.1145/3352460.3358283.

50. N. Shazeer, A. Mirhoseini, K. Maziarz, et al., "Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer," *arXiv preprint arXiv:1701.06538* (2017).

51. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), 234–241.

52. M. R. Kellman, E. Bostan, N. A. Repina, et al., "Physics-Based Learned Design: Optimized Coded-Illumination for Quantitative Phase Imaging," *IEEE Transactions on Computational Imaging* 5, no. 3 (2019): 344–353.

53. F. Wang, Y. Bian, H. Wang, et al., "Phase Imaging with an Untrained Neural Network," *Light: Science & Applications* 9, no. 1 (2020): 77.

54. F. I. Diakogiannis, F. Waldner, P. Caccetta, et al., "Resunet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data," *ISPRS Journal of Photogrammetry and Remote Sensing* 162 (2020): 94–114.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File 1: lpor71207-sup-0001-SuppMat.pdf.

Supporting File 2: lpor71207-sup-0002-Video1.mp4.