Deep-Learning-Based Single-Shot Fringe Projection Profilometry Using Spatial Composite Pattern

Yansong Jiang, Jiayi Qin, Yuankun Liu, Menglong Yang, *Member, IEEE*, and Yiping Cao

***Abstract*— Single-shot fringe projection profilometry (FPP) is crucial for real-time or dynamic 3-D measurement scenarios. In this regard, we propose a spatial composite FPP (SCFPP), which creatively fuses the spatial feature in different frequency fringe patterns. Inspired by image inpainting techniques in computer vision, SCFPP employs a novel adaptive space-division encoding strategy (ASES) to divide fringe patterns into several basic blocks and reaggregate partial basic blocks in the spatial domain to form a spatial composite fringe pattern. In the demodulation stage, we develop a deep-learning network model, the phase inpainting network (PIN), to restore the spatial composite fringe pattern to complete phase information. The absolute phase map is then obtained with high accuracy by applying the conventional three-frequency heterodyne phase unwrapping (CTHPU) algorithm with three known wrapped phase maps. To the best of the authors' knowledge, SCFPP is the first successful use of spatial multiplexing strategy in FPP. In addition, SCFPP can be seen as an application of image inpainting techniques in the fringe projection field, introducing new ideas for applying computer vision methods in fringe analysis.**

***Index Terms*— 3-D measurement, deep learning, fringe projection, image inpainting, single shot.**

I. Introduction

As an active and noncontact 3-D surface measurement technique, fringe projection profilometry (FPP) [1], [2], [3] has been the focus of considerable attention due to its high accuracy and flexibility. With the continuous advancement of electronic imaging sensors and computer vision technology, FPP has found widespread applications in various fields, such as mechanical engineering, on-line inspection, and machine vision [4], [5], [6]. These applications have not only facilitated the development of relevant fields but also necessitated improvements in FPP, particularly in terms of measurement efficiency.

The FPP measurement system is composed of three essential components: a projector, an industrial camera, and an algorithm processing unit, as shown in Fig. 1. In the process of measurement, the projector initially projects predesigned encoded patterns onto the tested object. Due to the modulation effect caused by the object's depth information, the projected patterns undergo a certain degree of deformation. This spatial deformation information resulting from the object's varying depths are embedded within the phase of the distorted patterns. Following this, the industrial camera is utilized to capture these pattern images, after which the phase demodulation takes place through the algorithm processing unit. Through procedures such as phase unwrapping and phase-to-height mapping, the deciphered phase information can reconstruct the 3-D shape of the object accurately. Theoretically, the measurement speed of the FPP system relies on the projector's projection speed, the camera's capture speed, and the number of encoded patterns projected. Despite the kilohertz-level speeds of modern hardware devices [7], [8], [9], the high cost and large size of high-speed projectors and cameras limit the cost and miniaturization of the measurement instrument. Thus, several researchers have attempted to reduce the number of encoded patterns to achieve an equivalent increase in measurement speed. Notably, single-shot FPP, which can recover the 3-D surface of an object from a single image, holds significant research and application value.

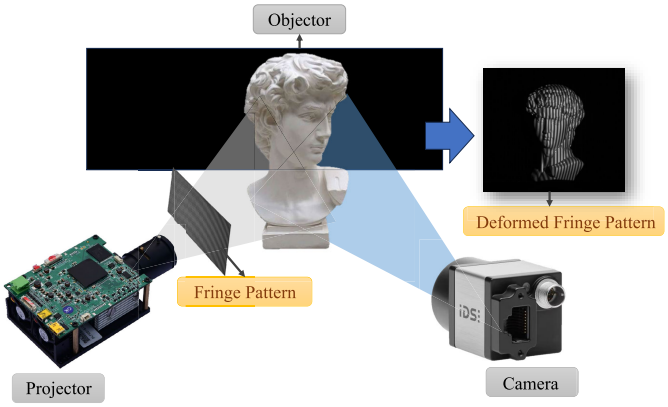


Fig. 1. Flowchart of FPP.

Traditional single-shot FPP algorithms can be classified into three categories based on the type of projected fringe patterns: single-frequency sinusoidal fringe projection, frequency-domain composite fringe projection, and spectral-domain composite fringe projection. The first algorithm projects a single-frequency sinusoidal fringe onto the object and retrieves the object's phase information through transformation algorithms. The second algorithm combines fringe patterns with different information in the frequency domain to achieve absolute phase unwrapping of the wrapped phase map. The

Manuscript received 3 January 2024; revised 15 April 2024; accepted 5 May 2024. Date of publication 28 June 2024; date of current version 23 July 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFF0712902, and in part by the National Natural Science Foundation of China (NSFC) under Grant 62271334. The Associate Editor coordinating the review process was Dr. Libing Bai. *(Yansong Jiang and Jiayi Qin are co-first authors.) (Corresponding author: Yuankun Liu.)*

Yansong Jiang and Menglong Yang are with the School of Aeronautics and Astronautics, Sichuan University, Chengdu, Sichuan 610065, China (e-mail: 657129773@qq.com; mlyang@scu.edu.cn).

Jiayi Qin, Yuankun Liu, and Yiping Cao are with the School of Electronic Information, Sichuan University, Chengdu, Sichuan 610065, China (e-mail: 809142027@qq.com; lyk@scu.edu.cn; ypcao@scu.edu.cn).

Digital Object Identifier 10.1109/TIM.2024.3420365

1557-9662 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

third algorithm compounds different information in the three channels of a color pattern to solve for a higher precision phase or to achieve absolute phase unwrapping. However, due to various physical factors, such as spectral aliasing, color imbalance, and color crosstalk, traditional single-shot measurement algorithms often struggle to FPP obtain high-precision reconstruction results in practical measurements.

In recent years, with the rapid advancement of artificial intelligence, the deep-learning technology has found widespread applications across numerous domains, including FPP. Researchers have leveraged deep-learning techniques to overcome the physical limitations inherent in traditional methods, resulting in significantly improved reconstruction outcomes. This development has also opened up the possibility of achieving a high-precision single-shot FPP algorithm. Especially, in the field of computer vision, deep-learning-based image inpainting techniques have been widely employed to restore missing pixel information and generate images that are highly similar to the original ones [10]. Regrettably, deep-learning-based image inpainting techniques have yet to fully integrate and make inroads into the domain of fringe analysis, thereby unquestionably constraining their potential for advancement.

To this end, this article proposes a novel spatial composite FPP (SCFPP) and treats the fringe decoding as an image inpainting task. We utilize the proposed phase inpainting network (PIN) to restore high-precision wrapped phase information. Subsequently, the phase ambiguity can be removed using conventional three-frequency heterodyne phase unwrapping (CTHPU) to obtain the continuous phase information of the target object. Experimental results demonstrate that our proposed algorithm achieves higher measurement accuracy compared with traditional single-frame measurement methods and can reconstruct complex isolated scenes. The main contributions of this work are as follows.

1) *Novel Fringe Composition Method, Termed Spatial Composition, Is Proposed for the First Time:* This method combines different frequency fringe patterns in the intensity domain. Compared with frequency composition, each pixel of a spatial composite pattern exclusively carries information from a single frequency, simplifying the demodulation process and achieving high precision in the demodulated phase. Furthermore, unlike spectral composition, this algorithm is not constrained by the object's color, thereby providing increased flexibility. To best of the authors' knowledge, our method is the first successful use of spatial composition in FPP.
2) *Image Inpainting Technology Is Introduced Into the Field of Fringe Analysis for the First Time:* During the demodulation process of spatial composite fringe patterns, we interpret the decoding step as an image inpainting task, ultimately achieving high-precision phase demodulation. This introduces new ideas for applying computer vision methods in fringe analysis.
3) *Fringe-to-Phase Hierarchical Convolutional Network, Named PIN, Is Proposed to Achieve High-Quality Reconstruction:* Specifically, our PIN is constructed in a layerwise manner, employing multiple hierarchical phase inpainting units (HPIUs), with optimal weights assigned through mixed attention mechanisms to achieve robust phase reconstruction. Furthermore, by leveraging enhanced attention units (EAUs) and phase extract blocks (PEBs), we obtain finer representations. PIN has demonstrated exceptional performance across various datasets, showcasing its effectiveness in both subjective and objective evaluations.
4) *Brand-New Dataset Is Gathered for the Proposed PIN:* We utilize our FPP experimental system to acquire an entirely novel dataset, distinguishable from publicly accessible datasets. This dataset comprises 958 image pairs, consisting of spatial composite fringe patterns and wrapped phase maps, and notably excludes any simulation image.

The remaining organization of this article is as follows. Related works will be reviewed in Section II; we will provide a detailed introduction to the proposed SCFPP in Section III; Section IV presents the experimental results; and finally, Section V describes the results and our future work.

II. Related Works

A. Traditional Single-Shot FPP

In 1983, Takeda et al. [11] proposed the first single-shot measurement algorithm, Fourier transform profilometry (FTP) [11], [12], [13], [14]. This method projects a high-frequency fringe pattern onto the tested surface and extracts the object's phase information through low-pass filtering after capturing the deformed fringe pattern with an imaging device. However, spectral aliasing occurs in FTP when processing the fringe spectrum due to interference between the background light intensity and the fundamental frequency component, making it difficult to fully filter out the phase information, thus affecting the accuracy of the obtained phase [15]. In addition, the phase value obtained by FTP is wrapped within the range of −pi to pi due to the characteristics of the arctangent function, requiring spatial phase unwrapping [16], [17], [18] techniques to obtain continuous phase by using the phase relationship between adjacent pixels. However, this phase unwrapping method is ineffective when measuring isolated scenarios [19]. In summary, while FTP has the advantage of single-frame measurement, it is only suitable for measuring smooth and continuous objects.

To tackle the aforementioned challenges while preserving the advantages of single-shot measurements, scholars have proposed composite fringe projection algorithms that fall into two categories: spectral-domain composite and frequency-domain composite. The former utilizes colored fringes to record the depth information and integrates diverse fringes, such as phase-shifting and multifrequency fringes, into the red, green, and blue channels of a color pattern. Huang et al. [20] embedded three patterns with different phase shifts into a color stripe and utilized spatial phase unwrapping to obtain relative phase information. Compared with the phase obtained

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

through FTP, the phase calculated through three-step phase shifting method was more accurate. However, during the phase unwrapping process, the algorithm faced challenges in measuring complex and isolated scenes due to the reliance on neighboring pixel phase values to remove phase discontinuities. Chen et al. [21] proposed the color composite pattern projection algorithm, where they used different colors to calibrate the order information in the stripe pattern and employed wavelet transform to solve the phase in sinusoidal stripes. As each pixel's phase was independently determined, this algorithm avoided the error diffusion problem encountered in spatial phase unwrapping algorithms. Nevertheless, this algorithm could not be applied to isolated measurement scenes since the order of the color stripe relied on neighboring pixel colors for determination. To address measurement issues in isolated scenes, Su [22] combined sinusoidal stripes, binary stripes, and color gratings into a single-color pattern. However, due to the incorporation of multiple information into one pattern, the decoded depth information often suffered from limited accuracy. Zhang et al. [23] embedded three fringes of varying wavelengths into the three channels of a color pattern and utilized the Fourier transform to extract the phase information. The use of color fringes introduced spectral problems, such as chromatic aberration, color imbalance, crosstalk, and color texture, which will hinder the algorithm's performance in practical scenarios.

The latter approach composites different fringes in the frequency domain of a grayscale image and demodulates these fringes using appropriately designed filters. Guan et al. [24] utilized four fringe patterns with varying frequencies as carriers to compose four single-frequency phase-shifting images. In the frequency spectrum, these four single-frequency phase-shifting fringe patterns are distributed at different positions and can be obtained through low-pass filtering. Takeda et al. [25] employed the frequency multiplexing to combine two fringe patterns for the measurement of spatially isolated and discontinuous scenes. Similarly, Sansoni and Redaelli [26] attempted to merge two fringes of different frequencies into a single pattern. Although these diverse encoding methods allow a grayscale pattern to contain more information, the accuracy of the demodulated phase is significantly low due to issues such as spectrum aliasing. As a result, these methods face challenges when applied in high-precision measurement scenarios.

B. Deep-Learning-Based Single-Shot FPP

In recent years, with the development of artificial intelligence, the deep-learning technology has brought new possibilities for realizing high-accurate single-shot FPP [27]. Feng et al. [28] utilized deep learning to analyze fringes, employing a convolutional neural network (CNN) to extract object phase information from a fringe pattern. While this approach achieves greater measurement accuracy than FTP, it is limited to obtaining phase information for a single frequency and does not address the phase unwrapping problem in isolated scenes. Yu et al. [29] proposed a fringe pattern transformation network to forecast phase-shifting fringe images, which allows for single-frame measurements of isolated scenes in conjunction with geometric constraints. Nonetheless, the application of geometric constraints proves to be unfeasible for measuring objects with large depths. Qian et al. [30] employed deep learning and stereo-phase unwrapping technology to attain single-frame measurements. Despite this technique's ability to generate dependable phase unwrapping outcomes, the requirement for an extra camera leads to elevated hardware expenditures. Li et al. [31], [32] introduced a novel approach that utilizes deep learning to extract rough absolute and wrapped phase maps from a frequency-domain composite fringe pattern, thus circumventing the spectral aliasing problem that conventional methods face. Nevertheless, frequency-domain composite algorithms, in which all frequency fringe patterns share the same bandwidth of the hardware including both the projector and sensor, will produce a lower result compared with single-frequency strategy even in ideal conditions. Qian et al. [33] devised a CNN that extracts phase maps from a color fringe image, effectively addressing issues like chromatic aberration and color imbalance that plague conventional approaches. However, it still keeps the limitation for measuring the colored objects. Alongside the aforementioned techniques, several scholars have endeavored to employ deep learning for end-to-end measurement by directly translating encoded patterns into depth information [34], [35], [36]. Nonetheless, in practical measurement scenarios, these techniques face challenges in producing stable and highly accurate outcomes.

III. Method

In this section, we will offer a comprehensive exposition of our innovative SCFPP methodology. The measurement process of SCFPP is similar to that of FPP. Initially, we employ an adaptive space-division encoding strategy (ASES) to generate a spatial composite fringe pattern, which is casted onto the target object. The presence of object height information induces deformations in the fringe pattern. Subsequently, the camera captures these deformed fringe patterns, which are then subjected to demodulation through our designed PIN. Finally, the wrapped phase maps, obtained through demodulation, are leveraged for phase unwrapping by the CTHPU. A detailed elucidation of the specific algorithmic processes will be presented in the forthcoming sections. Among these, Section III-A will expound upon the encoding principles underlying spatial composite fringe patterns; Section III-B will introduce the methodology for constructing a dataset using the FPP system; Section III-C will provide an in-depth introduction to the intricate details of the designed PIN; and Section III-D will elucidate the algorithmic process of phase unwrapping.

A. Adaptive Space-Division Encoding Strategy (ASES)

The spatial composite pattern differs from traditional spectral and frequency composite patterns by combining different fringe images in the spatial intensity domain. In this article, we utilize the proposed ASES to encode spatial composite fringe patterns for measurement, as shown in Fig. 2. The basic

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

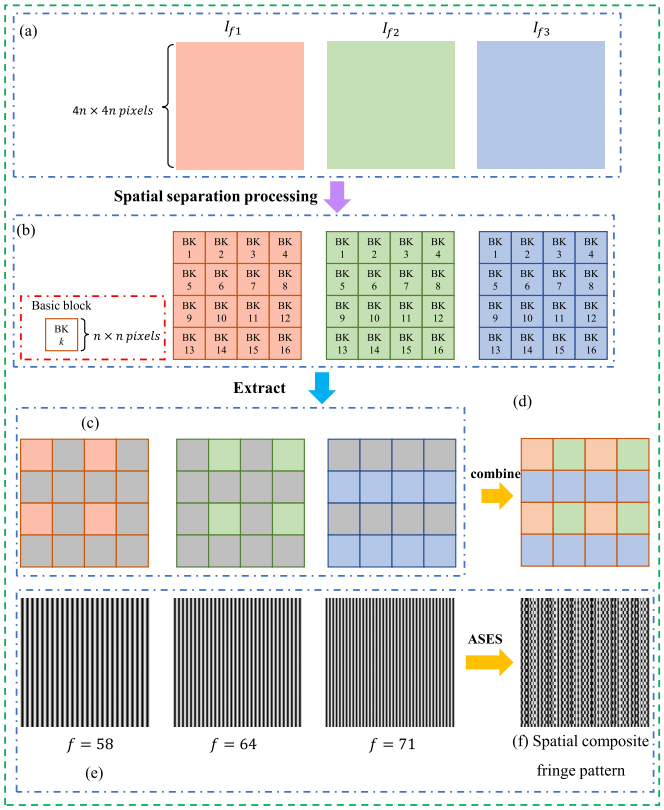


Fig. 2. Flowchart of encoding three fringe patterns with different frequencies using ASES. (a) Three patterns with different frequencies. (b) Three split patterns. (c) Extracted basic blocks from different patterns. (d) Spatial composite fringe pattern. (e) Three fringe patterns were generated by the computer. (f) Spatial composite fringe pattern.

idea of ASES is to create a new fringe pattern by using a set of small blocks, which are split from the three-frequency fringe images. As depicted in Fig. 2(a), the three patterns possessing different frequencies ($f_1 < f_2 < f_3$) are distinguished in red, green, and blue, respectively.

First, we segment the three complete images into split patterns through spatial separation processing [as illustrated in Fig. 2(b)]. For the sake of clarity, we assume that the entire image can be partitioned into 16 blocks. In practical scenarios, the number of basic blocks is dependent on the projected image size as well as the basic blocks. The selection of basic block size plays a critical role in ensuring the accuracy of reconstruction results, which will be further elucidated in the experimental section.

Second, ASES extracts a part of basic blocks from each split pattern. In ASES, there are three fundamental principles for extracting basic blocks. First, the extracted basic block from the split pattern should maintain its spatial position. For instance, in Fig. 2(b), BK_1–BK_{16} denotes 16 basic blocks at different locations. If we extract the basic block at BK_1 from I_{f1}, its position in the composite pattern must remain as BK_1. Second, only one fringe image can provide a basic block at a specific position in the composite pattern. For example, in Fig. 1(c), extracting the basic block at position BK_1 from the other split patterns is not allowed if we have extracted the same basic block from I_{f1}. Third, the interval extraction strategy is adopted, meaning each complete pattern can occupy only partial intensity in the composite pattern. Thus, existing basic blocks must be utilized to repair or predict the vacancy to recover complete information. The correlation between basic blocks from the same pattern determines the accuracy of recovered complete information. Hence, ASES applies the interval extraction strategy to ensure a high correlation between basic blocks. Based on these extraction principles, in Fig. 2, we extract BK_1, BK_3, BK_9, and BK_{11} for I_{f1}, BK_2, BK_4, BK_{10}, and BK_{12} for I_{f2}, and BK_5–BK_8 and BK_{13}–BK_{16} for I_{f3}. Finally, the spatial composite pattern can be generated by combining these basic blocks as illustrated in Fig. 2(d).

After generating and projecting spatial composite fringe patterns, the next step involves addressing the demodulation of these fringe patterns. As depicted in Fig. 2, it becomes evident that fringe patterns at each frequency occupy only a portion of the intensity information within the spatial composite fringe pattern. Consequently, conventional demodulation techniques, such as phase shifting or Fourier analysis, prove inadequate for extracting phase information. Therefore, we regard the demodulation of the composite pattern as an image inpainting task, which aims to recover the phase information through deep learning.

In comparison with the frequency-compound fringe pattern, the spatial-compound fringe pattern exhibits higher reliability in the demodulated phase since each pixel solely contains information from a single frequency. Moreover, unlike the color composite fringe pattern, the spatial composite fringe pattern is not constrained by the color of the tested object. In the ASES, an interval extraction strategy is adopted to ensure the correlation between the missing basic blocks and other basic blocks. In addition, neighboring basic blocks in the fringe pattern usually exhibit higher correlations. Thus, compared with other image inpainting tasks, restoring the fringe image is relatively easier.

B. Construct High-Quality Datasets

Before using the deep-learning technique to decode the spatial composite pattern, we must construct a high-quality dataset and train the designed CNN. In the dataset, we define the input as the spatial composite fringe pattern generated by ASES. As for the dataset's output, considering that the measurement scenarios were conducted under static conditions with diverse objects and poses, we opted to utilize the N-step phase-shifting method to calculate the highest precision wrapped phase maps. These wrapped phase images then serve as the ground truth for the dataset. Fig. 3 depicts the flowchart of building the training dataset. There are $3N + 1$ images in the projection sequence of the projector, the first $3N$ images are N-step phase-shifting images with three different frequencies, and the other image is the spatial composite fringe pattern formed by ASES. The N-step phase-shifting patterns are used to obtain the accurate wrapped phase information, the fringe patterns can be represented as follows:

$$I_f(i) = A + B\cos\left[\Phi_f + 2\pi(i-1)/N\right] \tag{1}$$

where f represents the fringe frequency, $I_f(i)$ is the ith fringe image, $i \in [1, N]$, A is the average intensity, B is the

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

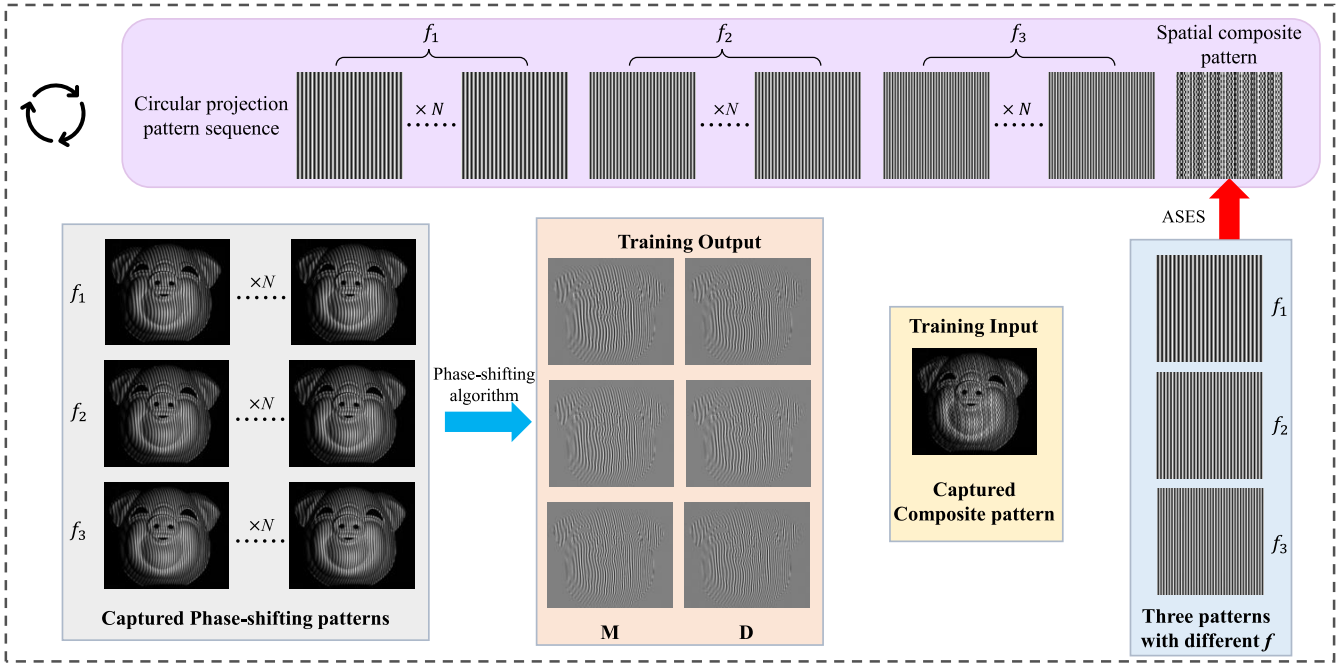


Fig. 3. Flowchart of generating training dataset.

amplitude, and Φ_f is the absolute phase. The wrapped phase ϕ_f can be obtained by

$$\phi_f = \arctan \frac{\sum_{i=0}^{N-1} I_f(i) \sin\left[2\pi(i-1)/N\right]}{\sum_{i=0}^{N-1} I_f(i) \cos\left[2\pi(i-1)/N\right]} = \arctan \frac{M_f}{D_f} \tag{2}$$

where M_f and D_f represent the numerator and denominator of the arctangent function and are labeled as training output. The captured spatial composite pattern is labeled as training input. By capturing a large number of different measurement scenes, a high-quality dataset can be built, and we can train the designed CNN by this dataset.

C. Phase Inpainting Network (PIN)

To improve the efficiency and accuracy of the 3-D reconstruction, PIN is designed to realize the nonlinear mapping from the composite image to the wrapped phase (numerator and denominator). As shown in Fig. 4, PIN follows the feature training strategy of U-Net and fully extracts the information of the composite-block input. Then, through hierarchical expansion and refinement, the fused features can be restored to images of the same size and channel as the input. Specifically, our CNN can be illustrated in the following three parts.

1) Phase Encoder Downsampling, Information Analysis, and Feature Extraction: In the encoding phase, we aim to enhance the capability of information representation through a series of steps, including down sampling, information analysis, and feature extraction. We achieve comprehensive feature extraction during the information encoding process. For the spatial composite image with a single-input channel, the relevant information is predominantly concentrated within the spatial pixels, making the efficacy of a single filter often limited. To address this, we employ a 64-channel convolutional layer in the encoding phase to initially extract spatial-level information. Subsequently, by leveraging the interplay between feature pixels and the channel's expandability, the encoder further unearths and extends the shallow information. We further expand the receptive field by extending the network with multiple convolutional layers, ensuring the extraction of all relevant information. As illustrated in Fig. 3, following four levels of expansion, the pertinent information is decomposed from a single channel into 1024 channels, leading to a corresponding reduction in image scale. At this stage, information at all levels (x_1, x_2, x_3, x_4, and x_5 in Fig. 4) has been progressively acquired and transferred, resulting in a collection of multichannel, compact feature sets.

The specific encoding process is as follows. As it is shown in Fig. 4, given the input of PIN is x and the x_i denotes the output of ith HPIU_E^i. In PIN, there are five HPIUs containing five similar structures with different input and output channel numbers. Assume that Pool$(\cdot)$ represents the max-pooling with a stride of 2. After the first HPIU-E, the operation of our PIN can be expressed as follows:

$$x_1 = \text{HPIU}_E^1(x) \tag{3}$$

where $\text{HPIU}_E^1(\cdot)$ denotes the functional function corresponding to the first module. Similarly, we can map the functions of functional modules into a unified standardized representation

$$x_i = \text{HPIU}_E^i(\text{Pool}(x_{i-1})), \quad (2 \le i \le 5). \tag{4}$$

Consequently, following the processing through five HPIU_E^s, we obtain the final output denoted by x_5. Through the multilayer convolution operations performed by the encoder, the profound features of the phase map are comprehensively extracted.

2) Phase Decoder–Upsampling, Information Fusion, and Feature Reconstruction: For the analytical features extracted by the encoder, the decoder plays a crucial role in restoring

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

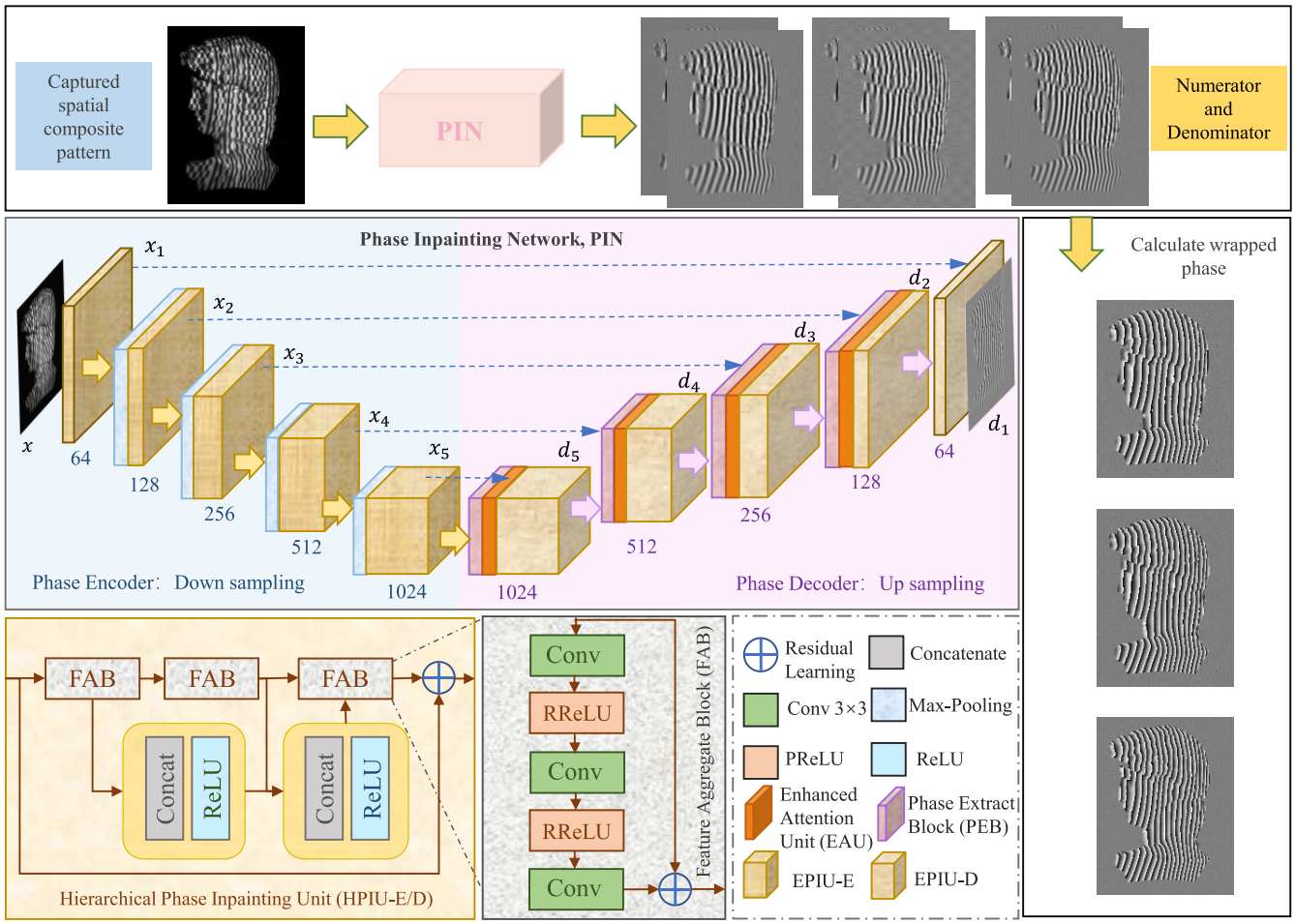


Fig. 4. PIN.

the wrapped phase numerator and denominator, encompassing the tasks of information reconstruction and fusion. Within the decoder, it is imperative to re-establish the channel count and specifications related to the processing object, ensuring their appropriateness and effectiveness. To mitigate the issue of short-term memory loss, during the information compression phase, the convolutions from the same level within the encoder are fed into the decoder for supervision (as illustrated by the blue dotted line in Fig. 4). After several stages of information fusion, the numerator and denominator features of the wrapped phase are reconstructed as needed. The concept of "self-adaptation" primarily manifests in the number and specifications of channels in the input image. In other words, the input image can have varying scales and channel configurations. This adaptability enhances PIN's potential for application and research in real-world 3-D measurement scenarios, significantly contributing to the ongoing advancement of optical image processing.

Given the output of the ith HPIU_D^i is d_n $(n = 6 - i)$, the function implementation process of the decoder is as follows:

$$d_5 = \mathrm{HPIU}_D^1(\mathrm{concat}[x_5, \mathrm{EAU}(\mathrm{PEB}(x_5))]) \tag{5}$$

where $\mathrm{EAU}(\cdot)$ and $\mathrm{PEB}(\cdot)$ represent the internal convolution operations of EAUs and PEBs, respectively, and $\mathrm{concat}[A, B]$ denotes the concatenate of A and B. When $2 \leq i \leq 4$, the operation of hierarchical dephasing can be expressed as follows:

$$d_n = \mathrm{HPIU}_D^1\big(\mathrm{concat}\big[x_n, EAU(\mathrm{PEB}(d_{n+1}))\big]\big), \quad 2 \leq n \leq 4. \tag{6}$$

Finally, after the calculation of the final layer EPIU_D^5, we can obtain the truncated phase of the target

$$d_1 = \mathrm{HPIU}_D^5(\mathrm{concat}[x_1, d_5]). \tag{7}$$

It is worth noting that there is no $\mathrm{EAU}(\cdot)$ and $\mathrm{PEB}(\cdot)$ operations in the final convolutional layer. This, to a certain extent, enhances the efficiency of feature extraction and aligns more closely with the principles of lightweight network design.

3) Feature Extraction and Phase Refinement: To achieve more efficient feature extraction and enhance phase recovery quality, we have meticulously crafted a hierarchical nested network architecture, enabling us to attain more precise representations. This design involves the implementation of distinct training strategies at different levels, accompanied by the development of corresponding functional modules. At the global level, we have harnessed the power of the PIN for encoder–decoder phase recovery, incorporating long-term memory elements at a macroscopic scale. In the intermediate tier, we have introduced HPIUs to refine feature granularity

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

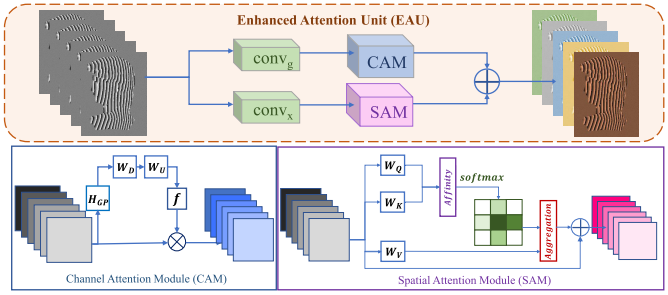


Fig. 5. Structural diagram of EAU.

through localized feature extraction and fusion. At the pixel level, we have integrated feature aggregate blocks (FABs) to explore and fuse point-to-point features of subpixels. By extracting and fusing features at multiple levels and fine-tuning the encoding–decoding learning strategy, we can achieve highly satisfactory phase recovery outcomes. To facilitate comprehension, we have defined the aforementioned modules.

a) Hierarchical phase inpainting unit (HPIU): Given the input of HPIU is H_{in} and the output of the ith FAB is FAB_i, the output H_{in} can be formulated as follows. Shallow extraction relies on the first two FABs and can be defined as follows:

$$\begin{cases} F_1 = \text{FAB}_1(H_{\text{in}}) \\ F_2 = \text{FAB}_2(F_1). \end{cases} \tag{8}$$

Here, $\text{FAB}(\cdot)$ is the operation of the ith FAB, and it can be used for constructing the high-level features HLF_i

$$\begin{cases} \text{HLF}_1 = \text{ReLU}(\text{concat}[F_1, H_{\text{in}}]) \\ \text{HLF}_2 = \text{ReLU}(\text{concat}[F_2, \text{HLF}_1]) \end{cases} \tag{9}$$

where $\text{ReLU}(\cdot)$ is activation function named rectified linear units (ReLUs). Further integration on the basis of shallow features can yield deep features

$$F_3 = \text{FAB}_3(\text{concat}[F_2, \text{HLF}_2]). \tag{10}$$

To alleviate the problem of gradient vanishing or exploding during the training process, we introduced residual learning operations and obtained the output H_{out}

$$H_{\text{out}} = F_3 + H_{\text{in}}. \tag{11}$$

Here, "+" denotes the operation of residual learning. In addition, the function of FAB can be defined as follows:

$$\text{FAB}_{\text{out}} = \text{conv}_3(\text{ReLU}(\text{conv}_3(\text{ReLU}(\text{conv}_3(\text{FAB}_{\text{in}}))))) + \text{FAB}_{\text{in}}. \tag{12}$$

Here, FAB_{in} and FAB_{out} are the input and output of FABs, respectively. $\text{conv}_3(\cdot)$ represents the 3×3 convolutions.

b) Enhanced attention unit (EAU): The structure of EAU is shown in Fig. 5, consisting of channel attention (CA) and spatial attention mechanisms ($\text{CA}(\cdot)$ and $\text{SA}(\cdot)$ are the channel and spatial attention mechanisms). According to the CA mechanism in [37], we can extract effective information channelwise and obtain the CA_{out}

$$\text{CA}_{\text{out}} = \text{CA}\big(\text{conv}_g(\text{EAU}_{\text{in}})\big). \tag{13}$$

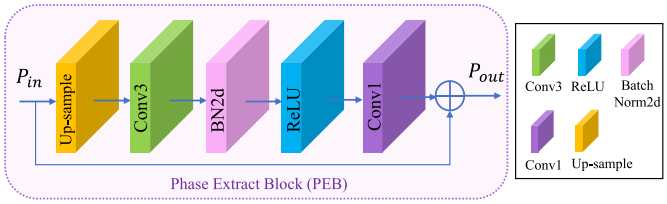


Fig. 6. Structural diagram of PEB.

Similarly, based on [38], spatial attention is conducted on the EAU_{in} and the SA_{out} also can be gained

$$\text{SA}_{\text{out}} = \text{SA}(\text{conv}_x(\text{EAU}_{\text{in}})). \tag{14}$$

Finally, the output EAU_{out} can be gained by conducting residual learning between CA_{out} and SA_{out}

$$\text{CA}_{\text{out}} = \text{CA}_{\text{out}} + \text{SA}_{\text{out}}. \tag{15}$$

c) Phase extract block (PEB): As depicted in Fig. 6, the pixel extract block (PEB) serves as the fundamental component of the PIN, responsible for conducting pixel-level feature extraction. To provide a deeper insight into the pivotal role played by PEBs in extracting phase information from images, we will delve into the architectural details of PEBs, as illustrated in Fig. 6. Given that the input to a PEB is denoted by P_{in}, and P_{out} represents the respective output, we have consciously pursued a strategy aimed at maintaining the lightweight nature of our PIN architecture. This strategy entails a deliberate reduction in the number of convolution layers within the network, while concurrently enhancing the structural capacity for feature representation. To elaborate, the process begins with upsampling and the extraction of the phase information map F_U, as follows:

$$F_U = \text{conv}_3(\text{UpSample}(P_{\text{in}})) \tag{16}$$

where $\text{conv}_3(\cdot)$ is the 3×3 convolution, and the upsample operation is represented by $\text{UpSample}(\cdot)$. Then, F_U is fed into the batch normalization layer $\text{BN}_{2d}(\cdot)$ and the activation function $\text{ReLU}(\cdot)$ for feature refinement

$$F_R = \text{conv}_1(\text{ReLU}(\text{BN}_{2d}(F_U))). \tag{17}$$

Here, $\text{conv}_1(\cdot)$ is a 1×1 convolution to match the numbers of channels. In addition, $\text{ReLU}(\cdot)$ represents the operations of ReLU and finally the refined features are defined as F_R. Finally, utilizing the residual learning, P_{out} can be obtained

$$P_{\text{out}} = F_R + P_{\text{in}}. \tag{18}$$

Obviously, through the extraction and refinement of image information, pixel-level features are fully utilized. This also lays a solid foundation for building a lighter and more efficient network.

d) Loss function: During the process of network training, we employ the L_1 function as our loss function. We can calculate L_1 loss as follows: given an LR image I_{LR} is fed into our PIN, by the processing of PIN, we can obtain its corresponding SR vision I_{SR}. Then, we can denote the target HR counterpart of I_{LR} as I_{HR}. To optimize our PIN,

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

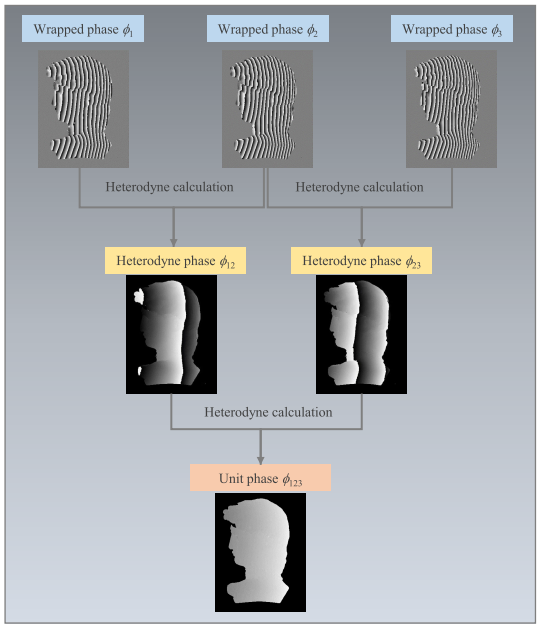


Fig. 7. Flowchart of calculating the unit phase map.

we introduce L_1 loss function. For an LR image whose batch size is set to S, we can get the responding value

$$L(\theta) = \frac{1}{S}\sum_{i=1}^{S}\left\| I_{\mathrm{LR}}^{i} - I_{\mathrm{HR}}^{i} \right\|_1. \tag{19}$$

Here, θ is the parameter set employed to train the backbone network. In detail, $\{I_{\mathrm{LR}}^{i}, I_{\mathrm{HR}}^{i}\}$ is the LR–HR pairs in the batch.

D. Phase Unwrapping

After obtaining the deformed spatial composite pattern, the numerator and denominator of the wrapped phase can be predicted by a pretrained PIN. Using (2), three wrapped phase maps ϕ_1, ϕ_2, and ϕ_3, can be obtained. Due to the mathematical properties of the arctangent function, the phase values calculated by (2) will be constrained to the range of −pi to pi. As the phase that records object depth information are continuous, a phase unwrapping algorithm is necessary to remove phase ambiguity. In this article, we use CTHPU to achieve phase unwrapping. In order to describe the process of phase unwrapping more clearly, take the fringes with frequency $f_1 = 58$, $f_2 = 64$, and $f_3 = 71$ as an example. Without the loss of generality, the proposed method can also apply to other fringe patterns with different frequencies.

First, by performing heterodyne calculations, three wrapped phase maps of different frequencies can be transformed into two heterodyne phase maps

$$\phi_{21} = \begin{cases} \phi_2 - \phi_1, & \phi_2 \geq \phi_1 \\ \phi_2 - \phi_1 + 2\pi, & \phi_2 < \phi_1 \end{cases} \tag{20}$$

$$\phi_{32} = \begin{cases} \phi_3 - \phi_2, & \phi_3 \geq \phi_2 \\ \phi_3 - \phi_2 + 2\pi, & \phi_3 < \phi_2 \end{cases} \tag{21}$$

where ϕ_1, ϕ_2, and ϕ_3 are the wrapped phase with fringe frequency f_1, f_2, and f_3, respectively. ϕ_{21} and ϕ_{32} are the heterodyne phase maps. The frequencies f_{21} and f_{32} corresponding to ϕ_{21} and ϕ_{32}, are now equal to 6 and 7, respectively. Using these phase maps, a unit phase image can be obtained by

$$\phi_{321} = \begin{cases} \phi_{32} - \phi_{21}, & \phi_{32} \geq \phi_{21} \\ \phi_{32} - \phi_{21} + 2\pi, & \phi_{32} < \phi_{21}. \end{cases} \tag{22}$$

The calculation process of the unit phase map is shown in Fig. 7. Since the frequency of the unit phase map is 1, phase unwrapping is unnecessary, and its wrapped phase is equal to the continuous phase. As a result, we can use the unit phase diagram to guide the high-frequency phase maps for phase unwrapping. Fig. 8 illustrates the phase values of a row extracted from the unit phase map. Second, the fringe orders k_{21} of ϕ_{21} will be found with the help of the unit phase

$$k_{21} = \mathrm{Round}\left[\frac{(f_{21}/f_{321})\Phi_{321} - \phi_{21}}{2\pi}\right]. \tag{23}$$

Here, the unit phase Φ_{321} can be rewritten as ϕ_{321}. Then, the continuous phase Φ_{21} is acquired by

$$\Phi_{21} = \phi_{21} + 2\pi k_{21}. \tag{24}$$

Similarly, with the assistance of Φ_{21}, the continuous phase Φ_3 can be acquired by

$$\Phi_3 = \phi_3 + 2\pi\,\mathrm{Round}\left[\frac{(f_3/f_{21})\Phi_{21} - \phi_3}{2\pi}\right]. \tag{25}$$

After obtaining the continuous phase, the phase information of the object can be obtained by subtracting the phase of the reference plane. The process of phase unwrapping and reconstruction is shown in Fig. 9.

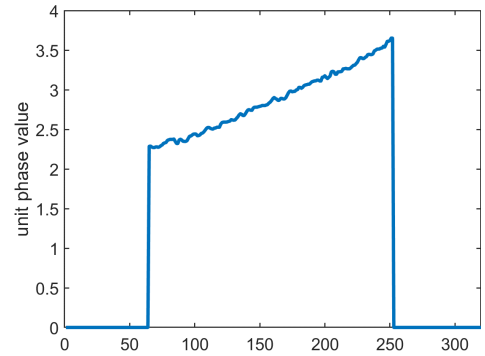


Fig. 8. One cross section of the unit phase map.

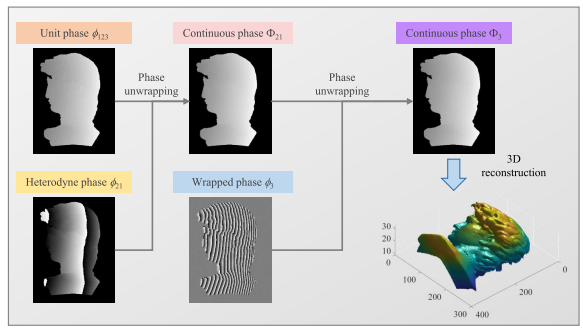


Fig. 9. Flowchart of phase unwrapping and 3-D reconstruction.

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

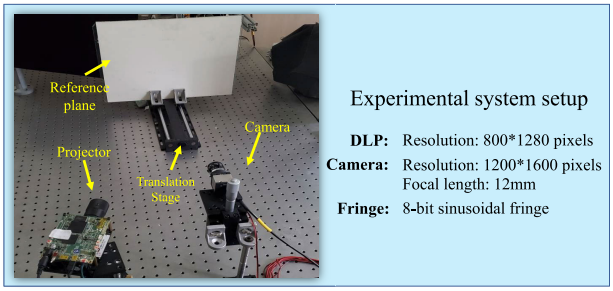


Fig. 10. Experimental system setup.

IV. Results and Discussion

In this section, Section IV-A addresses the optical experimental system's configuration and expounds upon the specifics of the network training arrangement. Section IV-B illuminates the underlying reasoning for selecting basic blocks and underscores the network architecture's effectiveness via the ablation study. Section IV-C empirically establishes the superiority of the proposed method through comparative experiments. Section IV-D presents the reconstruction outcomes of SCFPP within complex scenarios, thus providing additional evidence of the approach's efficacy.

A. Experimental System Setup

1) Measurement System Setup: We constructed an FPP system comprising an 800 × 1280 pixel projector and a 1200 × 1600 pixel camera fit with a 12-mm lens, to validate the effectiveness of our approach. The generated fringe frequencies for the experiment were set at 58, 64, and 71. We collected a total of 900 data sets for various scenes, encompassing simple, complex, and isolated objects, which were then split into training and validation sets at an 8:1 ratio. Fig. 10 shows the hardware setup.

2) Training Implementation Details: PIN was trained on NVIDIA 2080 Ti GPU. Every 16 images are trained as a batch during training, and their sizes depend on the sampling specification. For more proper training, we limit the initial learning rate of the Adam optimizer to 5×10^{-4} and halving for every 2×10^2 epochs.

To clearly illustrate the generalization capabilities and advantages of our proposed network, we incorporated another publicly available dataset [30] for training purposes and conducted a thorough comparison of the outcomes. In subsequent experiments, the captured dataset is named Dataset 1, and the publicly available one is Dataset 2.

While the object depth can be easily mapped from phase information in FPP, we use phase accuracy to evaluate the algorithm's reconstruction precision. Specifically, in our ablation and comparative experiments, we employed the wrapped phase maps calculated using the phase-shifting method as the ground truth. Subsequently, we utilized CNNs to predict the wrapped phase map and compared it to the ground truth to assess the algorithm's performance. In this experimental configuration, we utilized three distinct evaluation metrics: mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). This selection allowed for a more comprehensive and robust assessment of the predicted results.

TABLE I

Comparative Results Obtained on Two Datasets With/Without EAU for Wrapped Phase Retrieval

Method	DataSet1		DataSet2	
	MAE	PSNR/SSIM	MAE	PSNR/SSIM
w/o EAU	0.0412	34.72/0.9257	0.0347	**35.43**/0.9527
w EAU	**0.0366**	**35.63/0.9334**	**0.0326**	35.32/**0.9575**

B. Ablation Study

1) Verification of Basic Blocks: To determine the optimal basic block size for ASES, we captured nine datasets and compared their phase accuracy. Because FPP captures an object's depth information by analyzing its phase, the precision of the calculated phase can serve as an indicator of the overall measurement accuracy. The only difference between these datasets was their input for training and validation, which was a different spatial composite fringe pattern. The ground truth, measurement scenarios, and hardware used were identical for all datasets. The nine different spatial composite fringe patterns had basic block sizes of 1 × 1, 5 × 5, 10 × 10, 15 × 15, 20 × 20, 25 × 25, 30 × 30, 35 × 35, and 40 × 40. In this experiment, we used these nine datasets to train the proposed PIN and measured a complex sculpture to compare the phase accuracy predicted by PIN.

Fig. 11(a) shows the sculpture and the line plot of the MAE of phase error. The plot shows a U-shaped trend. Fig. 8(b) shows the specific phase error map. It can be seen from the figure that when the basic block size is 1, 5, or 40, the phase error is large, with an MAE value of around 0.2. When the basic block size is 10, 15, 25, 30, or 35, although the MAE value of phase error is less than 0.1, there is a larger error in the local area of the tested object. The phase error is smallest when the basic block size is 20. This experiment demonstrates that the selection of a basic block size has a significant impact on the final phase accuracy and an appropriate basic block size must be chosen to ensure the final measurement accuracy. In subsequent experiments, we chose a basic block size of 20.

2) Enhanced Attention Unit (EAU): The attention mechanism primarily serves the purpose of assigning a corresponding weight value to specific elements within the network structure, such as channels, pixels, or batches. It is typically employed in conjunction with application requirements to enhance the network's reconstruction quality. Our approach integrates both spatial attention and CA mechanisms, enabling the judicious allocation of weights for network components at both pixel and channel levels, thereby enhancing training efficiency. To substantiate the efficacy of our proposed EAUs, we conducted ablation experiments on the attention mechanisms using Dataset 1 and Dataset 2, respectively. The results are presented in Table I. Evidently, the network augmented with the attention mechanism exhibited superior reconstruction outcomes on Dataset 1 (ours), thereby affirming the rationale behind incorporating the attention mechanism.

Moreover, we substantiated the effectiveness of the attention enhancement mechanism, as detailed in Table II.

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

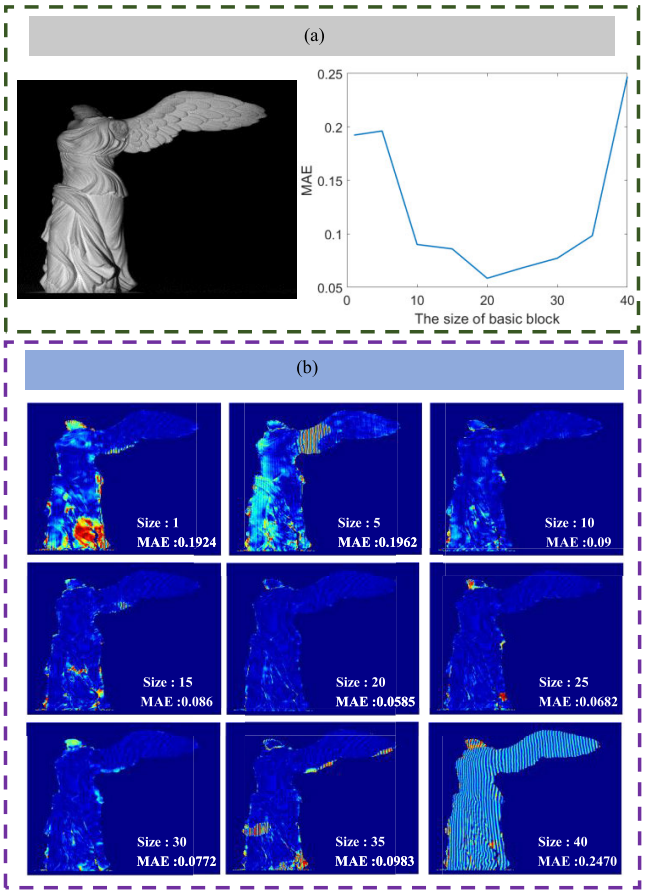


Fig. 11. (a) Measured object and the line map of the MAE value of the phase errors. (b) Wrapped phase errors of different spatial composite patterns.

TABLE II

COMPARATIVE RESULTS OBTAINED ON TWO DATASETS WITH/WITHOUT HYBRID EAU FOR WRAPPED PHASE RETRIEVAL

Method		DataSet1		DataSet2	
CA	SA	MAE	PSNR/SSIM	MAE	PSNR/SSIM
		0.0412	34.72/0.9257	0.0347	35.43/0.9527
✓		0.0409	35.1/0.9279	0.0346	34.83/0.9530
	✓	0.0417	34.7/0.9252	0.036	34.83/0.9543
✓	✓	**0.0366**	**35.63/0.9334**	**0.0326**	**35.32/0.9575**

We conducted a comparative analysis across different methods using two distinct test datasets, leading to the following conclusions: 1) the mixed attention mechanism (CA&SA) outperforms a single attention mechanism (CA/SA) in terms of reconstruction quality; 2) within the PIN framework, the spatial attention (SA) mechanism exhibits a slightly superior performance compared with the CA mechanism; and 3) the incorporation of EAUs enhances the resilience of PIN during reconstruction. As illustrated in Table II, with the inclusion of EAUs, the MAE decreased from 0.0412 to 0.0366, demonstrating a substantial improvement in performance.

3) Phase Extract Block (PEB): The PEB, serving as the fundamental building block within our PIN architecture, assumes a pivotal role in enhancing the overall reconstruction quality. To comprehensively illustrate the effectiveness of PEB, we conducted a comparative analysis utilizing objective evaluation metrics, including MAE, PSNR, and SSIM, on two datasets, as presented in Table III. Upon reviewing the data in Table III, it becomes evident that the incorporation of PEBs not only enhances the performance of PIN but also fortifies its robustness. PEBs contribute to a more refined extraction of phase information, as manifested by the reduction in MAE and the concurrent improvement in PSNR and SSIM scores.

TABLE III

COMPARATIVE RESULTS OBTAINED ON TWO DATASETS WITH/WITHOUT PEB FOR WRAPPED PHASE RETRIEVAL

Method	DataSet1		DataSet2	
	MAE	PSNR/SSIM	MAE	PSNR/SSIM
w/o PEB	0.276	18.72/0.7984	0.317	21.33/0.6222
w PEB	**0.0366**	**35.63/0.9334**	**0.0326**	**35.32/0.9575**

TABLE IV

COMPARATIVE RESULTS OBTAINED ON TWO DATASETS WITH/WITHOUT ECL FOR WRAPPED PHASE RETRIEVAL

Method	DataSet1		DataSet2	
	MAE	PSNR/SSIM	MAE	PSNR/SSIM
w/o ECL	0.0424	34.5/0.9247	**0.0287**	**36.57**/0.9499
w ECL	**0.0366**	**35.63/0.9334**	0.0326	35.32/**0.9575**

TABLE V

COMPARATIVE RESULTS OBTAINED ON TWO DATASETS WITH DIFFERENT ACTIVATION FUNCTIONS FOR WRAPPED PHASE RETRIEVAL

Activation Function	DataSet1		DataSet2	
	MAE	PSNR/SSIM	MAE	PSNR/SSIM
ReLU	**0.0366**	**35.63/0.9334**	**0.0326**	35.32/**0.9575**
PReLU	0.0432	34.43/0.9236	0.0335	**36.09**/0.9480
RReLU	0.6036	15.40/0.5039	0.3667	19.94/0.5750

4) Hierarchical Phase Inpainting Unit (HPIU): The HPIU stands as a pivotal functional module within PIN, strategically employed to extract essential phase information and globally redistribute key weights. An essential operation within HPIU is referred to as enhanced concatenate learning (ECL). When it comes to wrapped phase retrieval, the design and implementation of ECL undeniably contribute to improved accuracy, as evident from Table IV. Primarily, when evaluated on Dataset 1 (ours), the incorporation of ECL resulted in a comprehensive enhancement across various performance metrics. However, Dataset 2 did not exhibit a significant improvement or enhancement. This observation underscores the high compatibility and adaptability of ECL within the PIN framework, while also emphasizing the importance of tailoring network design to specific application scenarios.

5) Activation Function: As indicated in Table V, various activation functions exhibit distinct impacts on the reconstruction outcomes. We conducted a replacement of the activation function within the PIN (excluding EAB) and compared commonly used ones, namely, ReLU, parametric rectified linear unit (PReLU), and randomized leaky rectified linear unit (RReLU), leading to the following conclusions. First, in general, the optimization of ReLU yields the most satisfactory results. This improvement is particularly noteworthy in the case of Dataset 1, resulting in a substantial enhancement across all evaluation metrics. Second, the choice of activation functions should be tailored to specific tasks. While the utilization of RReLU within EAB (at the local level) enhances the network's reconstruction outcomes, it performs the worst within the global PIN. This discrepancy can be attributed, to some extent, to the varying task and functionality requirements at different levels (e.g., pixel level, local level, and global level) within the PIN architecture.

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

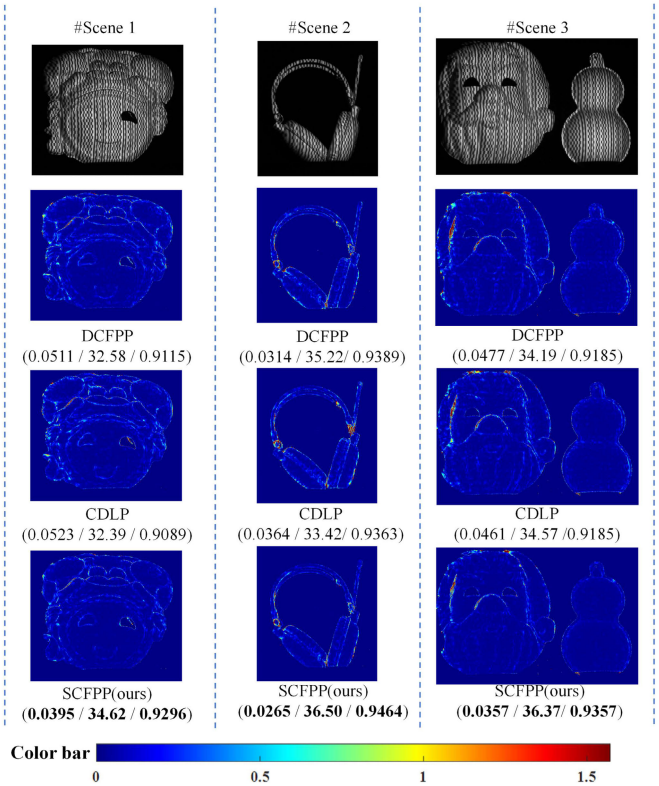


Fig. 12. Calculated wrapped phase maps and errors on Dataset 1. The evaluation index (MAE/PSNR/SSIM) for each error map is listed below the image.

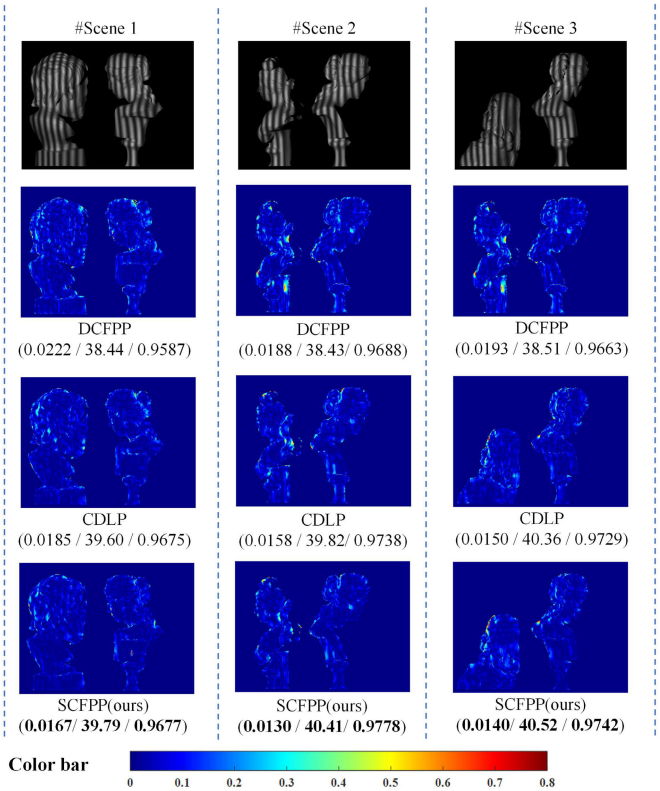


Fig. 13. Calculated wrapped phase maps and errors on Dataset 2. The evaluation index (MAE/PSNR/SSIM) for each error map is listed below the image.

TABLE VI

MEASUREMENT RESULTS OF THREE METHODS ON TWO DATASETS

Method	DataSet1		DataSet2	
	MAE	PSNR/SSIM	MAE	PSNR/SSIM
CDLP	0.0460	34.02/0.9185	0.0374	35.06/0.9525
DCFPP	0.0460	33.97/0.9194	0.0345	35.08/0.9459
SCFPP(ours)	**0.0366**	**35.63/0.9334**	**0.0326**	**35.32/0.9575**

C. Comparative Experiment

Two deep-learning-based single-frame FPP algorithms, i.e., CDLP [31] and DCFPP [32], which have shown their better performances than that of FTP, were used to comparatively demonstrate the advantages of the proposed algorithm. In this experiment, the accuracy of wrapped phase maps calculated by these three algorithms was assessed to compare their performance. Table VI displays the statistical measurement results for three methods on two datasets, with optimal evaluation index values highlighted. It is evident from the table that the proposed SCFPP consistently outperforms all other methods, achieving the best performance across both Dataset 1 and Dataset 2.

In Dataset 1, three typical measurement objects were selected: complex masks, headphones, and an isolated scene. The spatial composite fringe patterns of these three objects are presented in the first row of Fig. 12. After obtaining the deformed fringe images, wrapped phase information was obtained by inputting them into a pretrained CNN. The second row of Fig. 12 shows the wrapped phase error plots obtained by DCFPP for three scenarios. The MAE for the errors is: 0.0511, 0.0314, and 0.0477 rad, respectively. The PSNR for the predicted wrapped phase are: 32.58, 35.22, and 34.19, while the SSIM are: 0.9115, 0.9389, and 0.9185, respectively. The third row of Fig. 12 demonstrates the wrapped phase error plots computed by CDLP, with MAE values of 0.0523, 0.0364, and 0.0461 rad, respectively. The PSNR values for the wrapped phase are: 32.39, 33.42, and 34.57, while the SSIM values are: 0.9089, 0.9363, and 0.9185, respectively. CDLP exhibits smaller MAE values, and all three evaluation metrics surpass those of SPU. The last row of Fig. 12 showcases the error plots obtained by SCFPP, with MAE values of 0.0395, 0.0265, and 0.0357 rad, respectively. The PSNR values for the truncated phase are: 34.62, 36.50, and 36.37, while the SSIM values are: 0.9296, 0.9464, and 0.9357, respectively. In terms of error evaluation metrics, it becomes evident that the proposed method outperforms the other two approaches in terms of MAE, PSNR, and SSIM. From the phase error maps, a notable observation emerges: DCFPP and CDLP exhibit substantial phase errors along certain object edges, whereas SCFPP consistently maintains a comparatively lower overall phase error level. In addition, SCFPP demonstrates higher edge accuracy compared with DCFPP and CDLP.

In Dataset 2, we specifically selected three distinct scenarios to validate the performance of three different algorithms. In Fig. 13, the first row displays spatial composite fringe patterns for these measurement scenarios. The second row illustrates the error plot for the wrapped phases predicted by

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

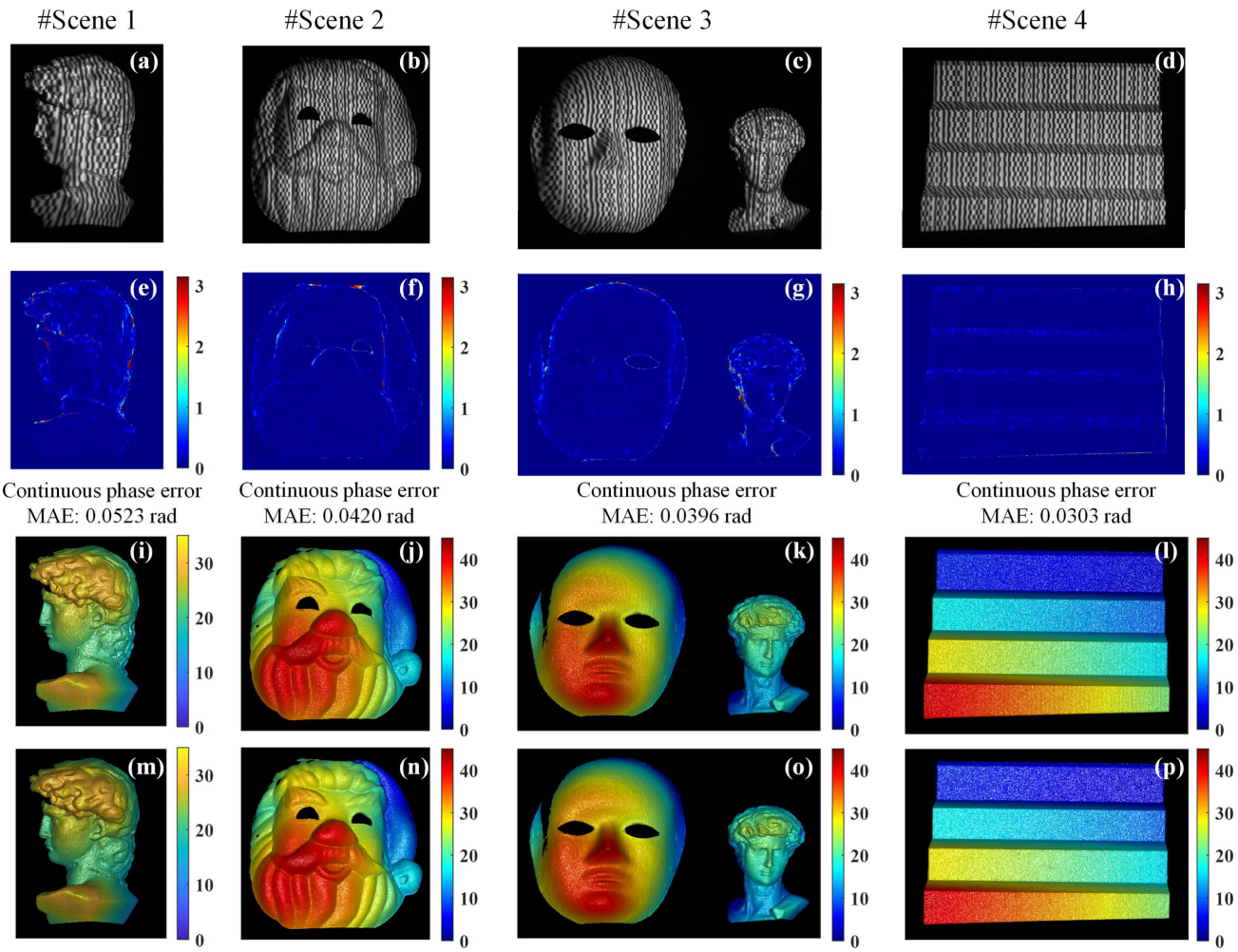


Fig. 14. (a)–(d) Spatial composite fringe patterns of four scenes. (e)–(h) Continuous phase error maps. (i)–(l) Reconstruction results of ground truth. (m)–(p) Reconstruction results of SCFPP.

the DCFPP algorithm, with corresponding MAE values of 0.0222, 0.0188, and 0.0193 rad. Furthermore, the PSNR values for these wrapped phase maps are 38.44, 38.43, and 38.51, and the SSIM values are 0.9587, 0.9688, and 0.9663, respectively. The third row of Fig. 13 showcases the phase error maps predicted by the CDLP algorithm, featuring MAE values of 0.01185, 0.0158, and 0.0150. The corresponding PSNR values for the wrapped phase maps are 39.60, 39.82, and 40.36, while the SSIM values are 0.9675, 0.9738, and 0.9729, respectively. Finally, in the last row of Fig. 13, we present the phase error plots generated by the proposed SCFPP algorithm, with MAE values of 0.0167, 0.0130, and 0.0140. The PSNR values for the wrapped phase maps are 39.79, 40.41, and 40.52, and the SSIM values are 0.9677, 0.9778, and 0.9742. It is noteworthy that all three methods yield small phase errors, but SCFPP consistently demonstrates superior performance.

D. Reconstruction Result

After demonstrating the superior performance of the SCFPP, we applied the proposed algorithm to reconstruct four distinct measurement scenarios, encompassing statues, intricate masks, isolated scenes, and step-like objects. In Fig. 14(a)–(d), we present the spatial composite fringe patterns for these four scenarios. These patterns were then processed using a pretrained PIN to convert them into wrapped phase information. Subsequently, the CTHPU was employed to mitigate phase blurring and recover the 3-D shape of the objects. To provide a more intuitive representation of the reconstruction accuracy, we employed the continuous phase calculated through the phase-shifting method and CTHPU as the ground truth, allowing us to compute the continuous phase error. Fig. 14(e)–(h) displays the resulting continuous phase error plots for the four scenes. The MAE of four error maps is 0.0523, 0.0420, 0.0396, and 0.0303 rad, respectively. Fig. 14(i)–(l) and (m)–(p) depicts the reconstruction outcomes for ground truth and SCFPP, respectively. These results illustrate that, even in complex and isolated scenarios, SCFPP consistently delivers reconstructions that closely resemble the ground truth. This reconstruction experiment conclusively demonstrates SCFPP's ability to reconstruct complex and isolated scenes, thereby reaffirming its exceptional performance.

V. Conclusion

In this work, we propose a novel spatial multiplexing algorithm that incorporates a PIN for calculating wrapped phases and employs CTHPU for phase unwrapping. The

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

proposed SCFPP method utilizes ASES to combine different information in the intensity domain, while ensuring measurement accuracy by appropriately selecting the size of the composite block. We conducted network training using 900 datasets and achieved single-frame measurements by applying the trained network in conjunction with the CTHPU algorithm. Compared with traditional compounding methods, spatial compounding ensures that each pixel contains information from only one frequency, thereby leading to higher phase accuracy.

In our future work, we will concentrate on two primary objectives.

1) Develop a swift and efficient algorithm for determining the basic block size. In SCFPP, we ascertain the basic block size through ablation experiments. These experiments have revealed a correlation between the basic block size and optical system parameters, such as camera and projector resolutions. Therefore, our forthcoming research will delve into understanding the relationship between optical system parameters and basic block size, ultimately leading to the creation of an algorithm for calculating basic block size.
2) Explore the application of traditional algorithms for demodulating spatial composite fringe patterns. The integration of the deep-learning technology has notably enhanced the accuracy and efficiency of FPP. On the other side, the physical models have been successfully used in other compound methods for higher accuracy. In our method, when projecting spatial composite fringe pattern to the object, each basic block within the fringe images undergoes a transformation in its mapping relationship with the original image pixels. This transformation complicates the task of distinguishing the composite fringe images with varying frequency components, a crucial limitation of traditional demodulation algorithms. Through simulation experiments, we have substantiated that segregating different information components allows for the use of conventional image restoration algorithms to recover complete information and enable 3-D reconstruction. Therefore, our future work will focus on build the physical model to identify the mapping relationship between composite fringe images and further designing the physics-based convolution neural networks.

REFERENCES

[1] S. S. Gorthi and P. Rastogi, "Fringe projection techniques: Whither we are?" *Opt. Lasers Eng.*, vol. 48, no. 2, pp. 133–140, Feb. 2010.

[2] P. Yao, S. Gai, and F. Da, "Toward real-world super-resolution technique for fringe projection profilometry," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–8, 2022.

[3] J. Zhu, C. Zhu, P. Zhou, Z. He, and D. You, "An optimizing diffusion kernel-based binary encoding strategy with genetic algorithm for fringe projection profilometry," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–8, 2022.

[4] S. Zhang, "High-speed 3D shape measurement with structured light methods: A review," *Opt. Lasers Eng.*, vol. 106, pp. 119–131, Jul. 2018.

[5] Z. Song, H. Jiang, H. Lin, and S. Tang, "A high dynamic range structured light means for the 3D measurement of specular surface," *Opt. Lasers Eng.*, vol. 95, pp. 8–16, Aug. 2017.

[6] C. Sun and X. Zhang, "Real-time subtraction-based calibration methods for deformation measurement using structured light techniques," *Appl. Opt.*, vol. 58, no. 28, p. 7727, 2019.

[7] S. Zhang, "Recent progresses on real-time 3D shape measurement using digital fringe projection techniques," *Opt. Lasers Eng.*, vol. 48, no. 2, pp. 149–158, Feb. 2010.

[8] S. Van der Jeught and J. J. J. Dirckx, "Real-time structured light profilometry: A review," *Opt. Lasers Eng.*, vol. 87, pp. 18–31, Dec. 2016.

[9] C. Zuo et al., "High-speed three-dimensional shape measurement for dynamic scenes using bi-frequency tripolar pulse-width-modulation fringe projection," *Opt. Lasers Eng.*, vol. 51, no. 8, pp. 953–960, Aug. 2013.

[10] Z. Qin, Q. Zeng, Y. Zong, and F. Xu, "Image inpainting based on deep learning: A review," *Displays*, vol. 69, Sep. 2021, Art. no. 102028.

[11] M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," *J. Opt. Soc. Amer.*, vol. 72, no. 1, pp. 156–160, 1982.

[12] X. Su and W. Chen, "Fourier transform profilometry: A review," *Opt. Lasers Eng.*, vol. 35, no. 5, pp. 263–284, 2001.

[13] J. Qian, T. Tao, S. Feng, Q. Chen, and C. Zuo, "Motion-artifact-free dynamic 3D shape measurement with hybrid Fourier-transform phase-shifting profilometry," *Opt. Exp.*, vol. 27, no. 3, pp. 2713–2731, 2019.

[14] L.-C. Chen, H.-W. Ho, and X.-L. Nguyen, "Fourier transform profilometry (FTP) using an innovative band-pass filter for accurate 3-D surface reconstruction," *Opt. Lasers Eng.*, vol. 48, no. 2, pp. 182–190, Feb. 2010.

[15] Z. H. Zhang, "Review of single-shot 3D shape measurement by phase calculation-based fringe projection techniques," *Opt. Lasers Eng.*, vol. 50, no. 8, pp. 1097–1106, Aug. 2012.

[16] H. Wu, Y. Cao, H. An, C. Xu, and H. Li, "Ultrafast spatial phase unwrapping algorithm with accurately correcting transient phase error," *Opt. Lett.*, vol. 46, no. 24, pp. 6091–6094, 2021.

[17] K. Chen, J. Xi, and Y. Yu, "Quality-guided spatial phase unwrapping algorithm for fast three-dimensional measurement," *Opt. Commun.*, vol. 294, pp. 139–147, May 2013.

[18] S. Zhang, X. Li, and S.-T. Yau, "Multilevel quality-guided phase unwrapping algorithm for real-time three-dimensional shape reconstruction," *Appl. Opt.*, vol. 46, no. 1, pp. 50–57, 2007.

[19] S. Zhang, "Absolute phase retrieval methods for digital fringe projection profilometry: A review," *Opt. Lasers Eng.*, vol. 107, pp. 28–37, Aug. 2018.

[20] P. S. Huang, C. Zhang, and F.-P. Chiang, "High-speed 3-D shape measurement based on digital fringe projection," *Opt. Eng.*, vol. 42, no. 1, pp. 163–168, 2003.

[21] H. J. Chen, J. Zhang, D. J. Lv, and J. Fang, "3-D shape measurement by composite pattern projection and hybrid processing," *Opt. Exp.*, vol. 15, no. 19, pp. 12318–12330, 2007.

[22] W.-H. Su, "Color-encoded fringe projection for 3D shape measurements," *Opt. Exp.*, vol. 15, no. 20, pp. 13167–13181, 2007.

[23] Z. Zhang, D. P. Towers, and C. E. Towers, "Snapshot color fringe projection for absolute three-dimensional metrology of video sequences," *Appl. Opt.*, vol. 49, no. 31, pp. 5947–5953, 2010.

[24] C. Guan, L. G. Hassebrook, and D. L. Lau, "Composite structured light pattern for three-dimensional video," *Opt. Exp.*, vol. 11, no. 5, pp. 406–417, 2003.

[25] M. Takeda, Q. Gu, M. Kinoshita, H. Takai, and Y. Takahashi, "Frequency-multiplex Fourier-transform profilometry: A single-shot three-dimensional shape measurement of objects with large height discontinuities and/or surface isolations," *Appl. Opt.*, vol. 36, no. 22, pp. 5347–5354, 1997.

[26] G. Sansoni and E. Redaelli, "A 3D vision system based on one-shot projection and phase demodulation for fast profilometry," *Meas. Sci. Technol.*, vol. 16, no. 5, pp. 1109–1118, May 2005.

[27] C. Zuo et al., "Deep learning in optical metrology: A review," *Light, Sci. Appl.*, vol. 11, no. 1, p. 39, Feb. 2022.

[28] S. Feng et al., "Fringe pattern analysis using deep learning," *Adv. Photon.*, vol. 1, no. 2, p. 1, Feb. 2019, Art. no. 025001.

[29] H. Yu et al., "Dynamic 3-D measurement based on fringe-to-fringe transformation using deep learning," *Opt. Exp.*, vol. 28, no. 7, pp. 9405–9418, 2020.

[30] J. Qian et al., "Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement," *APL Photon.*, vol. 5, no. 4, Apr. 2020, Art. no. 046105.

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.

[31] Y. Li, J. Qian, S. Feng, Q. Chen, and C. Zuo, "Composite fringe projection deep learning profilometry for single-shot absolute 3D shape measurement," *Opt. Exp.*, vol. 30, no. 3, pp. 3424–3442, 2022.

[32] Y. Li, J. Qian, S. Feng, Q. Chen, and C. Zuo, "Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3D shape measurement," *Opto-Electron. Adv.*, vol. 5, no. 5, p. 210021, 2022.

[33] J. M. Qian et al., "Single-shot absolute 3D shape measurement with deep-learning-based color fringe projection profilometry," *Opt. Exp.*, vol. 45, no. 7, pp. 1842–1845, 2020.

[34] Z.-W. Li, Y.-S. Shi, C.-J. Wang, D.-H. Qin, and K. Huang, "Complex object 3D measurement based on phase-shifting and a neural network," *Opt. Commun.*, vol. 282, no. 14, pp. 2699–2706, Jul. 2009.

[35] H. Nguyen, Y. Wang, and Z. Wang, "Single-shot 3D shape reconstruction using structured light and deep convolutional neural networks," *Sensors*, vol. 20, no. 13, p. 3718, Jul. 2020.

[36] Y. Zheng, S. Wang, Q. Li, and B. Li, "Fringe projection profilometry by conducting deep learning from its digital twin," *Opt. Exp.*, vol. 28, no. 24, pp. 36568–36583, 2020.

[37] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[38] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 603–612.

Yansong Jiang received the B.S. degree from the College of Science, China Three Gorges University, Yichang, China, in 2020, and the M.S. degree from Sichuan University, Chengdu, China, in 2023, where he is currently pursuing the Ph.D. degree.

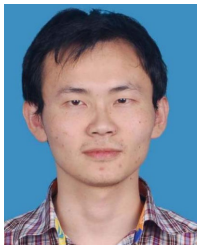
His research interests include image processing, deep learning, and structure light.

Jiayi Qin received the B.Sc. degree in information security and the M.S. degree in information and communication engineering from Sichuan University, Chengdu, China, in 2020 and 2023, respectively, where she is currently pursuing the Ph.D. degree in optical engineering.

Her research interests include 3-D shape measurement, fringe projection profilometry, deep learning, and image processing.

Yuankun Liu received the B.S., M.S., and Ph.D. degrees from Sichuan University, Chengdu, China, in 1998, 2004, and 2007, respectively.

He is currently an Associate Professor with the College of Electronics and Information Engineering, Sichuan University. His research interests include optical sensing, optical information processing, and machine vision.



Menglong Yang (Member, IEEE) received the B.S. degree from the School of Chemical Engineering, Sichuan University, Chengdu, China, in 2005, and the M.S. degree from the School of Computer Science and Engineering, Sichuan University, in 2008.

He is currently a Professor with the School of Aeronautics and Astronautics, Sichuan University. He is also an Engineer with Wisesoft Company Ltd., Chengdu, Sichuan. He has authored or co-authored more than 20 journal articles. His research interests include computer vision, pattern recognition, and machine learning.

Yiping Cao received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1984, the M.S. degree from the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, China, in 1988, and the Ph.D. degree in optics from Sichuan University, Chengdu, in 2003.

He is currently a Professor with the College of Electronics and Information Engineering, Sichuan University. His research interests include optical 3-D sensing, optical information processing, digital manufacturing, and optomechatronics.

Authorized licensed use limited to: GuangXi Normal University. Downloaded on September 09,2024 at 11:34:46 UTC from IEEE Xplore. Restrictions apply.