Super-Resolution Phase Retrieval Network for Single-Pattern Structured Light 3D Imaging

Jianwen Song, Kai Liu, *Senior Member, IEEE*, Arcot Sowmya, and Changming Sun

***Abstract*— Structured light 3D imaging is often used for obtaining accurate 3D information via phase retrieval. Single-pattern structured light 3D imaging is much faster than multi-pattern versions. Current phase retrieval methods for single-pattern structured light 3D imaging are however not accurate enough. Besides, the projector resolution in a structured light 3D imaging system is expensive to improve due to hardware costs. To address the issues of low accuracy and low resolution of single-pattern structured light 3D imaging, this work proposes a super-resolution phase retrieval network (SRPRNet). Specifically, a phase-shifting module is proposed to extract multi-scale features with different phase shifts, and a refinement and super-resolution module is proposed to obtain refined and super-resolution phase components. After phase demodulation and unwrapping, high-resolution absolute phase is obtained. A sine shifting loss and a cosine shifting loss are also introduced to form the regularization term of the loss function. As far as can be ascertained, the proposed SRPRNet is the first network for super-resolution phase retrieval by using a single pattern, and it can also be used for standard-resolution phase retrieval. Experimental results on three datasets show that SRPRNet achieves state-of-the-art performance on 1×, 2×, and 4× super-resolution phase retrieval tasks.**

***Index Terms*— Structured light, super-resolution, single-pattern, phase retrieval, phase-shifting.**

I. Introduction

STRUCTURED light (SL) three-dimensional (3D) imaging [1] is one of the most efficient active 3D vision techniques with outstanding accuracy. The principle of SL 3D imaging is similar to that of binocular stereo vision [2] and is based on triangulation. An SL system uses a projector that actively projects coded patterns onto objects, and the projection operation not only simplifies the matching process but also has superior ability to reduce noise and other error sources when compared to binocular stereo vision. Phase information is obtained from the patterns captured by the camera after the projector projects the coded patterns onto the object. By combining the phase information with the calibration parameters of the projector-camera system, the 3D coordinates of points on the object surface can be obtained. Therefore, the key issue in SL 3D imaging is the retrieval of phase information.

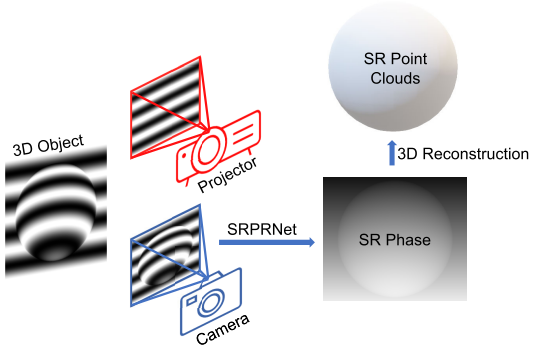


Fig. 1. Diagram of super-resolution single-pattern structured light 3D imaging.

Traditional SL 3D imaging methods involve various coding and decoding strategies to obtain the phase information. They can be classified into multi-pattern and single-pattern methods. Multi-pattern SL methods apply codification [1] or phase-shifting [3] via a series of patterns and then use the deformed patterns to decode the phase information. For codification-based multi-pattern SL methods, each point is assigned a unique code that is combined from all the patterns, where the code can be binary or gray-level values [4], [5], [6]. For phase-shifting multi-pattern SL methods, intensity variation information is coded in multiple patterns with phase shifts. Such intensity variations can be in the form of sinusoidal [7], [8], [9], [10], [11], [12], binary [13], [14], triangular [15], [16], hybrid [17], [18], or other shapes. Phase-shifting SL is the most representative method of SL because of its robustness and efficiency, and the sinusoidal pattern has the desirable properties of continuity and periodicity, and these properties make the sinusoidal pattern the most popular pattern for SL 3D imaging.

However, the use of multiple patterns requires long scanning time, which makes it challenging to use in a real-time system. Decreasing the number of projected patterns is the most straightforward way to improve the real-time capability of an SL system. Therefore, single-pattern SL methods are designed to speed up the scanning process. Apart

Manuscript received 30 March 2022; revised 26 October 2022 and 5 December 2022; accepted 8 December 2022. Date of publication 22 December 2022; date of current version 4 January 2023. The work of Jianwen Song was supported by the Australian Government Research Training Program Scholarship. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lisimachos Paul Kondi. *(Corresponding author: Changming Sun.)*

Jianwen Song is with the School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia, and also with CSIRO Data61, Epping, NSW 1710, Australia.

Kai Liu is with the College of Electrical Engineering, Sichuan University, Chengdu 610065, China.

Arcot Sowmya is with the School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia.

Changming Sun is with CSIRO Data61, Epping, NSW 1710, Australia, and also with the School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: changming.sun@csiro.au).

Digital Object Identifier 10.1109/TIP.2022.3230245

from color-based single-pattern SL, which combines multiple phase-shifting information within a single color-coded pattern, other single-pattern methods can be classified into Fourier-based [19] and indexing-based methods [20]. Fourier-based SL methods retain the principal components and remove other components in the frequency domain while retrieving phase information. Indexing-based SL methods use unique strips or points to construct the projected patterns and can distinguish different positions on the captured patterns based on the specific coding method when decoding the patterns. However, these single-pattern SL methods have low decoding accuracy.

Apart from the coding and decoding processes, some other issues are also important for improving the accuracy and efficiency of SL 3D imaging. Traditional methods are based on constructing various error models to improve accuracy. Recently, with the development of deep learning techniques, deep neural networks have been applied to tasks in SL 3D imaging, such as phase retrieval [21], [22], gamma distortion elimination [23], light saturation correction [24], intensity enhancement [25], and phase unwrapping [26].

To achieve high-quality phase retrieval for single-pattern SL 3D imaging, a novel super-resolution phase retrieval network (SRPRNet) is proposed in this work. The diagram of this work is shown in Fig. 1. The projector projects a single pattern onto the 3D object and the camera captures the deformed single low-resolution (LR) pattern. Then, SRPRNet retrieves the super-resolution (SR) phase from the single LR pattern. Finally, we can obtain SR point clouds using the SR phase. The core components of the network are a phase-shifting module (PSM) and a refinement and super-resolution module (RSRM). With the proposed SRPRNet, an SR phase map can be obtained using a single LR input pattern. Besides, SRPRNet can also be extended for standard-resolution (STR) phase retrieval. The main contributions of this work are as follows:

1) A phase-shifting module consisting of 4 shift blocks and 1 fusion block is designed for extracting features with information from different phase shifts.
2) A refinement and super-resolution module is designed to refine the phase-shifting features and generate SR phase components.
3) A novel SR phase retrieval network, namely SRPRNet, for single-pattern SL 3D imaging is proposed. Besides, the SR module can be integrated with any existing methods to form new SR phase retrieval methods for single-pattern SL 3D imaging, which further demonstrates the effectiveness and superiority of the proposed network.
4) A new dataset that contains 147 fringe patterns and phase components pairs is constructed for training phase retrieval networks, in addition to existing publicly available datasets.

The rest of the paper is organized as follows. Related work is reviewed in Section II. In Section III, the basic principles of phase-shifting SL are briefly introduced. The proposed SRPRNet is described in Section IV, and Section V provides the experimental results. Conclusion and future work are described in Section VI.

II. Related Work

Single-pattern SL methods are much faster than multi-pattern SL methods not only during scanning but also during computation. Traditional single-pattern SL methods decode a sinusoidal pattern based on Fourier analysis, or decode a uniquely labeled pattern with grids or strips according to specific coding methods. These methods are useful but have low accuracy. Deep learning techniques can also be used for phase retrieval in SL 3D imaging and can provide better performance than traditional methods. The projector resolution in an SL system limits the final resolution, and improving the resolution on hardware is expensive. It should be easier to improve the resolution of an SL system using a computational process, which is the motivation for this work.

A. Traditional Single-Pattern Structured Light Methods

Fourier profilometry [19], windowed Fourier profilometry [27], and wavelet transform profilometry [28] are three representative traditional transform-based single-pattern SL methods. These methods remove high-frequency components of the captured pattern in the frequency domain. Then, the phase information can be obtained using the remaining components. Li et al. [29] proposed a phase retrieval method that used an advanced shearlet transform to extract the fundamental frequency component of the single pattern. Zhu et al. [30] presented an image decomposition model, named TV-G-Shearlet, to remove noise and background from a single fringe pattern, and then the wrapped phase was obtained using Fourier transform. Dong and Chen [31] proposed an advanced Fourier transform method to retrieve phase information from a single spatial pattern. The method split the pattern into four shifted patterns with one-pixel difference, and Fourier transform was used to obtain Fourier spectra from the four patterns. Then, by subtraction between the Fourier spectra, the 0th harmonic component was filtered, and phase information was obtained.

A three-channel color-coded pattern [32] can be used to embed phase shifts or other features into a single pattern. Lin et al. [33] proposed a single-pattern SL method based on coding geometric information using color channels in one pattern. By using a two-step decoding process consisting of color decoding and geometric decoding, this method achieved high-quality 3D reconstruction. Budianto et al. [34] proposed a robust color-coded single-pattern SL method. They implemented an enhanced morphological component analysis method to separate texture and fringe patterns from a single RGB fringe pattern, and this method achieved better performance than the traditional single-pattern methods. Zhang et al. [35] combined two fringe patterns with a phase shift of π into the red and blue channels to form a single color-coded pattern. Then, the captured color-coded pattern can be used to extract two Moiré patterns [36] for retrieving phase information.

Strip or grid indexing SL methods [37] utilize uniquely coded strips or grid features to obtain the correspondences between the camera and projector spaces. Petković et al. [38] proposed a self-equalizing De Bruijn sequence method that eliminated influences such as ambient lighting and object

albedo. Wang and Yang [39] designed a single-line method to segment and cluster the single-line pattern for retrieving the LR phase map and then interpolated the LR phase map to obtain a full-resolution map.

Although Fourier transform-based single-pattern SL methods can successfully obtain phase information from a single pattern, their accuracy is much lower than that of multiple-pattern methods. While color-coded single-pattern SL methods can embed phase-shifting information into different color channels, it cannot be guaranteed that the color information from a projector is sufficiently accurate.

B. Deep Learning-Based Single-Pattern Structured Light Methods

Recently, deep learning techniques have been used for single-pattern SL 3D imaging. Feng et al. [21] first used a deep neural network to train a model for retrieving phase components that are then used for demodulating the phase from a single pattern. Yao et al. [40] trained a similar network to that of Feng et al. to retrieve the wrapped phase and achieved phase unwrapping from two extra patterns. Following Feng et al.'s idea, several other convolutional neural networks (CNNs) [41], [42] have been trained to obtain the phase components from a single pattern. Jeught and Dirckx [43] presented a CNN approach to extract height information directly from a single pattern, and this network was also used for tasks such as noise reduction and phase unwrapping. Nguyen et al. [22] trained a U-Net to extract depth for single-pattern SL 3D imaging. Zheng et al. [44] proposed a digital twin fringe dataset generation method based on deep learning and used the generated dataset to train a U-Net to obtain a depth map. Several encoder-decoder networks [45] or U-Nets [46] have been trained to retrieve phase information from a single pattern. Machineni et al. [47] designed a two-stage framework consisting of a probabilistic weighted synthesizer network for estimating the reference fringe patterns and a multi-resolution similarity assessment network for retrieving depth from the reference and deformed patterns. Yuan et al. [48] proposed a phase demodulation method for a single-frame interferogram based on a network combining a U-Net with dense blocks. This network can obtain a normalized wrapped phase using a normalized interferogram. Qian et al. [49] proposed a single-shot color-coded SL 3D imaging method based on deep learning. They used three shifted patterns extracted from a single color-coded pattern as inputs and trained the network to obtain the phase components.

Although these deep learning-based methods achieve better performance than traditional transform-based or color-coded methods, they only utilize common neural network architectures for phase retrieval and depth estimation tasks, and there is still much scope for improvement. Besides, direct estimation of depth from fringe patterns discards some information in SL. In addition, several previous methods use simulated patterns to train the networks, which may cause instability when tested on real scenes.

C. Super-Resolution Structured Light Methods

Limited by the projector resolution, improving the resolution of an SL system in hardware is expensive. SR techniques based on either hardware or software have been proposed to improve the resolution of an SL system. Kil et al. [50] proposed an SR laser scanning method that applied image SR with multiple scans. In their work, the collected multiple laser scans had random offsets so that each image would provide different information to the final model. Ouji et al. [51] proposed a 3D space-time non-rigid SR scanning method that used three calibrated cameras and an uncalibrated projection device. It was a hybrid stereo vision and phase-shifting method that used two shifted fringe patterns and one texture image. This method not only incorporated the advantages of SL and stereo vision but also overcame their shortcomings. Weinmann et al. [52] proposed a multi-camera and multi-projector SR SL framework. The system was used to scan scenes from different viewpoints, and the collected information in the form of multiple SL depth maps from different viewpoints was combined to perform SR reconstruction and obtain denser point clouds. Shiba et al. [53] proposed a multi-pattern SR SL method that captured multiple patterns to perform temporal SR on the depth maps.

These SR SL techniques work in a multi-sensor or multi-map fusion setup. They require expensive hardware or longer processing time. Currently there are no SR techniques that can be used in the phase retrieval process for SL 3D imaging.

D. Guided Depth Super-Resolution Methods

Depth maps generated by depth sensors such as those based on the time-of-flight approach typically have a low resolution because of hardware limitations. However, high-resolution and high-quality RGB images are usually available on depth sensors. Various guided depth map SR methods [54], [55], [56], [57], [58] have been proposed to fuse the details of RGB images to the depth maps for the same captured scenes and thus improve the resolution of the depth maps. Traditional guided depth SR methods [54], [55] use an energy function based on different priors and regularization terms to find the optimized SR depth. With deep learning techniques becoming popular in computer vision tasks, many deep learning-based guided depth SR methods [56], [57], [58] have been proposed to establish the mapping relationship between the LR depth map with the corresponding high-resolution RGB image and the SR depth map. These methods outperform traditional methods.

However, the depth generated using structured light imaging has high precision. Using an RGB image to guide the generation of SR depth may damage such precision severely. Besides, SL 3D imaging is more suitable for imaging objects that are not rich in color information. Under such conditions, the RGB-guided generation of SR depth is not suitable to be applied to SL 3D imaging directly.

III. Phase-Shifting Structured Light

In traditional phase-shifting SL techniques, the projected sinusoidal patterns along the vertical direction (with shift

information along vertical direction and stripes appearing along horizontal direction) are denoted as [12]:

$$I_n^p = A^p + B^p \cos\left[2\pi\left(f\frac{y^p}{H} - \frac{n}{N}\right)\right], \tag{1}$$

where p indicates the projector space, I_n^p represents the pattern intensities at projector coordinate (x^p, y^p), A^p and B^p are constants, f and H are the spatial frequency and the height of the projected pattern in the vertical direction respectively, n is the shift index, and N is the shift step, i.e., the total number of phase shifts. For conciseness, the coordinates (x^p, y^p) for I_n^p have been omitted in Eq. (1). If a horizontal mode is used to code the patterns, H will be substituted with the width of the projected pattern in the horizontal direction W, and y^p will be substituted with x^p. The patterns captured by the camera can be denoted as:

$$I_n^c = A^c + B^c \cos\left(\phi - \frac{2\pi n}{N}\right), \tag{2}$$

where c indicates the camera space, I_n^c represents the intensities of a captured pattern, and A^c, B^c, and ϕ are the average intensities, intensity modulation, and phase related to camera coordinate (x^c, y^c) respectively. To solve the three unknowns A^c, B^c, and ϕ, the smallest number of N for phase-shifting SL is three. With a larger N, the system will have a better ability to reduce errors. As for coordinates (x^p, y^p), the coordinates (x^c, y^c) are omitted in Eq. (2). The most important parameter for obtaining 3D coordinates, i.e., ϕ, can be computed by

$$\phi = \tan^{-1}\left(\frac{S_N}{C_N}\right), \tag{3}$$

where $\tan^{-1}$ is the arctangent operation, and S_N and C_N are two components of the wrapped phase (named as phase components), which can be computed by

$$S_N = \sum_{n=0}^{N-1} I_n^c \sin\left(\frac{2\pi n}{N}\right), \tag{4}$$

and

$$C_N = \sum_{n=0}^{N-1} I_n^c \cos\left(\frac{2\pi n}{N}\right). \tag{5}$$

After obtaining the phase information, the correspondences between the camera and projector for each pixel can be derived, i.e., $y^p = 2\pi/\phi$ for each (x^c, y^c). Then, the 3D coordinates can be obtained by combining the correspondences with the camera and projector parameters.

Parameter A^c can be computed by

$$A^c = \frac{1}{N}\sum_{n=0}^{N-1} I_n^c, \tag{6}$$

and B^c can be computed by

$$B^c = \frac{2}{N}\sqrt{S_N^2 + C_N^2}. \tag{7}$$

A^c is usually used as the texture for the final 3D point clouds and B^c is usually used as the background noise filter [8] so that the noise in shadow areas can be removed with a small B^c.

IV. Super-Resolution Phase Retrieval Network

The architecture of the proposed SRPRNet is shown in Fig. 2. The input is an STR or LR single fringe pattern of size $H \times W$. It is first fed into a PSM which consists of 4 shift blocks and 1 fusion block. The fused features generated by PSM are fed into an RSRM to estimate the STR or SR phase components. After phase demodulation and unwrapping, the final STR or SR absolute phase is obtained.

A. Phase-Shifting Module (PSM)

Traditional sinusoidal phase-shifting SL methods require at least three patterns with uniformly-spaced phase shifts. Training a network to map a single pattern into a wrapped phase is difficult because the distributions and scales are different for the pattern and phase. To address this issue, the proposed network maps a single input pattern into two phase components, i.e., S_N and C_N, rather than the wrapped phase directly. The phase components and the input pattern have similar scales and sinusoidal forms. A PSM is designed to extract features with different phase shifts and then these features are fused to retrieve the phase components. This is different from early research [21], [42] that trained neural networks to map a single pattern into S_N and C_N. Moreover, the PSM is a natural approach to retrieve phase information because it follows the process of computing S_N and C_N using Eq. (4) and Eq. (5).

In PSM, the single pattern $\mathbf{I} \in \mathbb{R}^{H\times W}$ is first fed into the first shift block consisting of a 3×3 convolution and a residual dense block (RDB) [59]. The RDB consists of 4 convolutions with a growth rate of 32 for extracting dense features and 1 convolution for fusing features. The activation function used in RDB is leaky ReLU. The number of branches in the first to the fourth shift block progressively increases from 1 to 4. Features generated by the first shift block are denoted as $\mathbf{F}^1 = CR_1^1(\mathbf{I})$ where CR_j^i is the jth combination of convolution and RDB (jth branch) in the ith shift block and $\mathbf{F}$ is in $\mathbb{R}^{H\times W\times C}$. The first shift block just extracts the features with zero shift information corresponding to the pattern represented by Eq. (2) with $n = 0$. Therefore, the features are close to the input pattern and single-scale feature extraction is used.

Then, $\mathbf{F}^1$ is fed into the second shift block that consists of two branches (with two scales). The first branch has the same resolution as $\mathbf{F}^1$, and then through a similar convolution followed by an RDB, the features generated are used as partial input to the next shift block. The second branch downsamples the input $\mathbf{F}^1$ to half resolution and then the features are fed into another convolution followed by an RDB. The output features are used as partial input to the third shift block. Simultaneously, the output features of the second branch are upsampled to the same resolution as the features of the first branch. Finally, features from these two branches are concatenated and denoted as $\mathbf{F}^2 = \text{Concat}\left[CR_1^2(\mathbf{F}^1), CR_2^2(\mathbf{F}^1\downarrow_2)\uparrow_2\right]$, where Concat represents a concatenation operation, $\downarrow_s$ is $1/s$ bicubic downsampling operation (s is a scale factor), $\uparrow_s$ is s times bicubic upsampling operation, and $\mathbf{F}^2$ is in $\mathbb{R}^{H\times W\times 2C}$. This second shift block extracts features with phase shift indexed 1 corresponding to the pattern with $n = 1$ in Eq. (2). This block

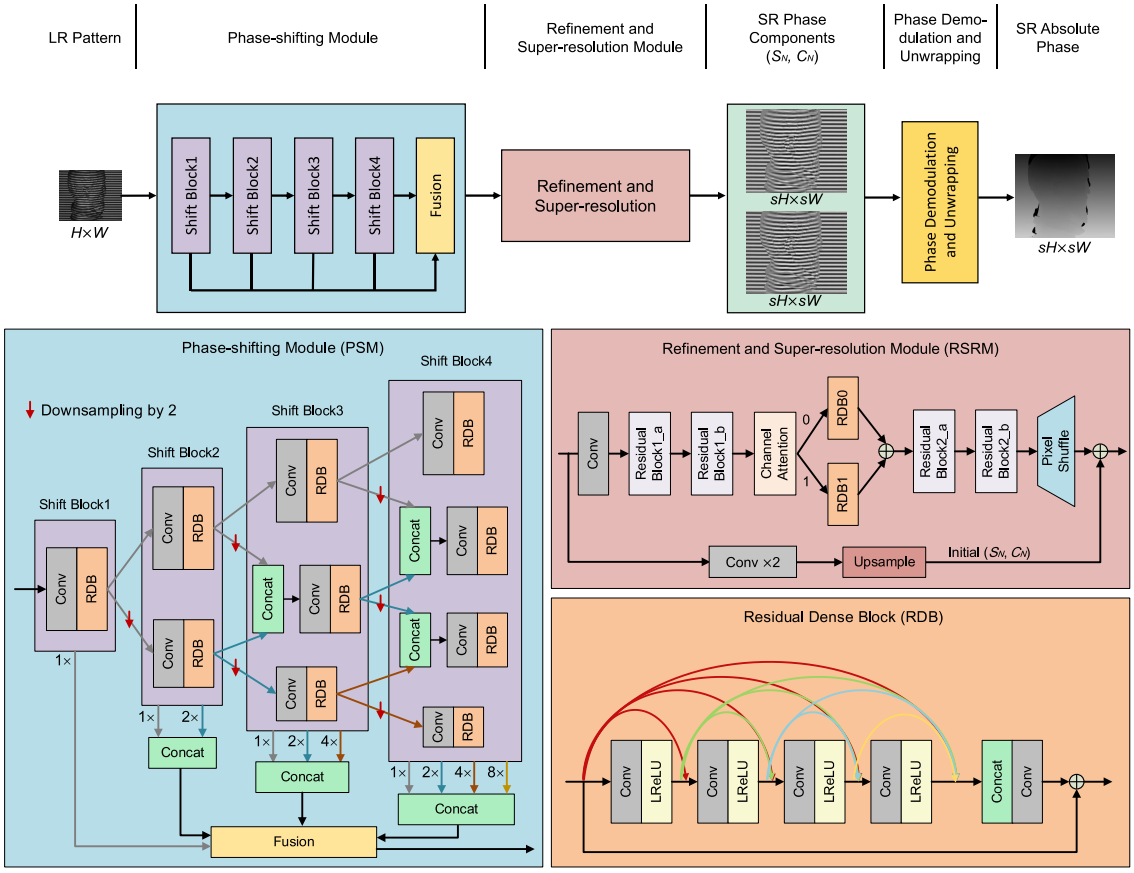


Fig. 2. Overall diagram of the proposed SRPRNet. Top: overall pipeline; Bottom: detailed modules for PSM, RSRM, and RDB.

requires more detailed extraction to obtain all the phase shift information. Therefore, we employ feature extraction at two scales to obtain the phase shift information.

As shown in Fig. 2, the third/fourth shift block has three/four branches to better deal with the phase shift and scaling information. Input features for the middle scales of the third and fourth shift blocks consist of two sets of output features from the previous shift blocks, some with and some without downsampling. Features generated by the third shift block is denoted as $\mathbf{F}^3$ and features generated by the fourth shift block can be written as

$$\mathbf{F}^4 = \text{Concat}\left\{CR_1^4\left(\mathbf{F}_1^3\right), CR_2^4\left[\text{Concat}\left(\mathbf{F}_1^3\downarrow_2, \mathbf{F}_2^3\right)\right]\uparrow_2, CR_3^4\left[\text{Concat}\left(\mathbf{F}_2^3\downarrow_2, \mathbf{F}_3^3\right)\right]\uparrow_4, CR_4^4\left(\mathbf{F}_3^3\downarrow_2\right)\uparrow_8\right\}, \tag{8}$$

where $\mathbf{F}_j^i$ represents the jth branch of features in the ith shift block. As the network becomes deeper, mapping the phase shift information from the previous layer into the phase shift information of the next layer is more complex. Therefore, the method extracts features with more scales in the following shift block than in the previous one. In addition, the upsampling and downsampling processes in shift blocks can provide more information for the final SR module. Although the same number of scales can be used for each shift block, it was found that the performance of such a design is worse (see Section V) and requires more parameters.

Each shift block generates a set of features with phase shift information and these features are fused together to obtain the final phase-shifting features denoted as

$$\mathbf{F} = \text{LReLU}\left\{\text{Conv}_{3\times3}\left[\text{Concat}\left(\mathbf{F}^1, \mathbf{F}^2, \mathbf{F}^3, \mathbf{F}^4\right)\right]\right\}, \tag{9}$$

where LReLU represents leaky ReLU, $\text{Conv}_{3\times3}$ represents a 3×3 convolution operation with $10C$ input channels and $2C$ output channels, and $\mathbf{F}$ is in $\mathbb{R}^{H\times W\times 2C}$.

B. Refinement and Super-Resolution Module (RSRM)

RSRM consists of two branches, one for initial coarse phase components estimation and the other for refining and super-resolving the features. The initial coarse phase component estimation branch (as shown in the lower part of RSRM in Fig. 2) consists of two 3×3 convolution operations that change the feature shapes from $H\times W\times 2C$ to $H\times W\times C$ and to $H\times W\times 2$ and a bicubic interpolation as an upsample block that upsamples the features to shape $sH\times sW\times 2$, yielding the initial phase components $\mathbf{SC}_{\text{ini}}\uparrow_s \in \mathbb{R}^{sH\times sW\times 2}$ where s is the upsampling factor.

The branch for refinement and super-resolution (as shown in the upper part of RSRM in Fig. 2) is used to generate the residual phase components. $\mathbf{F}$ is first fed into a 3×3 convolution that maps the features from $2C$ channels to C channels. Then, the features with reduced channels are fed into two residual blocks (RBs) [60] to generate $\mathbf{R}_1 \in \mathbb{R}^{H\times W\times C}$. A channel

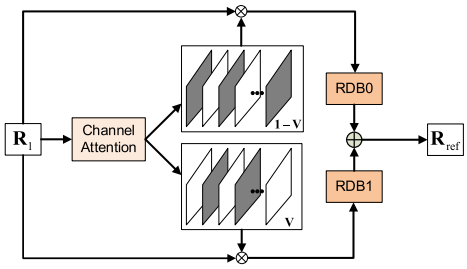


Fig. 3. Channel-attention mechanism to generate refined features.

attention (CA) layer is used for generating a channel attention weight:

$$\mathbf{V} = CA\left(\mathbf{R}_1\right), \tag{10}$$

where CA is the CA module [61] and $\mathbf{V} \in \mathbb{R}^{1\times C}$ contains C coefficients. Using the CA module, the weights for each channel are obtained. The original weighted channels (V) are fed into an RDB for extracting coarse features. Another RDB is used to extract additional information from the channels with weights $(\mathbf{1}-\mathbf{V})$. Then, the two sources of information are fused together to generate $\mathbf{R}_{\text{ref}}$, as shown in Fig. 3. The refined features $\mathbf{R}_{\text{ref}}$ are obtained by

$$\mathbf{R}_{\text{ref}} = RDB_1\left(\mathbf{R}_1 \otimes \mathbf{V}\right) + RDB_0\left[\mathbf{R}_1 \otimes (\mathbf{1}-\mathbf{V})\right], \tag{11}$$

where $\otimes$ is an element-wise product operation, $\mathbf{1}$ is a vector with the same dimensions as $\mathbf{V}$ and with values of all ones, and RDB_0 and RDB_1 are two RDBs. Using the channel attention mechanism, both coarse and fine information are extracted by the two RDBs, and the features in the phase components are fully used to retrieve a high-quality wrapped phase.

After obtaining the refined features $\mathbf{R}_{\text{ref}}$, two RBs followed by a pixel shuffle layer [62] that consists of a 3×3 convolution and a pixel shuffle operation are used to generate the SR residual phase components, which can be written as

$$\mathbf{SC}_{\text{res}} = PS\left\{\text{Conv}_{3\times3}\left[RBs\left(\mathbf{R}_{\text{ref}}\right)\right]\right\}, \tag{12}$$

where RBs represents the combination of two RBs, $\text{Conv}_{3\times3}$ is the convolution in the pixel shuffle layer that maps the features from shape $H\times W\times C$ to shape $H\times W\times s^2\times 2$, and PS is the pixel shuffle operation that rearranges the features from shape $H\times W\times s^2\times 2$ to shape $sH\times sW\times 2$.

Finally, the initial phase components are added to the residual component $\mathbf{SC}_{\text{res}}$ to obtain the final SR phase components:

$$\mathbf{SC}_{\text{out}} = \mathbf{SC}_{\text{ini}} \uparrow_s + \mathbf{SC}_{\text{res}}, \tag{13}$$

where $\mathbf{SC}_{\text{out}} \in \mathbb{R}^{sH\times sW\times 2}$ consists of an S_N of size $sH\times sW$ and a C_N of size $sH\times sW$.

When retrieving STR phases, i.e., $1\times$ phase retrieval, the upsampling of the initial estimation branch needs to be dropped and the pixel shuffle layer is substituted with a convolution in the refinement and super-resolution branch.

C. Phase Demodulation and Unwrapping

After obtaining the SR S_N and C_N, the wrapped phase ϕ can be obtained by using Eq. (3). Because the values of S_N and C_N may be either positive or negative, the quadrant of S_N/C_N should be found in order to obtain the correct value when using an arctangent function to compute the phase. For practical implementation, the atan2 function [63] is used to handle the cases of different quadrants. Unlike multi-pattern methods that are less error-prone, the single-pattern method only has access to limited intensity information. Therefore, the values of S_N and C_N may have incorrect signs when S_N and C_N are close to zero. In addition, because of the characteristics of the arctangent function, the phase is wrapped into $[-\pi, \pi]$, which may cause ambiguity during 3D reconstruction. To address these two issues, phase unwrapping can be used to obtain the correct and absolute phase.

Spatial phase unwrapping methods determine whether the phase of a pixel needs to be unwrapped, by comparing its value with that of its adjacent points. This process is easily influenced by issues such as phase jump and noise. Temporal phase unwrapping methods are much more robust and accurate although they require an additional low-frequency phase as reference [64]. Alternatively, deep learning techniques can be used to train a network for phase unwrapping by using the same single pattern [26]. Then, absolute phase can be obtained by

$$\Phi = \frac{\phi}{f} + \text{round}\left(\frac{f\phi^r - \phi}{2\pi}\right)\frac{2\pi}{f}, \tag{14}$$

where ϕ^r and Φ are a reference phase and an absolute phase respectively, and round$(\cdot)$ is a rounding operation. After phase unwrapping, the sign errors on S_N and C_N are eliminated, and the final absolute phase is continuous and normalized to $[0, 2\pi]$.

In this paper, the focus is on retrieving the wrapped phase and a reference phase is used to achieve phase unwrapping directly. Embedding a phase unwrapping module into the network will be studied in future.

D. Loss Function

Three losses are used to train the proposed network. The overall loss function is

$$\mathcal{L} = \mathcal{L}_{\text{SC}} + \lambda\left(\mathcal{L}_{\text{FS}_\text{N}} + \mathcal{L}_{\text{FC}_\text{N}}\right), \tag{15}$$

where $\mathcal{L}_{\text{SC}}$, $\mathcal{L}_{\text{FS}_\text{N}}$, $\mathcal{L}_{\text{FC}_\text{N}}$, and λ are phase component loss, sine shifting loss, cosine shifting loss, and weight of the last two losses, respectively.

The phase component loss measures the difference between the SR phase components and the ground truth, and is computed by

$$\mathcal{L}_{\text{SC}} = \left\|S_N^G - S_N\right\|_2^2 + \left\|C_N^G - C_N\right\|_2^2, \tag{16}$$

where S_N^G and C_N^G are the ground truths, and $\|\cdot\|_2^2$ represents the L_2 distance.

The PSM in SRPRNet implements the process of imitating the phase-shifting SL. Each shift block learns a relationship between the current phase shift and the next phase shift. In the proposed network, a four-step phase-shifting manner is adopted, i.e., with the number of shift steps N being 4.

According to Eq. (4) and Eq. (5), S_N and C_N are computed by

$$S_N = \sum_{n=0}^{3} I_n^c \sin\left(\frac{2\pi n}{4}\right) = I_2^c - I_4^c, \tag{17}$$

and

$$C_N = \sum_{n=0}^{3} I_n^c \cos\left(\frac{2\pi n}{4}\right) = I_1^c - I_3^c. \tag{18}$$

The quality of the phase-shifting features generated in PSM significantly influences the final demodulated phase. According to Eq. (17) and Eq. (18), regularization terms $\mathcal{L}_{\mathrm{FS_N}}$ and $\mathcal{L}_{\mathrm{FC_N}}$ are added to make each shift block learn the phase shift information effectively. Each shift block generates a single-channel feature after the group of features is obtained, and the single-channel feature of each shift block is obtained by

$$\mathbf{F}_g^i = \mathrm{Conv}_{3\times 3}\left(\mathbf{F}^i\right) \quad (i = 1, 2, 3, 4), \tag{19}$$

where $\mathbf{F}_g^i$ is in $\mathbb{R}^{H\times W\times 1}$. The sine shifting loss, which is used to evaluate the quality of the second and fourth phase-shifting features, is computed by:

$$\mathcal{L}_{\mathrm{FS_N}} = \left\| S_N^G \downarrow_s - \left(\mathbf{F}_g^2 - \mathbf{F}_g^4\right)\right\|_2^2. \tag{20}$$

The cosine shifting loss, which is used to evaluate the quality of the first and third phase-shifting features, is computed by:

$$\mathcal{L}_{\mathrm{FC_N}} = \left\| C_N^G \downarrow_s - \left(\mathbf{F}_g^1 - \mathbf{F}_g^3\right)\right\|_2^2. \tag{21}$$

The phase component loss, which contributes to the total loss predominantly, is used to control the network to obtain better parameters. The sine shifting loss and the cosine shifting loss are used to force the PSM to perform as expected on learning the relationships between features with different phase shifts.

V. Experiments

In this section, the datasets and implementation details are discussed, and ablation studies are carried out to verify the effectiveness of the proposed network architecture. Then, the proposed network is compared with several single-pattern phase retrieval methods. Finally, the phase-shifting results and 3D reconstruction results are shown to further verify the effectiveness of the proposed network.

A. Datasets

The dataset created by Qian et al. [42], named as FP1000, that contains 800 training pairs and 200 test pairs (single pattern I^c and phase component pairs S_N and C_N), the dataset created by Nguyen et al. [22], named as FP672, that contains 600 training pairs and 72 test pairs, and the dataset created for this work, named as FP147, that contains 120 training pairs and 27 test pairs were used to train and test the network. The ground truths on S_N and C_N are generated by a multi-pattern phase-shifting SL method. The frequencies of the used coded patterns for FP1000, FP672, and FP147 are 48, 100, and 32 respectively. The resolution of the images in the datasets is 640×480. The patterns in dataset FP1000 are in horizontal mode and the patterns in datasets FP672 and FP147 are in vertical mode. Our dataset FP147 with a different frequency further enriches the existing single-pattern structured light datasets. As described in Section IV-C, we focus on retrieving the phase components and omit phase unwrapping in this work. To avoid any issues that may be introduced by phase unwrapping, we directly apply the ground truth reference phase generated by the multi-pattern SL method for the phase unwrapping stage in all experiments. For the SR phase retrieval tasks, bicubic downsampling was used to generate the LR patterns.

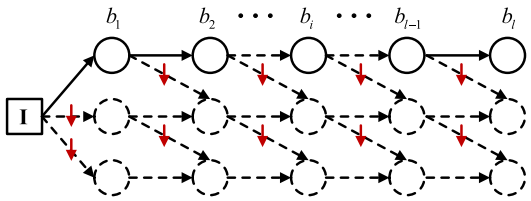


Fig. 4. Combinations of shift blocks.

B. Implementation Details

The network was implemented in PyTorch on a PC with an NVIDIA Tesla P100 GPU. The models were optimized via the Adam optimizer. The batch size was set to 1, and the initial learning rate was 2×10^{-4}. The learning rate decreased by half after every 100 epochs. For the FP1000 and FP672 datasets, the training stopped after 200 epochs. For the FP147 dataset, the training stopped after 400 epochs. The dataset FP147 was captured by an SL system which consists of a Casio XJ-M140 (projector resolution: 1024×768), an AVT Prosilica GC650 camera (camera resolution: 640×480), and a controlling circuit.

C. Ablation Study

1) Shift Blocks: The shift blocks in PSM were used to obtain representative features with phase shifts. To verify the effectiveness of the designed combination of the shift blocks, the network was retrained with different numbers of shift blocks and different numbers of scales in each shift block. In this section, RSRM is excluded, and the PSM is followed by a pixel shuffle layer to achieve the SR function. The combination of shift blocks is shown in Fig. 4, where b_i represents the number of branches in the ith shift block and a circle represents the convolution operation with an RDB. The number of shift blocks is l. In order to avoid too low resolutions, the largest number of b_i was set to four, i.e., the smallest downsampling scale is 1/8. The largest number of scale and the final combination is denoted as $(b_1, b_2, \cdots, b_l)$. The results are presented using: (1) the mean squared error (MSE) on S_N and C_N and (2) mean absolute error (MAE) and root mean squared error (RMSE) on absolute phase for $4\times$ SR phase retrieval. A more accurate phase can be obtained when the MSE on S_N and C_N is smaller.

TABLE I

COMPARATIVE RESULTS OBTAINED ON FP147 AND FP1000 USING DIFFERENT COMBINATIONS OF SHIFT BLOCKS WITHOUT RSRM FOR 4× SR PHASE RETRIEVAL. THE MSE IS EVALUATED ON S_N/C_N AND THE MAE/RMSE (10^{-4} RAD) IS EVALUATED ON ABSOLUTE PHASE

Combination	MSE (S_N/C_N)		MAE/RMSE		#P	FLOPs
	FP147	FP1000	FP147	FP1000		
(1,1,1)	14.14/11.83	20.79/20.31	12.86/32.03	5.47/16.60	0.81M	15.82G
(1,2,3)	13.24/10.88	20.41/20.19	12.55/31.46	5.50/16.65	1.61M	21.95G
(2,2,2)	13.62/11.27	20.21/19.98	12.66/32.07	5.41/16.52	1.61M	22.60G
(2,3,4)	13.73/11.48	20.36/20.03	12.62/31.82	5.50/16.97	2.40M	27.38G
(3,3,3)	13.67/11.64	20.45/19.97	12.49/31.71	5.51/16.76	2.40M	27.56G
(4,4,4)	13.33/11.60	20.38/19.98	12.68/32.19	5.53/16.89	3.19M	32.00G
(1,2,3,4)	12.46/**10.36**	19.83/19.58	12.00/**30.30**	5.36/**16.34**	2.69M	32.37G
(1,2,3,4,4)	**12.38**/10.45	**19.38/19.19**	**11.95**/30.49	**5.28**/16.38	3.81M	42.84G

TABLE II

COMPARATIVE RESULTS OBTAINED ON FP147 WITH AND WITHOUT RSRM FOR 4× SR PHASE RETRIEVAL. THE MSE AND MAE/RMSE REPRESENT THE SAME EVALUATIONS AS THOSE IN TABLE I

Model	MSE (S_N/C_N)	MAE/RMSE	#P	FLOPs
Without RSRM	12.46/10.36	12.00/30.30	2.69M	32.37G
RDB_0	11.92/10.04	11.70/30.75	3.48M	47.50G
RDB_1	11.91/9.97	11.62/30.27	3.48M	47.50G
RDB_0 + RDB_1	**11.53/9.67**	**11.52/30.16**	3.62M	50.22G

The MSE on S_N/C_N and MAE/RMSE on absolute phase for the 4× SR phase retrieval task with different combinations of shift blocks are shown in Table I. Besides, the number of parameters ('#P' in Table I) and the floating point operations (FLOPs) are listed. Among these combinations, the (1, 2, 3, 4) combination has the best performance on RMSE, and it is a noticeable improvement over other combinations. Although the (1, 2, 3, 4, 4) combination has similar MSE and MAE/RMSE as the (1, 2, 3, 4) combination, such a combination requires more parameters and FLOPs, which decreases efficiency. To balance efficiency and performance, we select the (1, 2, 3, 4) combination as the basic design in PSM. The optimal condition obtained on FP1000 is the same as or very similar to that obtained on FP147, and we only show the results obtained on FP147 in the following ablation studies.

2) Refinement and Super-Resolution Module (RSRM): The RSRM is used to obtain detailed information in areas around edges and the SR residual for compensating the initially estimated phase components. To demonstrate the effectiveness of RSRM, the proposed network was compared with and without this module. Besides, to demonstrate that the setup with two branches followed by the CA layer in RSRM is useful, the results of using either the RDB_0 or the RDB_1 branch alone were compared, as well as using two branches, i.e., both RDB_0 and RDB_1 in RSRM.

In Table II, the results are improved after including the RDB_1 branch in RSRM. Besides, the result on MSE and MAE/RMSE with two branches followed by the CA layer is better than that of using only one branch. Therefore, the design of RSRM is effective on obtaining detailed information for the initial estimation and on producing high-quality output.

3) Loss Function: To verify the effectiveness of the loss function with regularization and to find an appropriate weight, the network was trained without regularization and with different weights of the regularization term from 0.1 to 0.4.

TABLE III

COMPARATIVE RESULTS OBTAINED ON FP147 WITH DIFFERENT WEIGHTS ON REGULARIZATION FOR 4× SR PHASE RETRIEVAL. THE MSE AND MAE/RMSE REPRESENT THE SAME EVALUATIONS AS THOSE IN TABLE I

Weight on regularization	MSE (S_N/C_N)	MAE/RMSE
$\lambda = 0$	11.53/9.67	11.52/30.16
$\lambda = 0.1$	11.60/9.79	11.54/30.05
$\lambda = 0.2$	11.57/**9.47**	**11.48/29.76**
$\lambda = 0.3$	**11.56**/9.63	11.53/29.92
$\lambda = 0.4$	11.58/9.70	11.60/30.17

$\lambda = 0$ means that the network was trained by using the loss function without the regularization terms. As Table III shows, when the weight of the regularization terms is 0.2, the network has the best performance and obtains an improvement of 0.16 (MSE on S_N + MSE on C_N) compared with the network trained without the regularization terms.

D. Comparison With State-of-the-Art

To verify the effectiveness of SRPRNet for STR phase retrieval, SRPRNet was compared with several other phase retrieval methods for single-pattern SL 3D imaging, including Fourier transform profilometry [19] (FT) and the networks proposed by Feng et al. [21] (FPDL) and Qian et al. [42] (SPU). However, there is no other SR phase retrieval method for single-pattern SL 3D imaging in the literature. Bicubic interpolation was used to upsample the input pattern to the target resolution required for FT, FPDL, and SPU in order to verify the effectiveness and superiority of the SR function of SRPRNet. Besides, the combination of the pixel shuffle [62] layer with FPDL and SPU was also implemented for more comparisons. Furthermore, we compress our full SRPRNet by reducing the number of channels, substituting some 3×3 convolutions with 1×1 convolutions, and dropping two RBs. Specifically, the changes include the followings: (1) Channel number C is changed from 64 to 60. (2) In PSM, the convolutions ahead of the RDBs are changed from 3×3 convolutions to 1×1 convolutions. (3) In RSRM, the first convolution in the upper branch and the first convolution in the lower branch are changed from 3×3 convolutions to 1×1 convolutions, and 'Residual Block1_b' and 'Residual Block2_b' are dropped. Such a compressed model is named as SRPRNet-light.

1) Quantitative Results: In Table IV, quantitative results on the three datasets are shown. The note 1× in the 'Scale' column represents STR phase retrieval when the resolutions of the input and output are the same. 'Bicubic+' represents the combination of bicubic interpolation with another model, while '+Shuffle' represents the combination of another model with the pixel shuffle SR. MAE and RMSE on absolute phase are used to measure accuracy. When MAE and RMSE are smaller, absolute phase is more accurate and will lead to better 3D reconstruction. Because phase unwrapping can reduce phase errors [7] and absolute phase is normalized to $[0, 2\pi]$, the magnitude of absolute phase errors is small (10^{-4} rad level in the results). The frequencies of patterns in these three datasets are different, therefore the magnitudes of absolute phase errors are also different.

TABLE IV

COMPARATIVE RESULTS ON THREE DATASETS FOR STR OR SR PHASE RETRIEVAL TASKS. THE RESULTS SHOW THE MAE AND RMSE ON ABSOLUTE PHASE. THE BEST RESULTS ARE IN BOLD AND THE SECOND-BEST RESULTS ARE UNDERLINED

Method	Scale	#Params.	FLOPs	MAE/RMSE (unit:10^{-4}rad)			
				FP1000	FP672	FP147	Average
FT	1×	-	-	19.60/56.81	10.23/32.54	26.63/63.27	18.82/50.87
FPDL	1×	0.68M	146.53G	4.59/11.80	3.85/12.02	12.26/29.84	6.90/17.89
SPU	1×	1.34M	167.75G	4.21/10.99	3.37/10.41	11.72/27.87	6.43/16.42
SRPRNet-light (Ours)	1×	0.85M	174.69G	3.67/9.51	3.11/9.07	10.51/24.79	5.76/14.46
SRPRNet (Ours)	1×	3.60M	798.17G	**3.11/8.15**	**2.87/8.08**	**9.48/24.25**	**5.15/13.49**
Bicubic+FT	2×	-	-	19.42/56.12	10.49/33.31	26.42/63.29	18.78/50.91
Bicubic+FPDL	2×	0.68M	146.53G	5.92/16.37	4.35/14.22	12.97/32.09	7.75/20.89
FPDL+Shuffle	2×	0.69M	37.05G	4.76/13.17	4.03/13.14	14.59/37.89	7.79/21.40
Bicubic+SPU	2×	1.34M	167.75G	5.47/16.64	4.19/14.33	12.42/29.27	7.36/20.08
SPU+Shuffle	2×	1.35M	43.00G	4.80/13.56	3.70/11.92	12.98/31.75	7.16/19.08
SRPRNet-light (Ours)	2×	0.86M	43.92G	3.99/11.40	3.42/10.68	10.51/25.98	5.97/16.02
SRPRNet (Ours)	2×	3.61M	199.81G	**3.45/10.17**	**3.21/9.81**	**9.29/24.45**	**5.32/14.81**
Bicubic+FT	4×	-	-	19.36/55.63	71.10/104.95	24.79/64.46	38.42/75.01
Bicubic+FPDL	4×	0.68M	146.53G	8.05/24.09	19.94/42.33	16.26/38.78	14.75/35.07
FPDL+Shuffle	4×	0.69M	9.68G	5.70/16.55	5.45/18.16	17.79/47.02	9.65/27.24
Bicubic+SPU	4×	1.34M	167.75G	8.80/28.33	20.78/44.25	18.21/42.68	15.93/38.42
SPU+Shuffle	4×	1.41M	11.81G	6.34/18.76	5.62/18.61	15.82/40.98	9.26/26.12
SRPRNet-light (Ours)	4×	0.87M	11.23G	5.51/16.52	5.06/16.79	12.43/31.87	7.67/21.73
SRPRNet (Ours)	4×	3.62M	50.22G	**4.97/15.99**	**4.70/15.68**	**11.48/29.76**	**7.05/20.48**

As can be seen in Table IV, SRPRNet and SRPRNet-light achieve the best and second-best performances on the FP1000, FP672, and FP147 datasets for all 1×, 2×, and 4× SR phase retrievals. Specifically, the average MAE/RMSE of SRPRNet are 1.28/2.93×10^{-4} rad, 1.83/4.27×10^{-4} rad, and 2.21/5.64×10^{-4} rad smaller than SPU+Shuffle for 1×, 2×, and 4× SR phase retrievals. Although the results achieved by SRPRNet-light decreases slightly compared with SRPRNet, SRPRNet-light still has 0.67/1.96×10^{-4} rad, 1.19/3.06×10^{-4} rad, and 1.59/4.39×10^{-4} rad improvements on average MAE/RMSE compared with the SPU+Shuffle method. Apart from the FT method, the others are deep learning-based phase retrieval methods for single-pattern SL 3D imaging and have much better performance than the FT method. The number of parameters for SRPRNet-light is between that of FPDL+Shuffle and SPU+Shuffle, and the FLOPs for SRPRNet-light are similar to that of SPU+Shuffle. The full version SRPRNet provides further improvements on performance but with a larger number of parameters and FLOPs compared with SRPRNet-light.

The results for 2× SR FT (average MAE/RMSE: 19.42/56.12×10^{-4}) and 4× SR FT (average MAE/RMSE: 19.36/55.63×10^{-4}) on FP1000 are even better than that for 1× FT (average MAE/RMSE: 19.60/56.81×10^{-4}). This is because the accuracy of the FT method is related to removing high-frequency components and retaining fundamental component in the frequency domain, and it is difficult to extract the fundamental component completely. Although the upsampled input pattern is of lower quality, the interpolation is equivalent to a smoothing process that reduces the high-frequency components. After smoothing, it is easier to retain the fundamental component.

2) Qualitative Results: In Fig. 5, the ground truth absolute phases are shown. The first to third rows are from the FP1000, FP672, and FP147 datasets respectively. Qualitative comparisons for the 1×, 2×, and 4× SR phase retrieval are shown in Figs. 6, 7, and 8 respectively, which show the STR or SR absolute phase error maps. The values below each figure are the MAE/RMSE for those scenes. As can be seen, most of the phase errors are around areas with large depth discontinuities, such as the areas along object boundaries. The results obtained from SRPRNet show the smallest phase error numerically and visually, and SRPRNet performs better on areas with depth discontinuities.

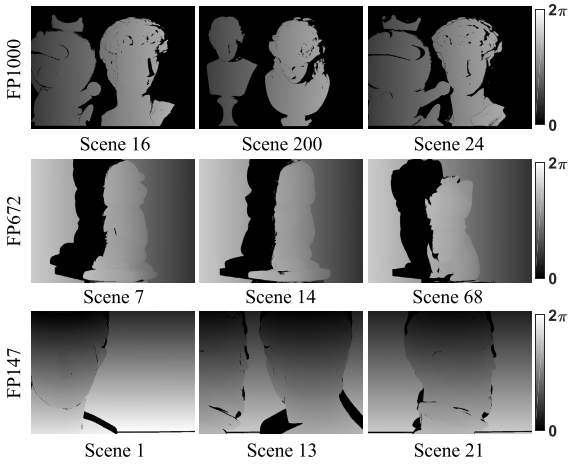


Fig. 5. Ground truth absolute phases from different scenes. The first to third columns correspond to the phase error maps in Figs. 6, 7, and 8, respectively.

It can be observed from Fig. 8 that the phase error maps using bicubic+FT, bicubic+FPDL, and bicubic+SPU have large errors even in areas without depth discontinuities. The reason for this phenomenon is that the frequency of the projected patterns is too high. As noted in Section V-A, the frequency of the patterns in FP672 is 100. Such a high frequency will result in aliasing when upsampling the pattern. For those areas where aliasing occurs, it is difficult to obtain the correct phase value using an incorrect intensity of the input pattern. For SRPRNet, the features for each image pixel

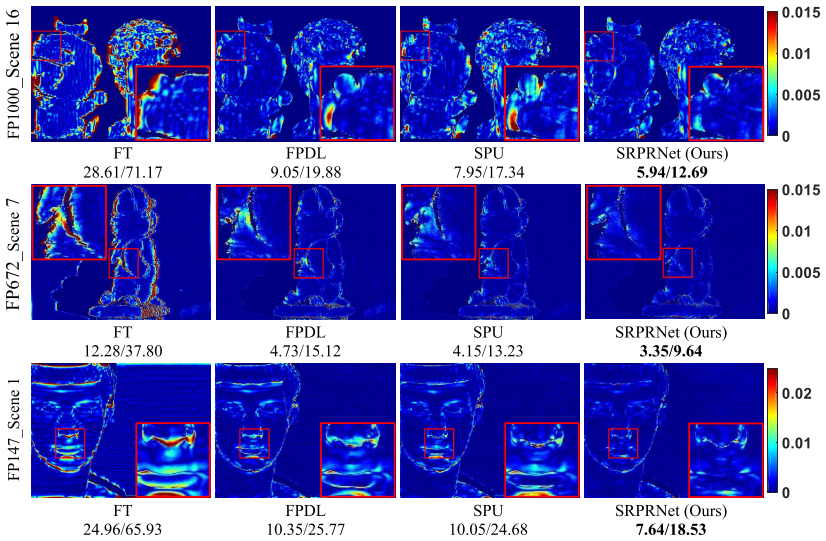


Fig. 6. Visual comparison of phase error maps for 1× phase retrieval. The numbers below each subfigure are the MAE/RMSE (unit: 10^{-4} rad) of that scene.

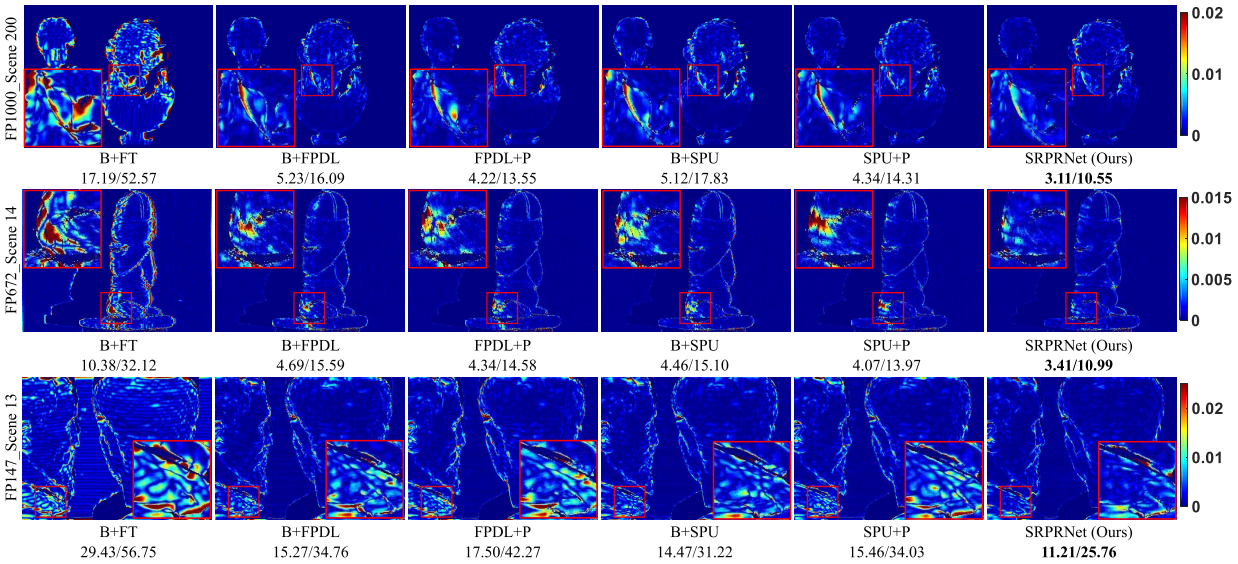


Fig. 7. Visual comparison of phase error maps for 2× SR phase retrieval. 'B+' represents the combination of bicubic interpolation with another model and '+P' represents the combination of another model with the pixel shuffle SR method. The numbers below each subfigure are defined to be the same as those in Fig. 6.

are extracted through multiple stages, and these features are combined with features of the surrounding pixels. Therefore, the phase information can be retrieved more accurately from the input intensity at that pixel.

E. Shifts and Refinement Results

The main motivation for SRPRNet comes from simulating the four-step phase-shifting SL by using a neural network with a single pattern as input. The single-channel features generated by the shift blocks represent the simulated patterns with different shift information, as illustrated in Fig. 9. The intensities of each feature map are normalized to $[0, 1]$. As can be seen, these single-channel features carry the phase shift information, and the four feature maps conform to the four-step phase-shifting patterns although the range of intensity for each feature map is different. The different ranges can be easily adjusted when fusing the features together.

Besides, two sample lines are extracted from each feature map to clearly illustrate the phase shift information. The bottom row in Fig. 9 shows the intensities of the two lines. The intensity at the same positions in each feature map fits the shape of the sinusoidal waveform. The phase shift information of each feature is reproduced in flat or discontinuous areas, which demonstrates that the approach is valid. As the features go deeper in the network, more convolution layers become involved, which makes the features smoother.

In Fig. 10, the feature maps of residuals generated within RSRM are illustrated. The residuals on S_N and C_N focus on the areas around edges or the positions where the gradient

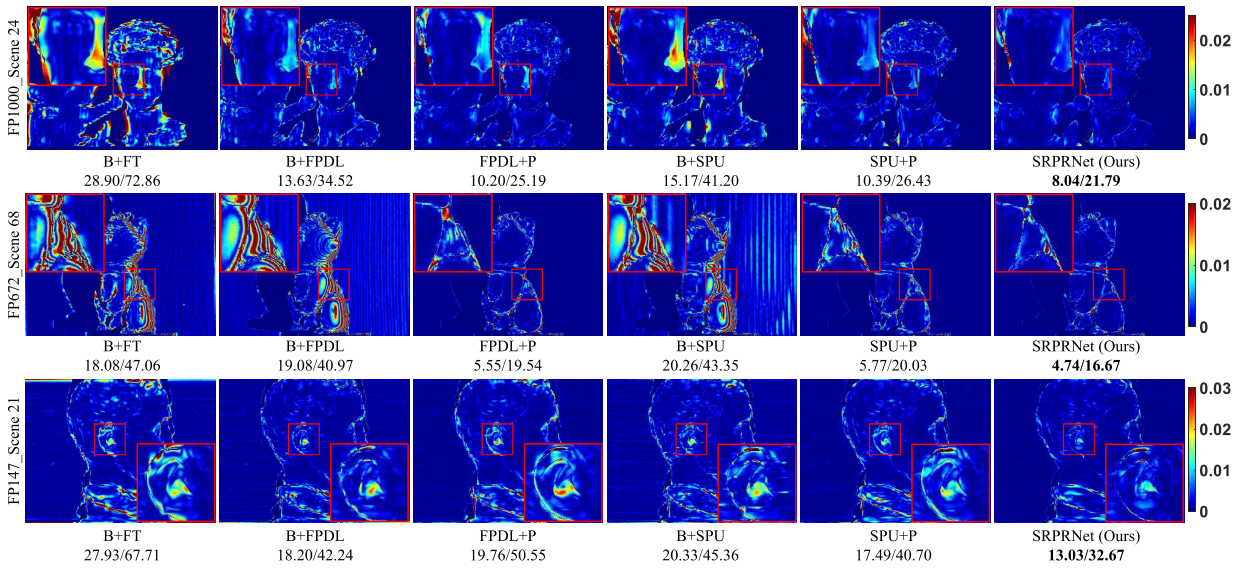


Fig. 8. Visual comparison of phase error maps for 4× SR phase retrieval. The method names are the same as those in Fig. 7. The numbers below each subfigure are defined to be the same as those in Fig. 6.

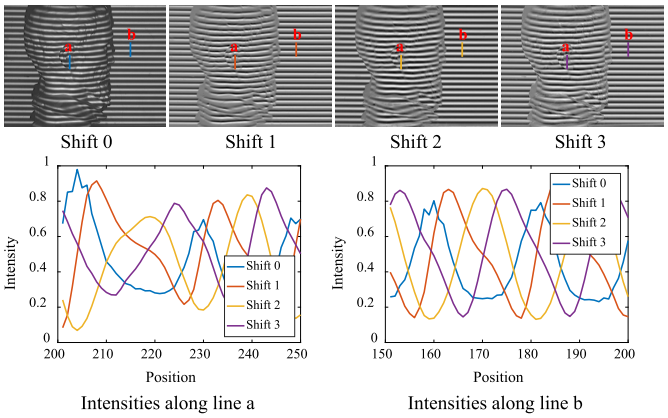


Fig. 9. Visual features with different phase shifts generated by PSM. Top: single-channel feature maps of each shift block; Bottom: two sample lines in each feature map.

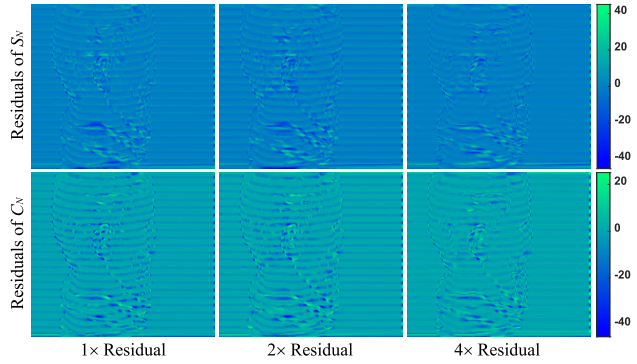


Fig. 10. Visualization of residuals generated within RSRM for tasks with different resolutions. The top and bottom rows show the residuals on S_N and C_N.

direction of the sine waveform changes. For flat areas, the features generated by PSM are effective enough for obtaining the final phase components. By using RSRM, more detailed information is obtained.

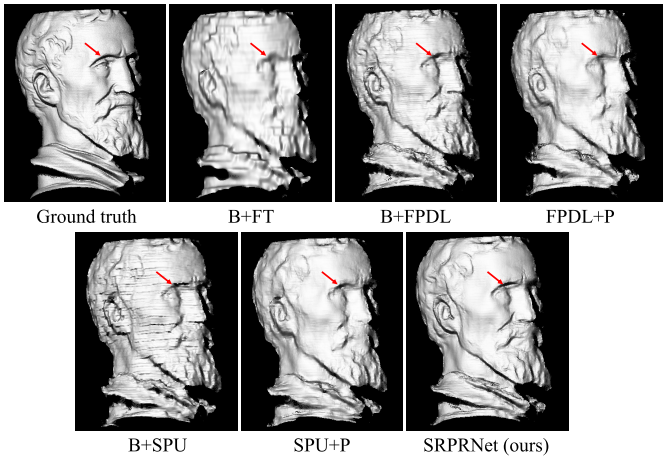


Fig. 11. 3D reconstruction results using 4× SR phases.

F. 3D Reconstruction

To illustrate the results more intuitively, a group of 3D reconstruction results using 4× SR phases is shown in Fig. 11. The 3D results using the phase generated by SRPRNet are significantly better than other methods, especially in areas with detailed textures, such as the local area indicated by the red arrow. Even by just using quarter-resolution input pattern, the quality of the final 3D result using SRPRNet maintains high. These 3D results demonstrate that the quality of 3D reconstruction is directly related to the quality of phase information, and SRPRNet is effective and superior. Besides, it can be seen that the 3D results obtained by deep learning-based methods are not influenced by gamma distortion [65] (with gamma distortion, the 3D results will be wavy), while multi-pattern 3D imaging with three patterns will be severely affected by gamma distortion. Multi-pattern 3D imaging with four or more patterns is much less affected by gamma distortion or noise, so the accuracy is typically higher than that of single-pattern

methods. Neural network methods have the ability to alleviate the influence of gamma distortion even with just a single pattern. We will fully investigate such phenomena in future.

VI. Conclusion and Future Work

In this paper, a super-resolution phase retrieval network (SRPRNet) for single-pattern SL 3D imaging has been proposed. Motivated by multi-pattern phase-shifting SL 3D imaging, a phase-shifting module (PSM) that exploits groups of features with different phase shifts is designed. The PSM consists of four shift blocks and one fusion block, and hierarchically extracts features at different scales. Then, a refinement and super-resolution module (RSRM) is designed for generating high-quality STR or SR phase components. Finally, the STR or SR absolute phase can be obtained after phase demodulation and unwrapping. Comparative results on $1\times$, $2\times$, and $4\times$ super-resolution phase retrieval show that the proposed SRPRNet achieves state-of-the-art performance. In future work, a network that can retrieve phases with different resolutions by using only one group of parameters will be designed.

References

[1] J. Geng, "Structured-light 3D surface imaging: A tutorial," *Adv. Opt. Photon.*, vol. 3, no. 3, pp. 128–160, Mar. 2011.

[2] L. Yang, B. Wang, R. Zhang, H. Zhou, and R. Wang, "Analysis on location accuracy for the binocular stereo vision system," *IEEE Photon. J.*, vol. 10, no. 1, pp. 1–16, Feb. 2018.

[3] C. Zuo, S. Feng, L. Huang, T. Tao, W. Yin, and Q. Chen, "Phase shifting algorithms for fringe projection profilometry: A review," *Opt. Lasers Eng.*, vol. 109, pp. 23–59, Oct. 2018.

[4] I. Ishii, K. Yamamoto, K. Doi, and T. Tsuji, "High-speed 3D image acquisition using coded structured light projection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 925–930.

[5] R. J. Valkenburg and A. M. Mcivor, "Accurate 3D measurement using a structured light system," *Image Vis. Comput.*, vol. 16, no. 2, pp. 99–110, Feb. 1998.

[6] J. L. Posdamer and M. D. Altschuler, "Surface measurement by space-encoded projected beam systems," *Comput. Graph. Image Process.*, vol. 18, no. 1, pp. 1–17,1982.

[7] J. Li, L. Hassebrook, and C. Guan, "Optimized two-frequency phase-measuring-profilometry light-sensor temporal-noise sensitivity," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 20, no. 1, pp. 106–115, 2003.

[8] K. Liu, Y. Wang, D. L. Lau, Q. Hao, and L. G. Hassebrook, "Dual-frequency pattern scheme for high-speed 3D shape measurement," *Opt. Exp.*, vol. 18, no. 5, pp. 5229–5244, 2010.

[9] Y. Wang, K. Liu, Q. Hao, D. L. Lau, and L. G. Hassebrook, "Period coded phase shifting strategy for real–time 3D structured light illumination," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3001–3013, Nov. 2011.

[10] M. Gupta and S. K. Nayar, "Micro phase shifting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 813–820.

[11] Y. Zhang, D. L. Lau, and Y. Yu, "Causes and corrections for bimodal multi-path scanning with structured light," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4431–4439.

[12] Y. Wang, K. Liu, Q. Hao, X. Wang, D. L. Lau, and L. G. Hassebrook, "Robust active stereo vision using Kullback–Leibler divergence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 548–563, Mar. 2012.

[13] L. Ekstrand and S. Zhang, "Three-dimensional profilometry with nearly focused binary phase-shifting algorithms," *Opt. Lett.*, vol. 36, no. 23, pp. 4518–4520, 2011.

[14] Y. Wang, S. Basu, and B. Li, "Binarized dual phase-shifting method for high-quality 3D shape measurement," *Appl. Opt.*, vol. 57, no. 23, pp. 6632–6639, 2018.

[15] P. Jia, J. Kofman, and C. English, "Intensity-ratio error compensation for triangular-pattern phase-shifting profilometry," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 24, no. 10, pp. 3150–3158, 2007.

[16] L. Lu, J. Xi, Y. Yu, and Q. Guo, "New approach to improve the performance of fringe pattern profilometry using multiple triangular patterns for the measurement of objects in motion," *Opt. Eng.*, vol. 53, no. 11, May 2014, Art. no. 112211.

[17] D. Moreno, K. Son, and G. Taubin, "Embedded phase shifting: Robust phase shifting with embedded signals," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2301–2309.

[18] Y. Zhang, Z. Xiong, Z. Yang, and F. Wu, "Real-time scalable depth sensing with hybrid structured light illumination," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 97–109, Jan. 2014.

[19] X. Su and Q. Zhang, "Dynamic 3D shape measurement method: A review," *Opt. Lasers Eng.*, vol. 48, no. 2, pp. 191–204, Feb. 2010.

[20] M. Maruyama and S. Abe, "Range sensing by projecting multiple slits with random cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 6, pp. 647–651, Jun. 1993.

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

[22] 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.

[23] C. Jiang, S. Xing, and H. Guo, "Fringe harmonics elimination in multi-frequency phase-shifting fringe projection profilometry," *Opt. Exp.*, vol. 28, no. 3, pp. 2838–2856, Jan. 2020.

[24] S. Feng, C. Zuo, L. Zhang, W. Yin, and Q. Chen, "Generalized framework for non-sinusoidal fringe analysis using deep learning," *Photon. Res.*, vol. 9, no. 6, pp. 1084–1098, 2021.

[25] H. Yu, D. Zheng, J. Fu, Y. Zhang, C. Zuo, and J. Han, "Deep learning-based fringe modulation-enhancing method for accurate fringe projection profilometry," *Opt. Exp.*, vol. 28, no. 15, pp. 21692–21703, 2020.

[26] G. E. Spoorthi, R. K. S. S. Gorthi, and S. Gorthi, "PhaseNet 2.0: Phase unwrapping of noisy data based on deep learning approach," *IEEE Trans. Image Process.*, vol. 29, pp. 4862–4872, 2020.

[27] Q. Kemao, "Two-dimensional windowed Fourier transform for fringe pattern analysis: Principles, applications and implementations," *Opt. Lasers Eng.*, vol. 45, no. 2, pp. 304–317, 2007.

[28] J. Zhong and J. Weng, "Phase retrieval of optical fringe patterns from the ridge of a wavelet transform," *Opt. Lett.*, vol. 30, no. 19, pp. 2560–2562, 2005.

[29] B. Li, C. Tang, X. Zhu, Y. Su, and W. Xu, "Shearlet transform for phase extraction in fringe projection profilometry with edges discontinuity," *Opt. Lasers Eng.*, vol. 78, pp. 91–98, Mar. 2016.

[30] X. Zhu, C. Tang, B. Li, C. Sun, and L. Wang, "Phase retrieval from single frame projection fringe pattern with variational image decomposition," *Opt. Lasers Eng.*, vol. 59, pp. 25–33, Aug. 2014.

[31] Z. Dong and Z. Chen, "Advanced Fourier transform analysis method for phase retrieval from a single-shot spatial carrier fringe pattern," *Opt. Lasers Eng.*, vol. 107, pp. 149–160, Aug. 2018.

[32] K. L. Boyer and A. C. Kak, "Color-encoded structured light for rapid active ranging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 1, pp. 14–28, Jan. 1987.

[33] H. Lin, L. Nie, and Z. Song, "A single-shot structured light means by encoding both color and geometrical features," *Pattern Recognit.*, vol. 54, pp. 178–189, Jun. 2016.

[34] Budianto, D. P. K. Lun, and Y.-H. Chan, "Robust single-shot fringe projection profilometry based on morphological component analysis," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5393–5405, Nov. 2018.

[35] H. Zhang et al., "Color-encoded single-shot computer-generated Moiré profilometry," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, May 2021.

[36] C. Li, Y. Cao, C. Chen, Y. Wan, G. Fu, and Y. Wang, "Computer-generated Moiré profilometry," *Opt. Exp.*, vol. 25, no. 22, pp. 26815–26824, 2017.

[37] A. O. Ulusoy, F. Calakli, and G. Taubin, "One-shot scanning using de Bruijn spaced grids," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops (ICCV) Workshops*, Sep. 2009, pp. 1786–1792.

[38] T. Petković, T. Pribanić, and M. Đonlić, "Single-shot dense 3D reconstruction using self-equalizing De Bruijn sequence," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5131–5144, Nov. 2016.

[39] Z. Wang and Y. Yang, "Single-shot three-dimensional reconstruction based on structured light line pattern," *Opt. Lasers Eng.*, vol. 106, pp. 10–16, Jul. 2018.

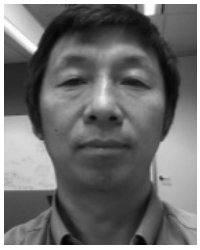
[40] P. Yao, S. Gai, and F. Da, "Coding-Net: A multi-purpose neural network for fringe projection profilometry," *Opt. Commun.*, vol. 489, Jun. 2021, Art. no. 126887.

[41] G. Qiao, Y. Huang, Y. Song, H. Yue, and Y. Liu, "A single-shot phase retrieval method for phase measuring deflectometry based on deep learning," *Opt. Commun.*, vol. 476, Dec. 2020, Art. no. 126303.
[42] 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.
[43] S. Van der Jeught and J. J. Dirckx, "Deep neural networks for single shot structured light profilometry," *Opt. Exp.*, vol. 27, no. 12, pp. 17091–17101, 2019.
[44] 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.
[45] T. Yang, Z. Zhang, H. Li, X. Li, and X. Zhou, "Single-shot phase extraction for fringe projection profilometry using deep convolutional generative adversarial network," *Meas. Sci. Technol.*, vol. 32, no. 1, Jan. 2020, Art. no. 015007.
[46] W. Hu, H. Miao, K. Yan, and Y. Fu, "A fringe phase extraction method based on neural network," *Sensors*, vol. 21, no. 5, p. 1664, Feb. 2021.
[47] R. C. Machineni, G. E. Spoorthi, K. S. Vengala, S. Gorthi, and R. K. S. S. Gorthi, "End-to-end deep learning-based fringe projection framework for 3D profiling of objects," *Comput. Vis. Image Understand.*, vol. 199, Oct. 2020, Art. no. 103023.
[48] S. Yuan, Y. Hu, Q. Hao, and S. Zhang, "High-accuracy phase demodulation method compatible to closed fringes in a single-frame interferogram based on deep learning," *Opt. Exp.*, vol. 29, no. 2, pp. 2538–2554, 2021.
[49] 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.
[50] Y. J. Kil, B. Mederos, and N. Amenta, "Laser scanner super-resolution," in *Proc. 3rd Eurographics/IEEE VGTC Conf. Point-Based Graph.*, Jul. 2006, pp. 9–16.
[51] K. Ouji, M. Ardabilian, L. Chen, and F. Ghorbel, "3D deformable super-resolution for multi-camera 3D face scanning," *J. Math. Imag. Vis.*, vol. 47, nos. 1–2, pp. 124–137, Sep. 2013.
[52] M. Weinmann, C. Schwartz, R. Ruiters, and R. Klein, "A multi-camera, multi-projector super-resolution framework for structured light," in *Proc. Int. Conf. 3D Imag., Modeling, Process., Visualizat. Transmiss.*, May 2011, pp. 397–404.
[53] Y. Shiba, S. Ono, R. Furukawa, S. Hiura, and H. Kawasaki, "Temporal shape super-resolution by intra-frame motion encoding using high-fps structured light," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 115–123.
[54] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 169–176.
[55] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014.
[56] Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Joint image filtering with deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1909–1923, Aug. 2019.
[57] L. He et al., "Towards fast and accurate real-world depth super-resolution: Benchmark dataset and baseline," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9229–9238.
[58] Z. Zhao, J. Zhang, S. Xu, Z. Lin, and H. Pfister, "Discrete cosine transform network for guided depth map super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5697–5707.
[59] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
[60] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[61] 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)*, Sep. 2018, pp. 286–301.
[62] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
[63] E. Organick, "Some processors also offer the library function called ATAN2 a function of two arguments (opposite and adjacent)," in *A FORTRAN IV Primer*. Reading, MA, USA: Addison-Wesley, 1966, p. 42.
[64] J. Song, D. Lau, Y. Ho, and K. Liu, "Automatic look-up table based real-time phase unwrapping for phase measuring profilometry and optimal reference frequency selection," *Opt. Exp.*, vol. 27, no. 9, pp. 13357–13371, Apr. 2019.
[65] K. Liu, Y. Wang, D. L. Lau, Q. Hao, and L. G. Hassebrook, "Gamma model and its analysis for phase measuring profilometry," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 27, no. 3, pp. 553–562, 2010.

Jianwen Song received the B.E. degree in electronic information engineering from Sichuan University, Chengdu, China, in 2017, and the M.E. degree in detection technology and automation equipment from Sichuan University. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, University of New South Wales, Sydney. His research interests focus on 3D vision, particularly on stereo matching and structured light 3D imaging.

Kai Liu (Senior Member, IEEE) received the B.S. and M.S. degrees in computer science from Sichuan University, Chengdu, China, and the Ph.D. degree in electrical engineering from the University of Kentucky, Lexington, KY, USA. He is currently a Professor with the College of Electrical Engineering, Sichuan University. His main research interests include computer/machine vision, active/passive stereo vision, and image processing.

Arcot Sowmya received the Ph.D. degree in computer science from IIT Bombay, besides other degrees in mathematics and computer science. She is currently a Professor with the School of Computer Science and Engineering, University of New South Wales, Sydney. Her research has been applied to extraction of linear features in remotely sensed images and feature extraction, recognition, and computer aided diagnosis in medical images. Her areas of research include learning in vision for segmentation, classification, and object recognition.



Changming Sun received the Ph.D. degree in computer vision from the Imperial College London, London, U.K., in 1992. Then, he joined CSIRO, Sydney, Australia, where he is currently a Principal Research Scientist, carrying out research and working on applied projects. He is also a Conjoint Professor with the School of Computer Science and Engineering, University of New South Wales, Sydney. His current research interests include computer vision, image analysis, and pattern recognition. He has served on the program/organizing committees for various international conferences. He is an Associate Editor of the *EURASIP Journal on Image and Video Processing*.