Optics and Lasers in Engineering 167 (2023) 107623

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Large depth-of-field microscopic structured-light 3D imaging with focus stacking

Liming Chen, Song Zhang*

School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Keywords:
Large depth-of-field
3d imaging
Structured light
Fringe projection
Microscopy

ABSTRACT

State-of-the-art microscopic structured-light (SL) three-dimensional (3D) imaging systems typically use conventional lenses with a fixed focal length achieving a limited depth of field (DOF). This paper proposes to drastically increase the DOF of the microscopic 3D imaging by leveraging the focus stacking technique and developing a novel computational framework. We first capture fringe images with various camera focal lengths using an electrically tunable lens (ETL) and align the recovered phase maps using phase constraints. Then, we extract the focused pixel phase using fringe contrast and stitch them into an all-in-focus phase map via energy minimization. Finally, we reconstruct the 3D shape using the all-in-focus phase map. Experimental results demonstrate that our proposed method can achieve a large DOF of approximately 2 mm and field of view (FOV) of approximately 4 mm × 3 mm with a pixel size at the object space of approximately 2.6 μm. The achieved DOF is approximately 10 × the DOF of the system without the proposed method.

1. Introduction

Microscopic structured-light (SL) three-dimensional (3D) imaging is an important inspection technique in many applications such as semiconductors and micro-mechatronics, owing to its advantages of high accuracy, high speed, high flexibility, and low cost. Over the past few decades, extensive research has been conducted, focusing on different aspects of this field including system design [1–3], calibration [4–7], measurement methods [8–10], and imaging speed [11,12]. These works greatly improved the performance of microscopic SL 3D imaging, and was well summarized by the review written by Hu, et al. [13]. Despite recent rapid advancements, there is still a major limitation in the state-of-the-art microscopic SL systems: they typically have a fixed focal length providing a limited depth of field (DOF). This limitation becomes more apparent with larger system magnifications. Though many systems used telecentric lenses [4–6,8], which can typically provide a larger DOF than pin-hole lenses under the same magnification, the original DOF is usually not sufficient for many applications.

The DOF limitation of microscopic SL 3D imaging has also attracted attention in the past few years. A representative idea is to extend the DOF via special devices or system configurations. For example, researchers introduced the Scheimpflug condition into microscopic SL systems [7,14,15], to ensure the maximum common focus area of the projector and the camera. Wang et al. [14]. proposed a multiple-camera SL system with telecentric lenses. Each camera was also set up following the Scheimpflug condition and focused on a different depth range. The DOF was extended by combining the 3D results from all cameras. Meng et al. [16]. recently proposed to employ light-field imaging in microscopic SL systems to extend the DOF. Though these methods worked effectively, they inevitably increased the system complexity and reduced the system flexibility.

Computational imaging has been demonstrated its promises to increase DOF. For example, our previous work [17] introduced the focal sweep technique in a microscopic SL system extend DOF. This method can effectively extend the DOF without drastically altering the system configuration and imaging model. However, since the deconvolution process involved with the focal sweep technique is sensitive to noise and cannot recover high spatial frequencies robustly, it would not be the best choice in applications with high-resolution and high-accuracy requirements [18].

Focus stacking is another computational imaging method that can increase DOF [19,20]. The basic idea is that the camera captures a set of images at different focus settings (e.g., focal length, image distance, or object distance), which is called *focal stack*; and the software algorithms analyze these images to generate an all-in-focus image. The algorithms mainly include two steps: The first step is image alignment (or image registration). This step transforms images with different focal settings since the magnification may vary with different focus settings. The second step is called image stitching. This step creates the all-in-focus image by extracting the pixels with the best focus from the aligned images

* Corresponding author.
E-mail address: szhang15@purdue.edu (S. Zhang).

https://doi.org/10.1016/j.optlaseng.2023.107623
Received 19 December 2022; Received in revised form 10 April 2023; Accepted 21 April 2023
Available online 25 April 2023

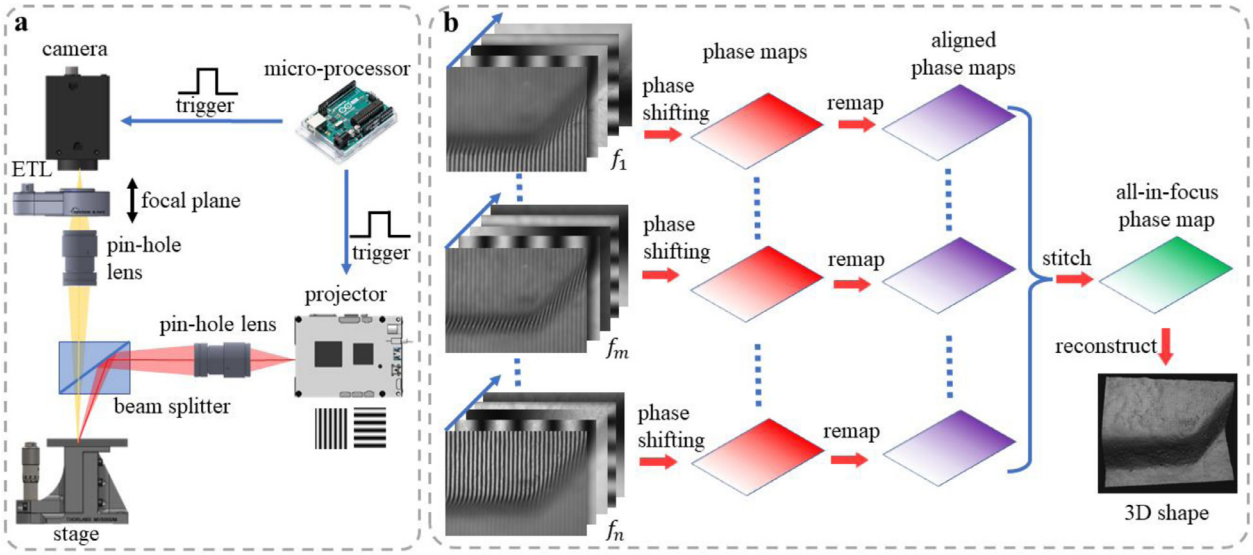

Fig. 1. Working principle of the proposed method for large DOF microscopic 3D imaging. (a) Image acquisition process: change the camera focal length by moving the focal plane while maintaining the projector's focus. (b) Image analysis process: compute high-frequency phase maps along the vertical direction and lower-frequency phase maps along both horizontal and vertical directions, remap all high-frequency phase maps using the affine warps obtained by the lower-frequency phase correspondence, and finally stitch all high-frequency phase maps into a single all-in-focus phase map for 3D reconstruction.

and combining these pixels into a full image. Both image alignment and image stitching have been extensively studied in 2D imaging [21,22], providing rich choices for the focus stacking techniques depending on different applications. Owing to the low cost and good performance, the focus stacking technique has raised interest in both microscopic [23] and macroscopic 2D imaging [24,25].

Though focus stacking has been widely used in 2D imaging, it is non-trivial to be applied to microscopic SL 3D imaging. One challenge is that microscopic 3D imaging is fundamentally different from 2D photography. For example, it has higher measurement accuracy and resolution requirements, so many algorithms developed for 2D imaging cannot be directly applied. Another challenge is that the imaging model of microscopic SL 3D imaging is highly related to the focal length, and thus it is very sensitive to the focus change. Liu et al. [26]. built a microscopic SL system using telecentric lenses to extend the DOF of microscopic 3D imaging using the focus stacking technique. In their study, they varied the object distance by moving the object vertically using a translation stage to form the focal stack. Since the telecentric lenses preserve magnification at different object distances, this method avoided the image alignment problem. However, the requirements of using telecentric lenses and high-precision mechanical motion increase system complexity and cost.

In this work, we propose a method to drastically enlarge the DOF of microscopic 3D imaging by leveraging the focus stacking technique and developing a novel computational framework. Fig. 1 summarizes the overall concept of our proposed method. Different from the previous work by Liu et al. [26], we capture the focal stack by changing the camera focal length using an electrically tunable lens (ETL) that allows for capturing multi-focus images without mechanical motion. Then we improve the basic flow of the focus stacking technique to process phase maps, in lieu of 2D images, for both alignment and stitching. Specifically, we capture fringe images with different frequencies and compute phase maps from these fringe images. We then use one of the lower-frequency phase maps to estimate affine warps that minimize the phase difference between different focus settings. Once the affine warps are determined, we apply them to the high-frequency fringe images to align the phase maps and fringe contrast maps. Then we use the aligned high-frequency fringe contrast maps to measure the focus levels for each pixel and create an all-in-focus high-frequency phase map using the aligned high-frequency phase maps based on the measured focus levels. The all-in-focus high-frequency phase map is finally used to reconstruct 3D shape of the scene. Since our proposed method considers the magnification differences among various focus settings, it does not require constant magnification for the hardware design. Therefore, it can be applied to generic microscopic SL systems with pin-hole lenses.

2. Principle

2.1. Phase-shifting algorithm

Phase-shifting algorithms are widely employed in SL systems because of their high accuracy, high resolution, and high speed. The intensity of the k-th fringe pattern can be mathematically described as,

$$I_k(x,y) = I'(x,y) + I''(x,y)\cos\left[\phi(x,y) + 2k\pi/N\right], k \in [1,N] \tag{1}$$

where $I'(x,y)$ is the DC component representing the background intensity, $I''(x,y)$ is the fringe modulation amplitude, $\phi(x,y)$ is the phase to be solved, If $N \geq 3$, the unknown phase can be computed by,

$$\phi(x,y) = -\tan^{-1}\frac{\sum_{k=1}^{N} I_k(x,y)\sin(2k\pi/N)}{\sum_{k=1}^{N} I_k(x,y)\cos(2k\pi/N)} \tag{2}$$

The phase map produced by the above equation is wrapped within a range from $-\pi$ to π with 2π discontinuities because of the nature of arctangent function. Hence, a temporal, spatial, or alternative phase unwrapping algorithm that has been extensively studied over the past few decades [27,28] is needed to recover a continuous phase map,

$$\Phi(x,y) = \phi(x,y) + 2\pi \times \kappa(x,y) \tag{3}$$

where $\kappa(x,y)$ is an integer number obtained from the phase unwrapping algorithm. In this work, a three-frequency phase unwrapping algorithm is adopted. Meanwhile, the DC component $I'(x,y)$ and the fringe modulation $I''(x,y)$ can also be computed by,

$$I'(x,y) = \frac{1}{N}\sum_{k=1}^{N} I_k(x,y) \tag{4}$$

$$I''(x,y) = \frac{2}{N}\sqrt{\left[\sum_{k=1}^{N} I_k(x,y)\cos\delta_k\right]^2 + \left[\sum_{k=1}^{N} I_k(x,y)\sin\delta_k\right]^2} \tag{5}$$

where

$$\delta_k = \frac{2k\pi}{N} \tag{6}$$

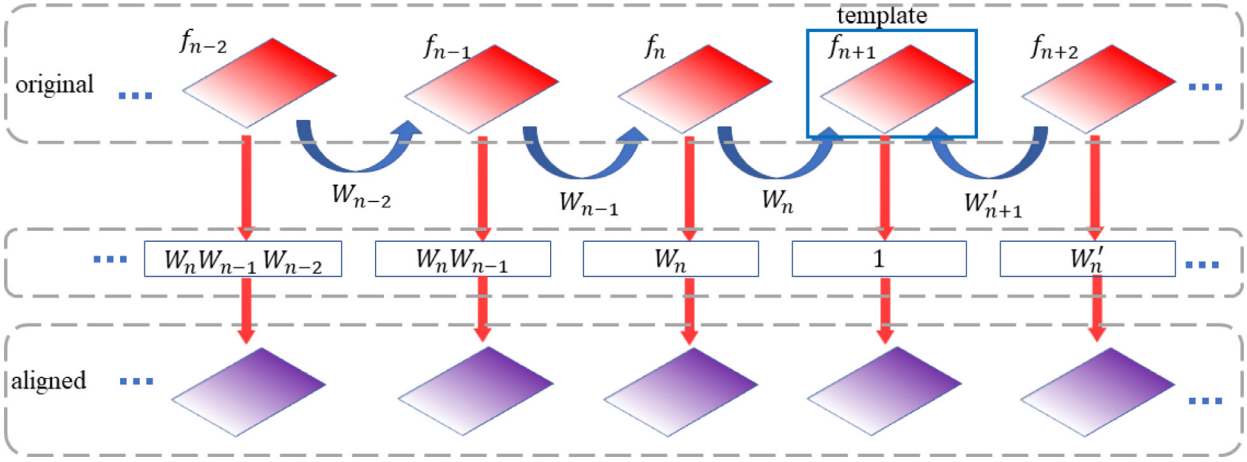

Fig. 2. Recursive phase alignment method to alleviate the impact of the phase errors caused by camera defocus. Compute the global warps consecutively in the focal stack, i.e., compute $\boldsymbol{W_n}$ which aligns the phase map of the focus setting $\boldsymbol{f_n}$ to $\boldsymbol{f_{n+1}}$ (or W'_{n+1} for $\boldsymbol{f_{n+2}}$ to $\boldsymbol{f_{n+1}}$, etc.), and then adopt the computed global warps recursively to transform each focus setting to the template. For example, we use global warp W_nW_{n-1} to align f_{n-1} with f_{n+1}.

Then the fringe contrast can be defined as the normalized fringe modulation with the DC component,

$$\gamma(x,y) = \frac{I''(x,y)}{I'(x,y)} \tag{7}$$

Since the fringe contrast varies with the blurring effect, it can be used as a focus measure [26]. We will describe the way of using this focus measure in Section 2.3.

2.2. Phase map alignment

Our phase map alignment algorithm utilizes the phase maps along the horizontal and the vertical directions obtained by the phase-shifting algorithm introduced in Section 2.1. With the recovered phase maps, it is a natural idea that we align two phase maps by remapping each pixel. However, this method assumes that the phase difference is only produced by the magnification change. But this assumption is not always valid unless we address two critical alignment problems associated with the camera defocus that introduces phase error or completely fails to recover phase.

Different levels of camera defocus introduce varying errors to the recovered phase. Due to camera defocus, each camera pixel receives light coming from several neighboring projector pixels. If the object surface reflection varies locally or/and the surface geometric shape changes quickly, the defocus blur brings error to the phase [29–31]. To alleviate the impact of the phase errors caused by camera defocus, we propose to compute the affine warps consecutively in the focal stack, i.e., align focus setting f_n to f_{n+1} (or f_{n-1}) and adopt the computed affine warps W_n (or W'_n) recursively to transform each phase map to the template. The mathematical definition of the affine warp is given in Section 2.2.2. This process is visualized in Fig. 2. Since the defocus levels of the consecutive focus settings do not change much, this method can greatly reduce the impact of the phase error caused by camera defocus.

The second problem is that if the camera is defocused so much that all fringes are washed out, it is impossible to recover phase from fringe patterns. From the Fourier optics, the defocus effect is considered as a low-pass filter (typically a Gaussian filter). The signal of the projected fringes may be filtered out if the filter size is too large, and this problem is more problematic for high-frequency fringes. Fig. 3(a) shows one of the fringe images of an angled flat plane. This image clearly shows that the fringes at the top and bottom regions nearly disappear due to the defocus effect. Fig. 3(b) shows the phase map computed from such high-frequency fringe images. The phase lines are obviously curved, which should not happen since the object is flat. To solve this problem, we compute the global warps using the phase map with a lower frequency.

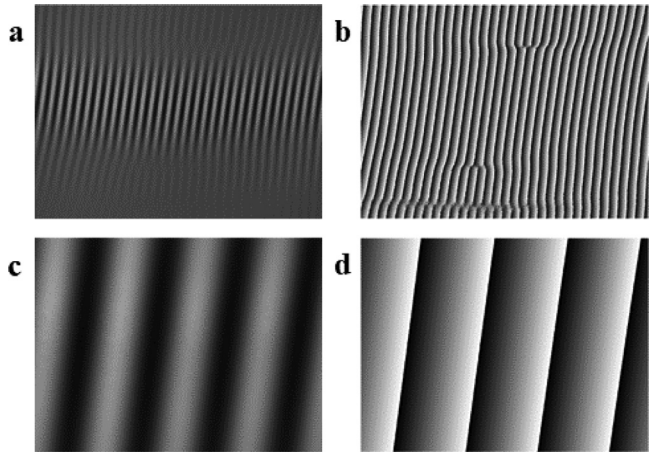

Fig. 3. Camera defocus effect on fringe images. (a) One of the high-frequency fringe images under camera defocus. (b) Recovered wrapped high-frequency phase map. (c) One of the lower-frequency fringe images under camera defocus. (d) Recovered wrapped lower-frequency phase map.

From the fringe image and the corresponding wrapped phase map shown in Fig. 3(c) and (d), we can clearly see that the lower-frequency phase is recovered well even at the region the camera is significantly defocused.

After addressing the problems associated with camera defocus, the phase maps between consecutive focus settings can be aligned by a pixel remapping process. Given two focus settings f_n and f_{n+1}, without loss of generality, we assume f_{n+1} as the template here, the alignment can be mathematically described as a pixel remapping process that seeks to find,

$$X_{n+1}, Y_{n+1} = arg \min_{X_{n+1},Y_{n+1}} \left\{ \left| \varphi_h^{f_n}(x,y) - \varphi_h^{f_{n+1}}\left(X_{n+1}, Y_{n+1}\right) \right| + \left| \varphi_v^{f_n}(x,y) - \varphi_v^{f_{n+1}}\left(X_{n+1}, Y_{n+1}\right) \right| \right\}, \forall (x,y) \tag{8}$$

where (x,y) are original coordinates, $\varphi_h^{f_n}(x,y)$and $\varphi_v^{f_n}(x,y)$ are horizontal and vertical unwrapped lower-frequency phase maps under f_n, and $\varphi_h^{f_{n+1}}(x,y)$ and $\varphi_v^{f_{n+1}}(x,y)$ are horizontal and vertical unwrapped lower-frequency phase maps under f_{n+1}, and (X_{n+1}, Y_{n+1}) are the remapping target for (x,y) that aligns f_n to f_{n+1}.

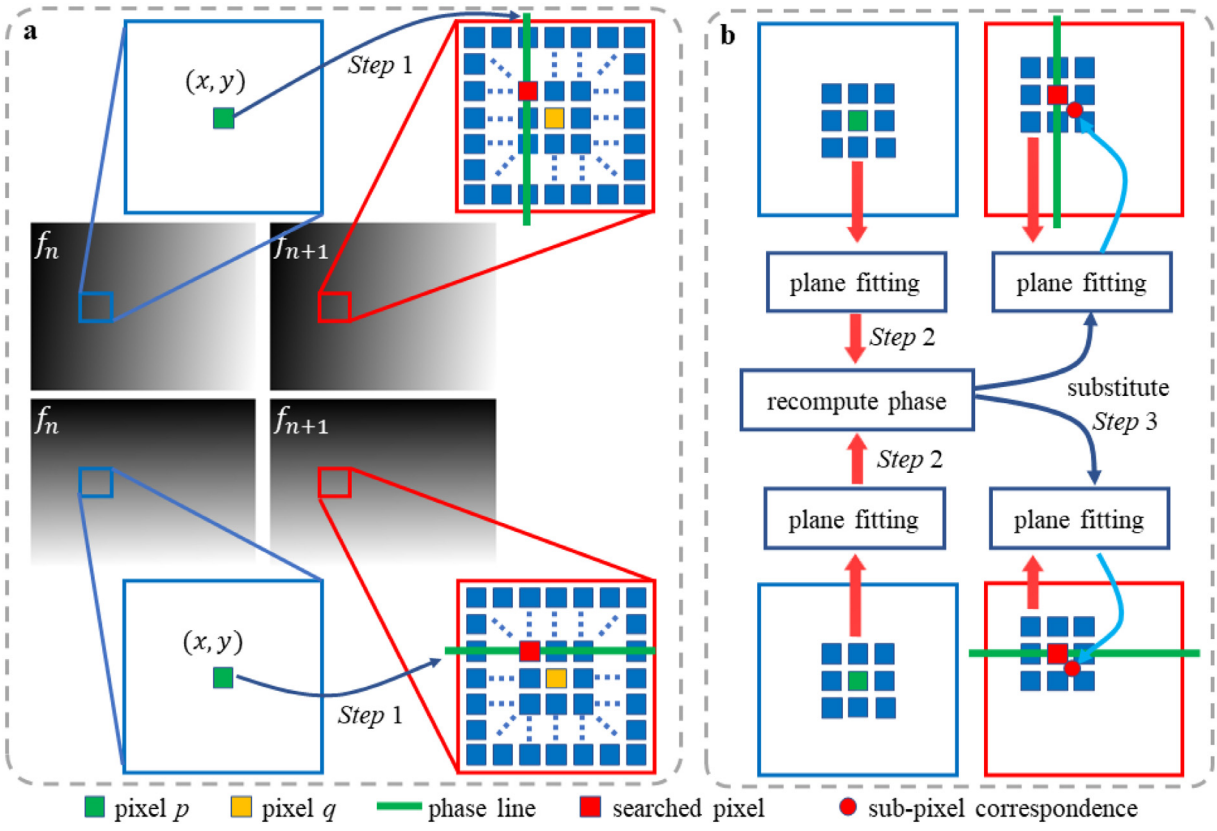

Fig. 4. Proposed sub-pixel correspondence method. (a) Pixel-wise search: for a pixel p in the lower-frequency unwrapped phase map of f_n, we search a window in lower-frequency unwrapped phase map of f_{n+1} centered at pixel q which has the same coordinates with p. With the horizontal and vertical phase maps, we can uniquely determine a pixel that matches p best. (b) Sub-pixel relocation: fit planes for both p and the searched pixel from (a) using both horizontal and vertical phase maps, then recompute the phases of p. Finally, substitute the recomputed phase into the fitted planes around the searched pixel and we can get the sub-pixel location with the minimum phase difference with p.

2.2.1. Sub-pixel correspondence

We propose to find the remapping pixel coordinates (X_{n+1}, Y_{n+1}) for pixel $p(x, y)$ by phase comparison and plane-fitting. Specifically,

(1) search a local window at f_{n+1} centered at pixel q, which has the same coordinates (x, y), to find a pixel with a minimum phase difference, which can be visualized as Fig. 4(a). The pixel with the minimum phase difference can be found by a pixel-by-pixel comparison with all pixels in the local window. The phase difference can be computed by,

$$\Delta\varphi = \left|\varphi_h^{f_n}(x,y) - \varphi_h^{f_{n+1}}(x',y')\right| + \left|\varphi_v^{f_n}(x,y) - \varphi_v^{f_{n+1}}(x',y')\right|,$$
$$|x'-x| \le \frac{r_m}{2}, |y'-y| \le \frac{r_m}{2} \tag{9}$$

where (x', y') represents the coordinates of the pixel inside the local window at f_{n+1}, *and* r_m represents the size of the local window, which can be set as a constant.

(2) for the pixel $p(x, y)$, fit two plane functions within a small pixel window around $p(x, y)$, one uses the horizontal phase map and the other uses vertical phase map,

$$\varphi_h^{f_n}(u,v) = a_0^{f_n} + b_0^{f_n}u + c_0^{f_n}v \tag{10}$$

$$\varphi_v^{f_n}(u,v) = a_1^{f_n} + b_1^{f_n}u + c_1^{f_n}v \tag{11}$$

where (u, v) are coordinates of the neighboring pixels used for plane fittings; $a_0^{f_n}$, $b_0^{f_n}$, $c_0^{f_n}$ and $a_1^{f_n}$, $b_1^{f_n}$, $c_1^{f_n}$ are the coefficients solved by a least-squares method. The window size w needs to be adjusted according to the noise level. We then recompute the phase of pixel $p(x, y)$ using the fitted plane equation, denoting as $\tilde{\varphi}_h^{f_n}(x, y)$ and $\tilde{\varphi}_v^{f_n}(x, y)$.

(3) fit planes using the same method as *step* (2) for the matched pixel (x', y') at f_{n+1} to obtain coefficients $a_0^{f_{n+1}}$ to $c_1^{f_{n+1}}$. Within the fitted planes, we relocate the sub-pixel correspondence using the recomputed phase values in *step* (2),

$$x'' = \frac{c_1^{f_{n+1}}\left[\tilde{\varphi}_h^{f_n}(x,y) - a_0^{f_{n+1}}\right] - c_0^{f_{n+1}}\left[\tilde{\varphi}_v^{f_n}(x,y) - a_1^{f_{n+1}}\right]}{c_1^{f_{n+1}}b_0^{f_{n+1}} - c_0^{f_{n+1}}b_1^{f_{n+1}}} \tag{12}$$

$$y'' = \frac{b_1^{f_{n+1}}\left[\tilde{\varphi}_h^{f_n}(x,y) - a_0^{f_{n+1}}\right] - b_0^{f_{n+1}}\left[\tilde{\varphi}_v^{f_n}(x,y) - a_1^{f_{n+1}}\right]}{b_1^{f_{n+1}}c_0^{f_{n+1}} - b_0^{f_{n+1}}c_1^{f_{n+1}}} \tag{13}$$

The *step* (2) and *step* (3) can be visualized as Fig. 4(b).

2.2.2. Affine warp estimation

The whole phase map can be aligned by running the same sub-pixel correspondence for every pixel. However, the computation complexity will be very high since 4 plane fittings are needed for each pixel. To simplify the algorithm, we model the pixel remapping function as an affine warp [32]. Here we ignored potential non-linear deformations caused by inconsistent lens distortions since our experimental results demonstrated the inconsistent lens distortion is very small. Specifically, we down-sample the phase map at f_n and only compute the sub-pixel correspondences for the sampled pixels of f_n. We also masked out invalid pixels using the lower-frequency fringe contrast. Based on these

sub-pixel correspondences, we estimate an affine warp that can be mathematically described as [32],

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{14}$$

where vector $[u', v', 1]^T$ represents the homogeneous pixel coordinates after the transformation; the p_1 through p_6 are parameters of the affine transformation; vector $[u, v, 1]^T$ denotes the original homogeneous pixel coordinates; and superscript T is matrix transpose operator. These parameters can be solved by the least-squares method. However, the sub-pixel correspondence could still contain incorrect mappings, especially around sharp discontinuities. To address this, we employ the RANSAC algorithm [33] in this least-squares to get the best estimation. With the affine warp, the remapping can be computed by,

$$\left[X_{n+1},\ Y_{n+1},\ 1\right]^T = W_n[x, y, 1]^T \tag{15}$$

Where

$$W_n = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ 0 & 0 & 1 \end{bmatrix} \tag{16}$$

As shown in Fig. 2, we mathematically describe the recursive alignment procedure from f_{n-m} to f_{n+1} as

$$\left[X_{n+1},\ Y_{n+1},\ 1\right]^T = W_n W_{n-1} \cdots W_{n-m}[x, y, 1]^T \tag{17}$$

2.2.3. Aligned phase map generation

After getting the affine warp in Eq. (17) for each focus setting, we can apply the obtained affine warps to the high-frequency phase map under the corresponding focus setting to generate a new phase map which will be aligned with the template. For each pixel, we use the bilinear interpolation algorithm to compute the phase value of the sub-pixel location indicated by its affine warp in the template high-frequency phase map and assign the phase value to the pixel. After running the same procedure for all focus settings, we can get aligned high-frequency phase maps under all focus settings that will finally be used to create the all-in-focus high-frequency phase map.

2.3. Phase map stitching

The first step in phase map stitching is to measure the focus level for each pixel under different focus settings. Since the projector focus does not change, we can use the pixel-wise fringe contrast $\gamma(x, y)$ computed by Eq. (7) as the focus measure. Since this method does not use image gradient, it also works well for texture-less scenes, which are always challenging to existing texture-based sharpness measure operators [34]. To further improve the sensitivity of this focus measure, we use the high-frequency fringe contrast as the focus measure. The fringe contrast maps also need to be aligned. For each focus setting, we use the same global warps with the phase map alignment discussed in Section 2.2.2 and the same procedure discussed in Section 2.2.3 to generate an aligned high-frequency fringe contrast map.

With the focus measure, we can directly select pixels with the largest focus measure value. However, such a method could introduce issues since the pixel-wise fringe contrast computation inevitably contains noises and errors, especially in dark regions. Instead, we adopt the method proposed by Suwajanakorn et al. [35]. to formulate the stitching problem as maximum a posteriori (MAP) estimation in a multi-label Markov Random Fields (MRF) on a regular 4-connected grid using the fringe contrast map instead of the image intensity. This MAP estimation can be further derived to an energy minimization problem [36]. Given $\mathcal{V}$ as the set of pixels, $\mathcal{N}$ as the neighborhood pixels defined by the 4-connected grid, and S as the focal stack size, we seek to minimize the energy function:

$$E(l) = \sum_{i \in \mathcal{V}} E_i(l_i) + \lambda \sum_{(i,j) \in \mathcal{N}} E_{i,j}(l_i, l_j) \tag{18}$$

where $l_i, l_j \in [0, S-1]$ are integer numbers representing the labels for pixel i(i.e., (x_i, y_i)) and j (i.e., (x_j, y_j)). The labels here mean the focus setting indices from which each pixel should be extracted. $l(x, y)$ is the output denoting the index map that minimizes the energy function. Here we simplify $l(x_i, y_i)$ as l_i to represent the label for pixel i whose coordinates are (x_i, y_i). λ is a weighting constant balancing the contribution of the two parts. The larger λ yields a smoother final index map with more filtering effect.

Eq. (18) contains two parts. The first part is a term that can be considered as the maximum likelihood estimation that only considers the focus measure. We mathematically model this part on the fringe contrast under each focus setting $\gamma_n(i)$ where $n \in [0, S-1]$ as,

$$E_i(l_i) = \exp\left(-\gamma_{l_i}(i)\right) \tag{19}$$

The second part considers the prior constraint to make the labels of neighboring pixels smooth. Here, we use the total variation (TV) operator to form this part,

$$E_{i,j}(l_i, l_j) = \left|l_i - l_j\right| \tag{20}$$

The energy function in this formulation is large-dimensional and non-convex. We employed the α-expansion algorithm based on the graph cuts method [36–38] to find the solution.

With the index map $l(x, y)$, we compute a weighted-average phase for each pixel (x, y). We count the number of different indices within a small window centered at (x, y) in the index map, then compute the weighted average using the phase values of the pixel with the same coordinates (x, y) under the focus settings specified by the counted indices. The weight is set as the ratio of the counts. This alternative approach can be mathematically described as:

$$\hat{\Phi}(x, y) = \frac{\sum_{l \in \mathcal{U}(x,y)} n_l \Phi^l(x, y)}{n_{\mathcal{U}}} \tag{21}$$

where l represents an index inside a small window $\mathcal{U}(x, y)$ centered at pixel (x, y) in the index map, n_l represents the counts of index l, $\Phi^l(x, y)$ represents the aligned high-frequency phase of pixel (x, y) under the focus setting l, and $n_{\mathcal{U}}$ represents the total pixel number of the window $\mathcal{U}(x, y)$.

2.4. 3D reconstruction

Once the all-in-focus phase map is created, we can reconstruct the 3D shape if the system is calibrated. In this research, we employed the calibration method discussed in Ref. [17] to calibrate our system. When calibrating the system, the focal length of the camera needs to be set to the same value as the template in the phase alignment algorithm. Since pin-hole lenses are used for both the camera and the projector, 3D shape can be reconstructed using the standard phase-to-coordinate conversion algorithm [39].

3. Experimental results

We built a prototype system, as shown in Fig. 5, to evaluate the performance of our proposed method. In this system, a complementary metal-oxide-semiconductor (CMOS) camera (model: PointGrey GS3-U3–23S6M) was attached with a lens system consisting of a 35 mm fixed aperture ($f/1.6$) lens (model: Edmund Optics #85–362), an equivalent 20 mm extension tube, a circular polarizer (model: Edmund Optics CP42HE), and an ETL (model: Optotune EL-16–40-TC). Note that the lens was mounted reversely to increase image distance. A digital-light-processing (DLP) projector (model: Shanghai Yiyi D4500) was equipped with a lens system composed of a 35 mm lens (model: Fujinon HF35HA-1B), a circular polarizer (model: Edmund Optics CP42HE) and an ETL (model: Optotune EL-16–40-TC). Each ETL in this system was tuned by a lens driver controller (model: Optotune Lens Driver 4i) within the range of −2 to +3 dpt. A beam splitter (model: Thorlabs BP145B1) was used to

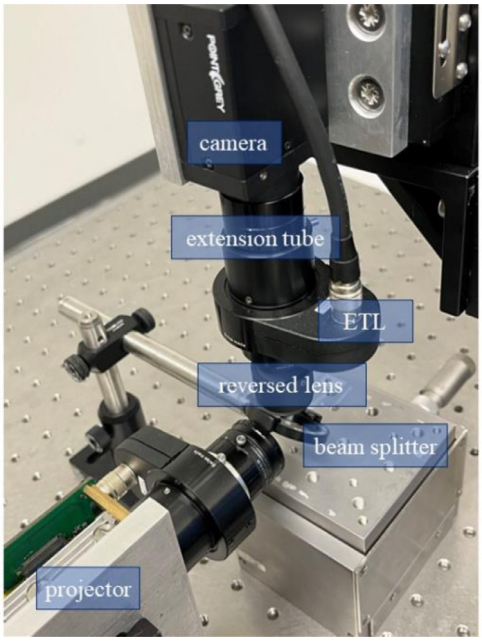

Fig. 5. A photograph of our prototype system.

adjust the projector light path [17]. The camera and the projector were synchronized by a microprocessor (model: Arduino Uno).

For all the experiments, the camera resolution was set as 1536 × 1140 pixels and the projector resolution was 912 × 1140 pixels. We used 11 different focal lengths to form the focal stack by changing the camera ETL driving current from −143.00 mA to −113.00 mA with a 3.00 mA incremental step. Hence, the focal stack will include 11 focal lengths (i.e., $S = 11$). Without loss of generality, we set the focal length f_1 produced by the −140.00 mA ETL current as the template in the alignment process. The projector ETL was held at 20.74 mA during the whole process. We set the aperture of the projector lens as $f/5.6$ and the camera exposure time as 9.9 ms. We captured three-frequency fringe images for each focal length with 9, 3, and 3 steps along the vertical direction and two-frequency fringe images with 3 and 3 steps along the horizontal direction. For the vertical direction, we set the fringe periods as 18, 144, and 912 pixels for the three frequencies. For the horizontal direction, we set the fringe periods as 216 and 1140 pixels. We used the middle-frequency phase maps computed by the fringe images with the period of 144 pixels along the vertical direction and 216 pixels along the horizontal direction to compute global warps in the phase map alignment algorithm.

For the phase map alignment, we set the middle-frequency fringe contrast threshold as 0.40 (i.e., $\tau_{mf} = 0.40$), the pixel-wise search window size as 11 × 11 pixels (i.e., $r_m = 11$), the plane fitting window as 11 × 11 pixels (i.e., $w = 11$) and the reprojection error in the RANSAC algorithm as 0.3 pixels. We applied a Gaussian filter (size: 21 × 21 pixels, standard deviation: $\sigma = 7$ pixels) to the middle-frequency phase maps at each focal length before the alignment to reduce the random noise. For the phase map stitching, we set the weighting constant in the energy minimization as 0.25 (i.e., $\lambda = 0.25$), and the size of the window to compute the weighted-average phase (i.e., U) as 21 × 21 pixels.

We first verified the performance of our proposed method by measuring a white plane. The white plane was tilted to the camera imaging plane, creating a depth of approximately 2 mm, which is approximately 10 × the DOF of our system. Fig. 6(a) - (e) show examples of captured fringe images with different frequencies. Using the fringe images shown in Fig. 6(b) and Fig. 6(d), we can compute middle-frequency wrapped phase maps and the wrapped phase maps can be unwrapped by the phase maps computed by fringe images shown in Fig. 6(c) and Fig. 6(e). We then used the computed middle-frequency unwrapped phase map to run our proposed phase alignment algorithm discussed in Section 2.2. Fig. 7(a) shows the difference map between the middle-frequency phase map computed from the middle-frequency vertical fringe images as shown in Fig. 6(b) captured at the focus setting f_1 (i.e., ETL current = −140.00 mA) and f_2 (i.e., ETL current = −137.00 mA) before alignment. The phase difference map clearly shows that they had different magnifications. We then applied our proposed phase map alignment method to compute an affine warp for each focus setting and then generated aligned phase maps. We computed the difference map again using the aligned phase maps under the same focus settings (i.e., f_1 and f_2) as shown in Fig. 7(b). The difference was greatly reduced, demonstrating that the phase maps were well aligned. Fig. 7(d) shows the difference distribution after phase map alignment that is close to the normal distribution, and the root-mean-square (RMS) value is 0.0047 rads. In comparison, without applying our proposed method, Fig. 7(c) shows the difference distribution that is not normal with a large RMS value of 0.0286 rads.

We then applied the same global warps to high-frequency phase maps and fringe contrast maps to align them. We used the aligned fringe contrast maps to generate the index map using the energy minimization algorithm discussed in Section 2.3. Fig. 8(a) shows the generated index map. Based on the index map, we stitched the high-frequency phase maps using our proposed weighted-average method to create an all-in-focus high-frequency phase map as shown in Fig. 8(b). With the all-in-focus high-frequency phase map, we reconstructed the 3D shape of the white plane, as shown in Fig. 8(c). This result demonstrated that the plane was reconstructed properly within the depth range of approximately 2 mm. We also evaluated the random noise by analyzing one cross section of the reconstructed 3D shape (the 800th column). Fig. 8(d) shows the cross section. We then removed the gross profile of the cross section to show the random noise. Fig. 8(e) shows the random noise of the 3D profile. The RMS value is approximately 1.2 μm, demonstrating that our proposed method can achieve a high depth resolution within the measurement range of approximately 2 mm.

We also evaluated our weighted-average method discussed in Section 2.3. The 3D shape shown in Fig. 9(a), and its zoom-in view of a region, where the focus setting will change, shown in Fig. 9(b) demonstrated that the reconstructed surface was smooth. It can be further verified by the cross section after detrending the depth along the 150th column, which is the center column of Fig. 9(b), as shown in Fig. 9(c).

Fig. 6. Examples of our captured fringe images. (a) One of the 9-step high-frequency vertical fringe images with a period of 18 pixels. (b) One of the 3-step middle-frequency vertical fringe images with a period of 144 pixels. (c) One of the 3-step low-frequency vertical fringe images with a period of 912 pixels. (d) One of the 3-steps middle-frequency horizontal fringe images with a period of 216 pixels. (e) One of the 3-step low-frequency horizontal fringe images with a period of 1140 pixels.

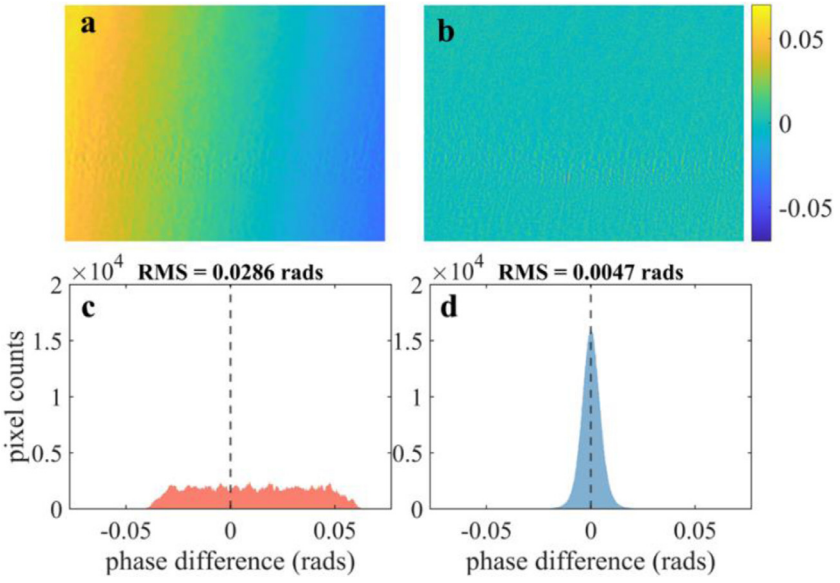

Fig. 7. Experimental results of phase map alignment for a white plane. (a) Phase difference between the middle-frequency unwrapped phase map using the vertical fringe images captured at the focus setting f_1 (i.e., ETL = −140.00 mA) and f_2 (i.e., ETL = −137.00 mA) before phase alignment. (b) Phase difference between the middle-frequency unwrapped phase map using the vertical fringe images captured at the focus settings as (a) after phase alignment. (c) Phase difference distribution of (a) (RMS = 0.0286 rads). (d) Phase difference distribution of (b) (RMS = 0.0047 rads).

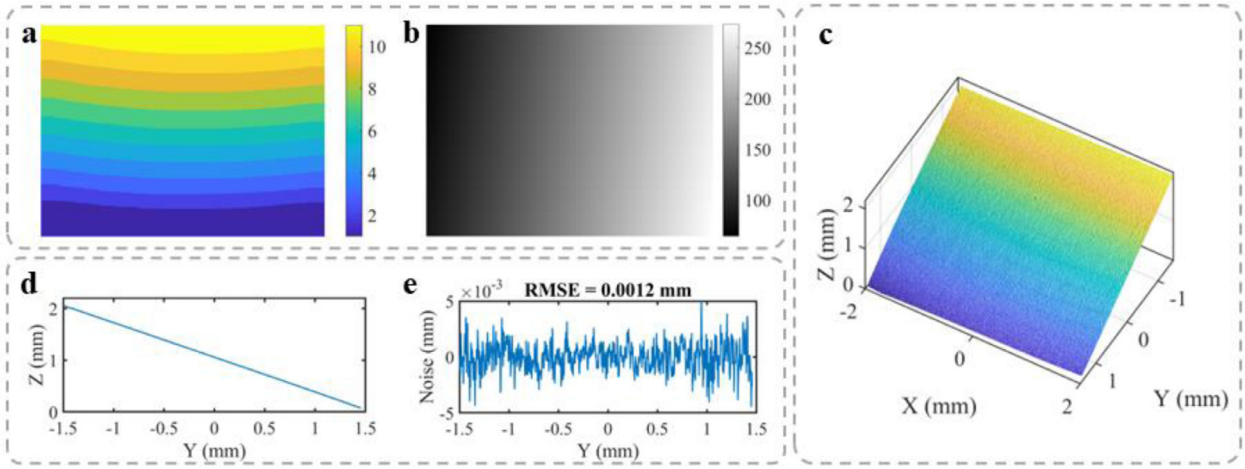

Fig. 8. Experimental results for a white plane measurement. (a) Generated index map for phase map stitching. (b) Stitched all-in-focus high-frequency phase map. (c) Reconstructed 3D shape. (d) A cross section of the reconstructed 3D shape along the 800th column. (e) Random noise after removing the gross profile of (d).

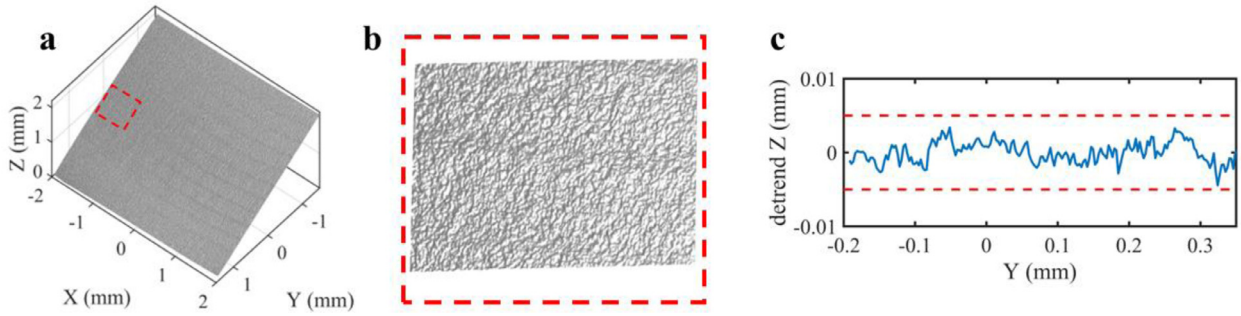

Fig. 9. Experimental evaluation of our proposed stitching method. (a) Reconstructed 3D shape using our proposed method. (b) A zoom-in view of (a). (c) A cross section after detrending the depth along the 150th column which is the center column of (b).

The detrended cross section shows that the surface roughness was less than 5 μm. Considering the random noise band shown in Fig. 8(e), the roughness was small, demonstrating that the reconstructed surface was smooth.

We further evaluated our proposed method by measuring a scene with complex geometry. Fig. 10(a) shows a photograph of the sample (i.e., the white flower) and a dime. We ran the whole process of our proposed method to measure the sample. Fig. 10(b) shows one of the captured high-frequency fringe images in the focal stack. Fig. 10(c) shows the generated index map and Fig. 10(d) shows the all-in-focus high-frequency phase map. The reconstructed 3D shape shown in Fig. 10(e) and a cross section along 550th row shown in Fig. 10(h) demonstrate that the shape of the sample was properly reconstructed. In comparison, some parts of the shape cannot be correctly reconstructed using only one single focal length as shown in Fig. 10(f) and (g). Here we masked out the invalid pixels whose high-frequency fringe contrast was less than

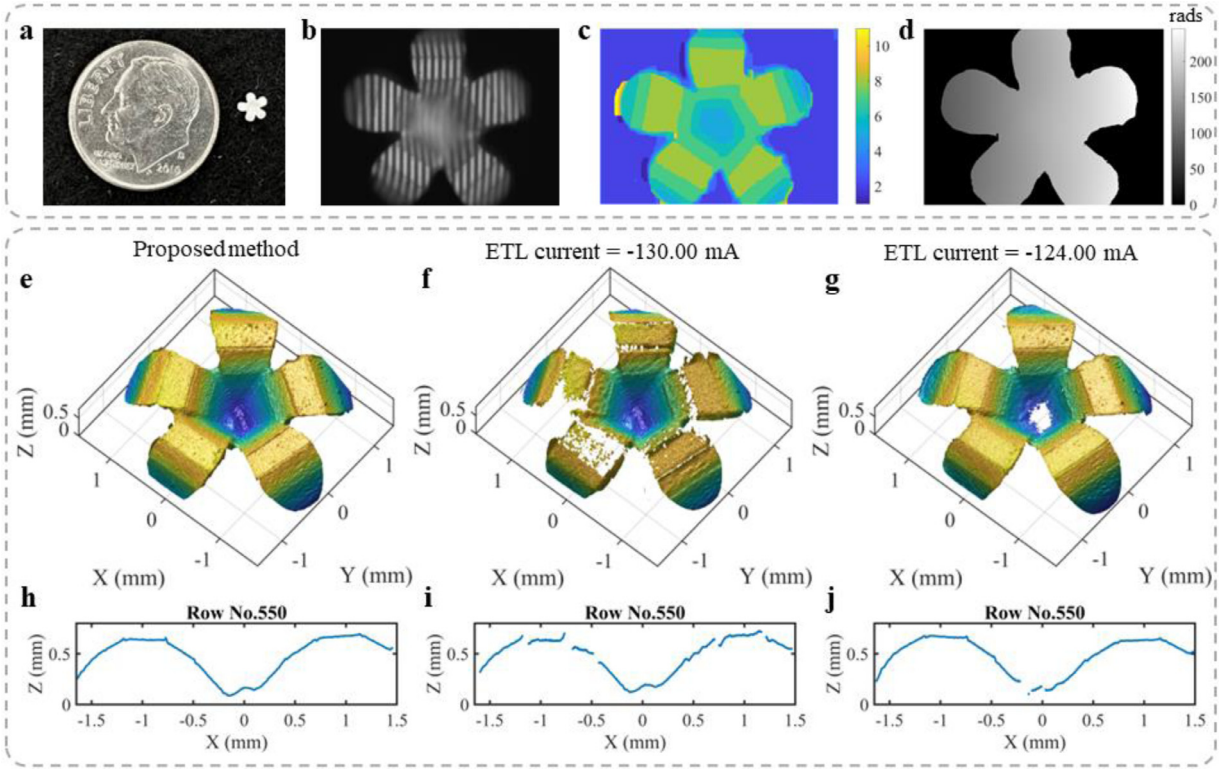

Fig. 10. Experimental results of measuring a scene with complex geometry. (a) A photograph of the sample put together with a dime. (b) One of the captured high-frequency fringe images in the focal stack. (c) Generated index map. (d) Stitched all-in-focus high-frequency phase map. (e) Reconstructed 3D shape using our proposed method (i.e., using the all-in-focus high-frequency phase map in (d)). (f) Reconstructed 3D shape using a single focal length (i.e., ETL current = 130.00 mA). (g) Reconstructed 3D shape using a single focal length (i.e., ETL current = 124.00 mA). (h) A cross section along the 550th row of (e). (i) A cross section along the 550th row of (f). (j) A cross section along the 550th row of (g).

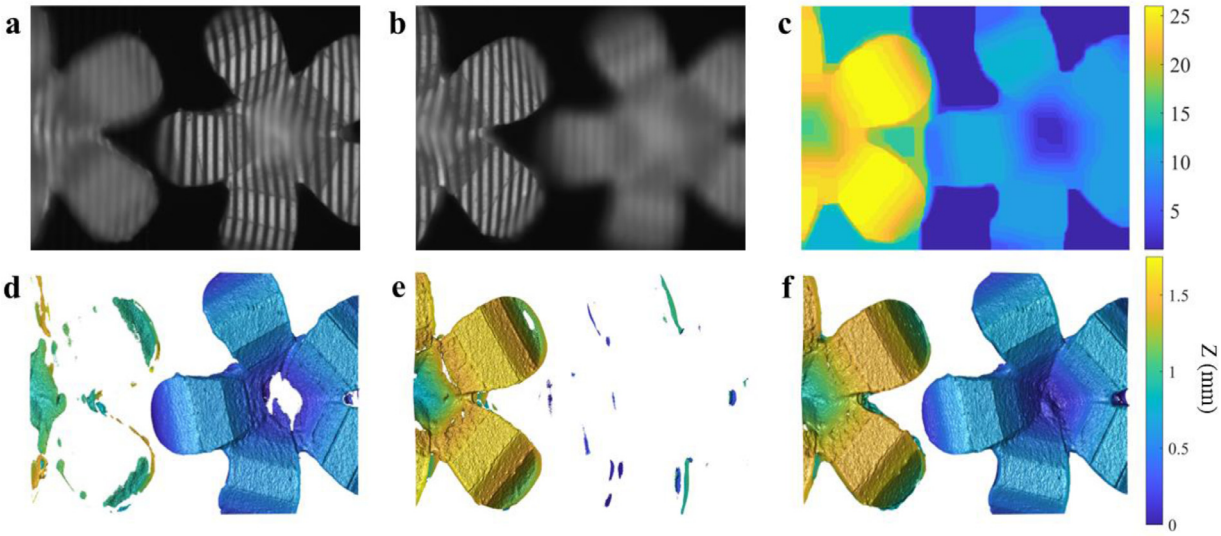

Fig. 11. Experimental results of measuring two samples at different heights. (a) One of the captured high-frequency fringe images at the focus setting f_3 (i.e., ETL current = −134.00 mA). (b) One of the captured high-frequency fringe images at the focus setting f_9 (i.e., ETL current = −119.00 mA). (c) Generated index map using our proposed method. (d) Reconstructed 3D shape using the fringe images shown in (a). (e) Reconstructed 3D shape using the fringe images shown in (b). (f) Reconstructed 3D shape using our proposed method.

0.08 (i.e., $\gamma(x, y) < 0.08$). The corresponding cross sections shown in Fig. 10(i) and (j) also demonstrate that some pixels were lost after 3D reconstruction due to the camera defocus.

We measured a scene with two samples similar to the previous experiment placed at different depths to test the enlarged DOF of our proposed method. Fig. 11(a) shows one of the captured high-frequency fringe images at f_3 when the ETL current was set as −134.00 mA. Similarly, Fig. 11(b) shows the image with the same pattern captured at f_9 when the ETL current was −119.00 mA. The reconstructed 3D shapes at these two focus settings are shown in Fig. 11(d) and (e), respectively. These results show that the system DOF was not enough to simultaneously measure these two objects. We then applied our proposed method to measure this scene. Fig. 11(c) shows the generated index map. The final reconstructed result is shown in Fig. 11(f), clearly showing that the two

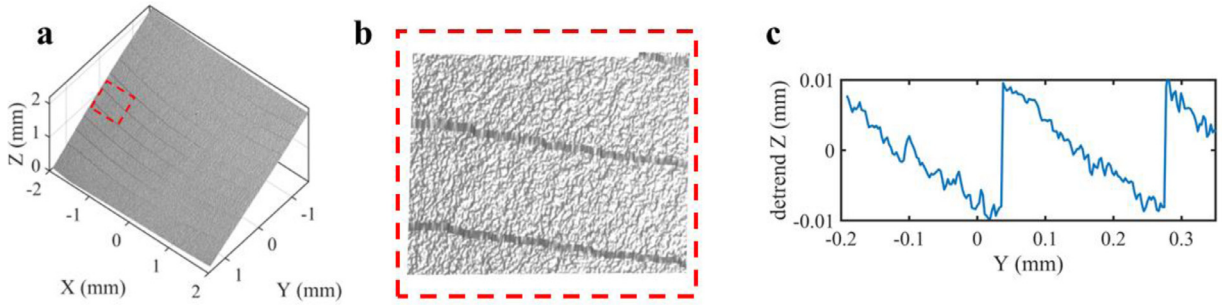

Fig. 12. 3D Reconstruction of the white flat plane measurement using the intensity-based ICIA algorithm. (a) Reconstructed 3D shape. (b) A zoom-in view of (a). (c) A Y-Z profile of one column of (b) after removing gross slope.

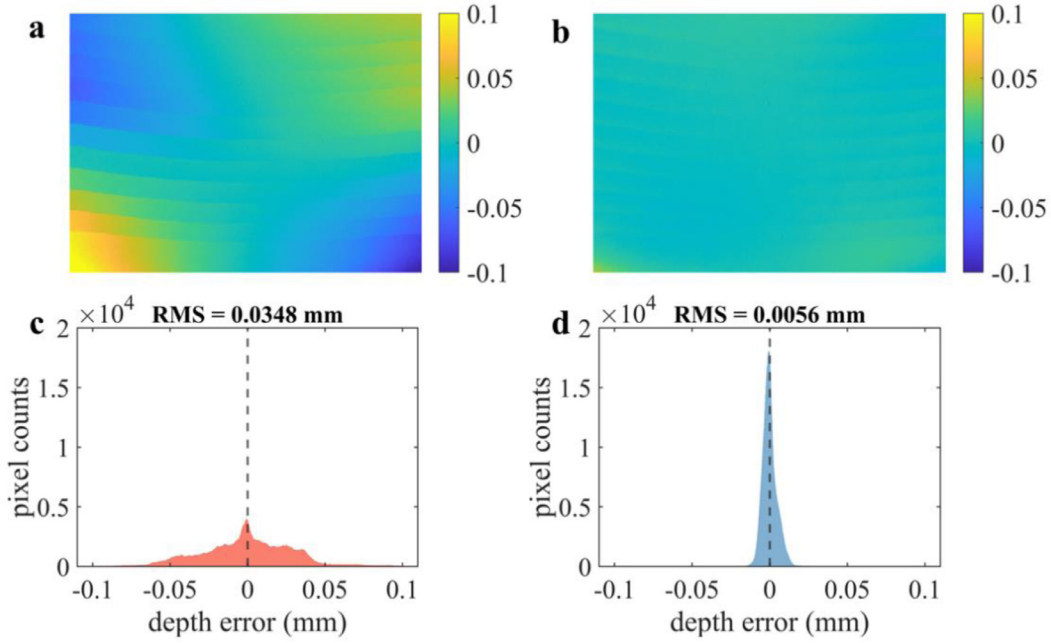

Fig. 13. Plane shape error analysis of using ICIA algorithm for the focal stack alignment and our proposed method. (a) Plane measurement error map from the reconstructed result from the ICIA algorithm. (b) Plane measurement error map from the reconstructed result from our proposed method. (c) The error distribution of (a). (d) The error distribution of (b).

samples were properly recovered. This experiment demonstrated that our proposed method worked well for the complex scene with a large DOF.

4. Discussion

In this paper, we proposed an accurate image alignment method using the phase of projected fringe patterns with optimized fringe frequency and an accurate image stitching method. Since lower-frequency fringe patterns are a less impact by more camera defocus, we used the phase maps from lower-frequency fringe patterns for image alignment. Comparing with image intensity, the phase is less sensitive to surface texture or the camera focus level. Our proposed phase-based image alignment method could achieve higher alignment accuracy than using the image intensity-based method. We aligned the same images of the white flat plane measurement shown in Fig. 9 using one of the most popular intensity-based image alignment method, the inverse compositional image alignment (ICIA) algorithm [40], then stitched the aligned images together using the same stitching algorithm. Fig. 12(a) shows the overall 3D reconstruction, Fig. 12(b) shows the zoom-in view and Fig. 12(c) shows the Y-Z profile of one column after removing its gross slope. Clearly, the surface is not smooth, demonstrating the white flat plane was not properly reconstructed if the intensity-based image alignment algorithm is used. This experiment demonstrated that the intensity-based image alignment method could fail without rich textures, whereas our proposed phase-based method still works well.

We also analyzed the overall shape error to further demonstrate the advantage of our proposed phase-based method comparing to the intensity-based method. The shape error is computed by the depth difference between the point cloud and its fitted ideal plane. The results shown in Fig. 13(a) clearly demonstrate that the reconstruction of using the ICIA algorithm contains large shape errors, as expected. Fig. 13(c) shows the error distribution with the RMS value of approximately 34.8 μm. In comparison, Fig. 13(b) shows the error map from our reconstructed result and Fig. 13(d) shows the error distribution that is close to be normal with the RMS value of approximately 5.6 μm, demonstrating that our proposed method achieved better performance.

We further proposed a high-accuracy phase image stitching method by developing optimization algorithms using both the fringe contrast and phase information. We propose to measure the focus level using the aligned high-frequency fringe contrast because it is more sensitive to camera defocus than the lower-frequency fringe contrast. Compared with focus measure operators based on image gradients, the fringe contrast is more robust to texture and camera defocus. To reduce the impact of large fringe contrast noise near the sharp discontinuities and/or dark regions, we combined the aligned high-frequency fringe contrast with

an energy minimization algorithm to obtain an optimized index map. We computed the weighted-average phase based on the optimized index map for each pixel to create the all-in-focus accurate phase map. Finally, we reconstructed 3D image of the scene from the all-in-focus phase map.

Our proposed method successfully enlarged the DOF of our 3D imaging system by approximately 10 times, which is limited of the projector's DOF. Therefore, future works can focus on simultaneously enlarging the projector DOF to further increase the system DOF.

5. Conclusion

We proposed a novel focus stacking method that can effectively enlarge the DOF of microscopic structured-light 3D imaging while preserving high depth resolution. We experimentally verified that our proposed method achieved a DOF of approximately 2 mm, which is approximately 10 × the camera DOF. The system has a FOV of approximately 4 mm × 3 mm with a pixel size at the object space of approximately 2.6 μm.

Author statement

This paper presents a novel focus stacking method that can effectively enlarge the depth of field of microscopic structured-light 3D imaging to approximately 10 times of the camera DOF, while preserving high depth resolution. The materials presented in this paper have not been published elsewhere. To the best of the author's knowledge, methodologies, data, and results discussed in this paper are accurate and original.

Declaration of Competing Interest

Corresponding author Professor Song Zhang serves as the CTO and co-founder of Vision Express Optics Inc (Auburn Hills, Michigan, USA). He serves as a consultant for Orbbec3D Inc (Troy, Michigan, USA), and ORI LLC (Salt Lake City, Utah, USA). This work conducted in this research was not supported by any of these entities. Co-author Liming Chen has no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Liming Chen: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Song Zhang:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Data availability

Data will be made available on request.

Acknowledgments

This work was sponsored by National Science Foundation (NSF) under the Grant No. IIS-1763689. Views expressed here are those of the authors and not necessarily those of the NSF.

References

[1] Windecker R, Fleischer M, Tiziani HJ. Three-dimensional topometry with stereo microscopes. Opt Eng 1997;36:3372–7.
[2] Zhang C, Huang PS, Chiang F-P. Microscopic phase-shifting profilometry based on digital micromirror device technology. Appl Opt 2002;41:5896–904.
[3] Proll K-P, Nivet J-M, Körner K, Tiziani HJ. Microscopic three-dimensional topometry with ferroelectric liquid-crystal-on-silicon displays. Appl Opt 2003;42:1773–8.
[4] Li B, Zhang S. Flexible calibration method for microscopic structured light system using telecentric lens. Opt Express 2015;23:25795–803 .
[5] Li D, Liu C, Tian J. Telecentric 3D profilometry based on phase-shifting fringe projection. Opt Express 2014;22:31826–35.
[6] Liu H, Lin H, Yao L. Calibration method for projector-camera-based telecentric fringe projection profilometry system. Opt Express 2017;25:31492–508.
[7] Yin Y, Wang M, Gao BZ, Liu X, Peng X. Fringe projection 3D microscopy with the general imaging model. Opt Express 2015;23:6846–57.
[8] Hu Y, Chen Q, Feng S, Tao T, Asundi A, Zuo C. A new microscopic telecentric stereo vision system - Calibration, rectification, and three-dimensional reconstruction. Opt Lasers Eng 2019;113:14–22.
[9] Li A, Peng X, Yin Y, Liu X, Zhao Q, Körner K, et al. Fringe projection based quantitative 3D microscopy. Optik 2013;124:5052–6.
[10] Li B, Zhang S. Microscopic structured light 3D profilometry: binary defocusing technique vs. sinusoidal fringe projection. Opt Lasers Eng 2017;96:117–23.
[11] der JSV, Soons JAM, Dirckx JJJ. Real-time microscopic phase-shifting profilometry. Appl Opt 2015;54:4953–9.
[12] Hu Y, Chen Q, Feng S, Tao T, Li H, Zuo C. Real-time microscopic 3D shape measurement based on optimized pulse-width-modulation binary fringe projection. Meas Sci Technol 2017;28:075010.
[13] Hu Y, Chen Q, Feng S, Zuo C. Microscopic fringe projection profilometry: a review. Opt Lasers Eng 2020;135:106192.
[14] Wang M, Yin Y, Deng D, Meng X, Liu X, Peng X. Improved performance of multi-view fringe projection 3D microscopy. Opt Express 2017;25:19408–21.
[15] Peng J, Wang M, Deng D, Liu X, Yin Y, Peng X. Distortion correction for microscopic fringe projection system with Scheimpflug telecentric lens. Appl Opt 2015;54:10055–62.
[16] Meng W, Quanyao H, Yongkai Y, Yang Y, Qijian T, Xiang P, et al. Large DOF microscopic fringe projection profilometry with a coaxial light-field structure. Opt Express 2022;30:8015–26.
[17] Chen L, Hu X, Zhang S. Calibration method for an extended depth-of-field microscopic structured light system. Opt Express 2022;30:166–78.
[18] Ramm R, Mozaffari-Afshar M, Höhne D, Hilbert T, Speck H, Kühl S, et al. High-resolution 3D shape measurement with extended depth of field using fast chromatic focus stacking. Opt Express 2022;30:22590–607.
[19] Burt PJ, Kolczynski RJ. Enhanced image capture through fusion. In: Proceedings of the 1993 (4th) international conference on computer vision; 1993. p. 173–82.
[20] Agarwala A, Dontcheva M, Agrawala M, Drucker S, Colburn A, Curless B, et al. Interactive digital photomontage. ACM Trans Graph 2004;23:294–302.
[21] Brown LG. A survey of image registration techniques. ACM Comput Surv 1992;24:325–76.
[22] Szeliski R. Image alignment and stitching: a tutorial. Found Trends Comput Graph Vis 2006;2:1–104.
[23] Clark D, Brown B. A rapid image acquisition method for focus stacking in microscopy. Microscopy Today 2015;23:18–25.
[24] Hasinoff SW, Kutulakos KN. Light-Efficient Photography. IEEE Trans. Pattern Anal. Mach. Intell. 2011;33:2203–14.
[25] Hasinoff SW, Kutulakos KN, Durand F, Freeman WT. Time-constrained photography. In: Proceedings of the 2009 IEEE 12th international conference on computer vision; 2009. p. 333–40.
[26] Liu Y, Fu Y, Zhuan Y, Zhou P, Zhong K, Guan B. Large depth-of-field 3D measurement with a microscopic structured-light system. Opt Commun 2021;481:126540.
[27] Zuo C, Huang L, Zhang M, Chen Q, Asundi A. Temporal phase unwrapping algorithms for fringe projection profilometry: a comparative review. Opt Lasers Eng 2016;85:84–103.
[28] Zhang S. Absolute phase retrieval methods for digital fringe projection profilometry: a review. Opt Lasers Eng 2018;107:28–37.
[29] Rao L, Da F. Local blur analysis and phase error correction method for fringe projection profilometry systems. Appl Opt 2018;57:4267–76.
[30] Gupta M, Nayar SK. Micro phase shifting. In: Proceedings of the 2012 IEEE conference on computer vision and pattern recognition; 2012. p. 813–20.
[31] Gupta M, Agrawal A, Veeraraghavan A, Narasimhan SG. Structured light 3D scanning in the presence of global illumination. CVPR 2011;2011:713–20.
[32] Hartley R, Zisserman A. Multiple view geometry in computer vision. 2nd ed. Cambridge: Cambridge Univ. Press; 2004.
[33] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 1981;24:381–95.
[34] Pertuz S, Puig D, Garcia MA. Analysis of focus measure operators for shape-from-focus. Pattern Recognit 2013;46:1415–32.
[35] Suwajanakorn S, Hernandez C, Seitz SM. Depth from focus with your mobile phone. In: Proceedings of the 2015 IEEE conference on computer vision and pattern recognition (CVPR); 2015. p. 3497–506.
[36] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Trans Pattern Anal Mach Intell 2001;23:1222–39.
[37] Kolmogorov V, Zabin R. What energy functions can be minimized via graph cuts? IEEE Trans Pattern Anal Mach Intell 2004;26:147–59.
[38] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. IEEE Trans Pattern Anal Mach Intell 2004;26:1124–37.
[39] Zhang S. High-Speed 3D imaging with digital fringe projection techniques. CRC Press; 2018.
[40] Baker S, Matthews I. Equivalence and efficiency of image alignment algorithms. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition; 2001. p. 1097. I-1090-I–.