

## RESEARCH ARTICLE

# Wafer Surface Reconstruction Based on Shape From Focus

SHAOHANG WANG<sup>1</sup>, JIQIANG CHEN<sup>2</sup>, HUA SHI<sup>2,3</sup>, AND MINGYE LI<sup>2,3</sup><sup>1</sup>School of Mechanical, Electrical and Information Engineering, Shandong University, Jinan, Shandong 264209, China<sup>2</sup>College of Mechanical and Energy Engineering, Beijing University of Technology, Beijing 100124, China<sup>3</sup>College of Engineering, China Agricultural University, Beijing 100083, China

Corresponding author: Jiqiang Chen (chenjiqiang2022@emails.bjut.edu.cn)

This work was supported by the National Natural Science Foundation of China.

**ABSTRACT** Scanning electron microscope, atomic force microscope and other equipment play an important role in the fields of topography restoration and detection. However, these devices are generally used in nanometer-scale measurement scenarios. For wafer topography quality control scenarios ranging from microns to hundreds of microns, these technologies have problems such as high cost and slow detection speed. Therefore, developing new, low-cost, and high-precision methods is necessary. To address this problem, a wafer surface reconstruction framework is proposed based on the shape-from-focus principle. In view of the characteristics of the large area and micro-small height of the wafer, to solve the limitations of the existing shape from focus framework, which is generally based on a single field, we created a multi-field image sequence rapid acquisition system and proposed the use of pulse control methods to achieve rapid acquisition of large area images. On the other hand, this paper proposes a dual filtering framework combining the Levy flight filtering principle with the SOR algorithm in point cloud filtering to achieve a balance between smoothing the depth map and maintaining the detailed structure, reducing the impact of noise, and improving the morphology restoration accuracy. To avoid splicing seams between fields, the progressive detection multifield stitching technique is used to complete large-area depth data stitching. Experiments were conducted on both synthetic and real objects to verify the effectiveness of the proposed method. In terms of synthesized images, the accuracy of the three methods significantly improved after applying the proposed method framework. After applying the Tenenbaum method framework, its correlation and peak signal-to-noise ratio improved by 7.5% and 38.2%, respectively, and its root mean square error was reduced by 40.7%. The excellent accuracy reconstruction results of the proposed method was verified through accuracy evaluation experiments. The height errors of the three methods used were all higher than 1  $\mu\text{m}$ . However, after using the proposed method framework, the maximum error was only 0.24  $\mu\text{m}$ . The experimental results indicated that this method overcomes the area limitation of traditional SFF and is suitable for applying wafer surface morphology measurements.

**INDEX TERMS** Wafer, shape from focus, Levy filtering, defect detection.

## I. INTRODUCTION

Dies are the core of a chip, which is cut from a wafer. Wafers are made by semiconductor technology, with numerous micro-sized dies arranged on the surface. After the wafer is made, there are inherent yield issues, and there may be various types of defects on its surface. Different detection

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaojie Su <sup>1</sup>.

techniques need to be used to identify defects on the wafer surface, classify and label them, and assist in sorting the wafers to prevent defective wafers from entering the subsequent packaging process. Detecting the three-dimensional geometric dimensions and shapes for surface processing control is necessary.

In existing wafer surface topography detection technologies, commonly employed methods include scanning electron microscopy, atomic force microscopy, laser confocal

microscopy, probe techniques, and various optical interference measurement technologies [1], [2]. However, these techniques are more commonly applied in environments requiring nanoscale surface quality control, such as roughness detection after wafer surface polishing, bow (BOW) detection, global flatness back ideal range (GBIR) detection, site flatness back ideal range (SBIR) detection, and percentage of local thickness variation [3]. These technologies achieve good detection accuracy, but their cost is too high. When used in wafer surface morphology detection environments with heights ranging from micrometres to hundreds of micrometres, their cost-effectiveness will be significantly reduced. Therefore, it is crucial to explore new, cost-effective, and high-precision methods for wafer topography quality control with heights ranging from micrometers to hundreds of micrometers. Shape from focus (SFF) is a noncontact optical measurement method with great development potential, requiring the use of only a single ordinary optical microscope and motion control system to achieve shape reconstruction. By controlling the microscope to move along the direction of the optical axis, the focusing surface scans the object surface along the optical axis direction, and optical slices are generated at different positions. The images of the slices are obtained by the camera to construct a focused image sequence. Then, the focus value of the image sequence is analyzed, and the maximum focus value is obtained in the optical axis direction to obtain the depth [4]. This method balances the convenience and accuracy of measurement.

The focus measure (FM) and depth estimation are important topics in the study of SFF. The former uses a certain mathematical standard to measure the focus value of pixels in the image space and generates a three-dimensional focus volume (FV), in which the discrete shapes of the object surface are hidden. The latter obtains discrete shapes through depth estimation algorithms. Currently, various types of FM operators have been reported. Pertuz et al. published an article analyzing FM operators and summarized the previously reported FM operators. They conducted experiments using simulations and real objects to analyze the sensitivity of existing operators to noise, image contrast, image saturation, and neighborhood window size [5]. According to different mathematical principles, Pertuz divided early focus operators into six types: gradient-based operators, Laplacian-based operators, statistics-based operators, DCT-based operators, wavelet-based or curved wave-based operators, and miscellaneous operators.

In recent years, researchers have proposed various distinctive and innovative approaches, placing greater emphasis on the application of new mathematical principles. There is also increased attention to the adaptability, robustness and accuracy of algorithms. Examples include the method of combining the spatial domain and frequency domain [6] and the idea of a tensor matrix [7]. Along the optical axis, the distribution of focus data is similar to a parabola, where the pixel points corresponding to the maximum value on the

parabola have the best clarity. Curve fitting techniques such as Gaussian fitting [8] and polynomial fitting [9] are often used to estimate the peak value. By using these methods to identify all clear pixel positions, discrete 3D shapes can be reconstructed. In addition, many other types of shape-from-focus algorithms have emerged, such as focus surface fitting algorithms [10], [11] and optimization-based algorithms [12], [13].

In recent years, deep learning has gained increasing attention and application in intelligent recognition, detection, and other areas, leading to the emergence of a growing number of deep learning algorithms [14]. For example, Tan et al. investigated the problem of reachable set estimation for delayed Markov jump neural networks with finite disturbances [15] and  $H_\infty$  state estimation for neural networks with time-varying delays [16]. They proposed improved inverse convex inequalities and the Lyapunov-Krasovskii functional (LKF), obtaining an accurate ellipsoidal description of the reachable set for delayed Markov jump neural networks. This has made significant contributions to the study of reachable set estimation in the field of control. Wu et al. addressed the issue of tomographic image reconstruction and proposed two networks: the deep embedding-attention-refinement (DEAR) network [17] and the dual-domain residual optimization network (DRONE) [18]. Both networks demonstrated unique advantages in tomographic image reconstruction, including edge preservation, feature recovery, and reconstruction accuracy.

The existing shape-from-focus framework generally focuses on a single field, which reconstructs only the area covered by the image sensor and restores only discrete shapes hidden in a single image sequence. Therefore, the area that these methods can detect is limited. In addition, a single image sequence contains fewer images, so existing shapes from focus frameworks pay less attention to time efficiency issues. However, the wafer area is large, and a single field cannot cover the measured area. The reconstruction mode relying solely on a single image sequence cannot meet the requirements. Therefore, it is necessary to divide the horizontal plane into multiple fields of view and use the same step size to capture image sequences in each field. All positions where the images are collected form a three-dimensional spatial grid. At this point, the number of images may be excessively large, which poses new requirements for the time efficiency of the shape from the focus framework.

We propose a wafer surface reconstruction framework based on the shape-from-focus principle to address the issues of wafer surface defect detection and size measurement. This framework includes large-area rapid image acquisition, 3D surface shape restoration, and postdata processing. This surface reconstruction framework is versatile, and the obtained 3D shape can be used for defect detection and 3D geometric size measurement in the latter stage. The main contributions of this method are reflected in the following three aspects: (1) By reconstructing the three-dimensional surface to obtain

depth data and utilizing these three-dimensional depth data for defect detection, the method captures a greater amount of information. This is superior to commonly used defect detection methods based on two-dimensional images. (2) This paper presents a low-cost, high-performance framework that can be built using only a conventional optical microscope and motion control. It is accessible to a wide range of users, in contrast to expensive options such as scanning electron microscopes, atomic force microscopes, confocal microscopes, and contour measurement devices, making it more affordable for a broader user base. (3) The traditional shape-from-focus methods are mostly based on a single field for reconstruction, and the reconstruction range is limited. The method proposed in this paper can realize the fast reconstruction of large-format wafer surfaces, which is very practical.

The rest of this paper is organized as follows. The next section provides a brief overview of the related work. Section III introduces the motivation behind the work. The proposed method of shape reconstruction for wafers is described in detail in Section IV. In Section V, the experimental results are analyzed to verify the effectiveness of the proposed method in this paper. The conclusions are given in Section VI.

## II. RELATED WORK

Shape-from-focus technology is used to restore the 3D shape of objects from multiple 2D image slices. These image slices are obtained by moving objects along the optical axis in a constant step size. The image slices at different focal positions contain all the object details. The purpose of the SFF method is to obtain the depth map of an object by calculating the optimal focusing position of each point on the object, thereby restoring the 3D shape of the object. For a given point  $P$ , its optimal focusing position satisfies the thin lens imaging law, and its principle is as follows:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \quad (1)$$

where  $f$  represents the focal length of the lens,  $u$  represents the distance from the object point  $P$  to the lens, and  $v$  represents the distance between the lens tube and the image plane.

The measurement of pixel focus is a very important step in the entire SFF, which is related to the quality and accuracy of SFF. Previous studies have shown that for different applications, such as SFF, autofocus, and depth of field stacking, various focus operators have been proposed in the literature. Pertuz et al. [5] summarized and analyzed these FM operators and classified them into six different types based on their different mathematical principles. They are gradient-based operators, Laplacian-based operators, statistics-based operators, DCT-based operators, wavelet-based or curved wave-based operators, and miscellaneous operators. Our goal is to achieve rapid detection of large-area wafer image sequences using shape-from-focus methods. Therefore, we chose several commonly used FM

operators, such as summarized modified Laplacian (SML), Tenenbaum (TEN), and gray-level-variance (GLV). SML is a modified Laplace operator that performs better with less time cost compared to other FM operators. Therefore, it is widely used for image clarity calculation. SML involves convolving the image sequence  $I_k^n(x, y)$  with the Laplacian operator. However, the result obtained from the Laplacian operator is signed, and under certain conditions, there is a possibility of opposite signs, leading to a phenomenon of mutual cancellation between them. To address the impact of this issue, Nayar and Nakagawa [8] proposed calculating the focus value for each pixel by calculating the sum of the squared absolute values of the second-order derivatives in the  $x$  and  $y$ -directions. Therefore, for a pixel point  $p(x_0, y_0)$  in the  $k$ -th frame image, the expression for calculating the focus value is:

$$FM_{SML}^{(k)} = \sum_{(x,y) \in U(x_0,y_0)} \left( \left| \frac{\partial^2 I_k^n(x, y)}{\partial x^2} \right|^2 + \left| \frac{\partial^2 I_k^n(x, y)}{\partial y^2} \right|^2 \right) \quad (2)$$

where  $I(x,y)$  is the input image frame,  $U(x_0,y_0)$  is the neighborhood centered around the pixel point  $(x_0,y_0)$ , and  $(x,y)$  is a pixel point in the neighborhood  $U$ . The TEN technique [19] is based on the Sobel operator, which is defined as the sum of squares of the gradients of the pixel points. The gradient values in the horizontal and vertical directions are extracted using the Sobel operator, and for a pixel point  $p(x_0, y_0)$  in the  $k$ -th frame image, its focus value can be expressed as:

$$FM_{TEN}^{(k)} = \sum_x \sum_y G_x^2(x, y) + G_y^2(x, y) \quad (3)$$

where  $x$  and  $y$  belong to the neighborhood  $U(x_0,y_0)$ , respectively, and  $G_x$  and  $G_y$  are gradient values computed using the Sobel operator. The GLV operator is a statistically based evaluation operator, which is one of the more accurate FM operators in sharpness detection methods, by calculating the variance in the gray values, with a higher variance representing higher image sharpness [20], and has been widely used in the application of FMs for SFF. For a pixel point  $p(x_0, y_0)$  in the  $k$ -th frame image, its focus value can be calculated using the following formula:

$$FM_{GLV}^{(k)} = \sum_{(x,y) \in U(x_0,y_0)} \frac{1}{N^2} [I_k^n(x, y) - \mu]^2 \quad (4)$$

where  $\mu$  is the average of all pixels in the neighborhood,  $U(x_0, y_0)$  is the neighborhood in which the point  $(x_0, y_0)$  is located, and  $N$  is the number of pixels in the neighborhood  $U$ .

## III. MOTIVATION

In recent years, many methods have been proposed for defect detection and size measurement on wafer surfaces with a height range of microsmall heights. For example, Zheng et al. proposed a wafer surface defect detection method based on background subtraction and faster R-CNN for wafer surface defect detection [21]. These methods have shown excellent

effects in defect detection and two-dimensional dimension measurement. However, with the advancement of semiconductor technology, there is a new demand for defect detection on wafer surfaces and the measurement of three-dimensional geometric dimensions. In existing wafer surface shape detection technologies, commonly used devices include scanning electron microscopes, atomic force microscopes, laser confocal microscopes, contact probe technology, and various optical interference measurement technologies. These devices offer high precision and are often employed in nanoscale detection environments. However, their high cost poses a challenge. Given the need for surface morphology detection and three-dimensional dimension measurement in the height range of micrometres to a few hundred micrometres, exploring novel, low-cost, high-precision methods is crucial.

For the problem of three-dimensional geometric dimension measurement, Chen et al. proposed a surface contour measurement technique [22]. This technique utilizes the Mohr projection scanning technology based on the principles of triangulation, combined with a dual optical microscopy system, to achieve a three-dimensional reconstruction of the wafer surface. It effectively addresses the reconstruction challenges of wafers with both diffuse and specular reflections. However, the overall system is complex and requires intricate calibration work. In contrast, SFF methods do not necessitate a complex system structure, offering a balance between measurement convenience and accuracy.

## IV. PROPOSED METHOD

### A. SYSTEM AND RECONSTRUCTION PROCESS

The principle of the conventional single-field SFF is given in Fig. 1. The core of the hardware system is the vision system and the Z-axis motion system; the vision system is used for image acquisition, and the Z-axis motion system is used to control the vision system to move along the Z-direction. The focusing surface of the objective lens of the vision system scans the surface of the object while moving. The single-field image acquisition process follows the following flow:

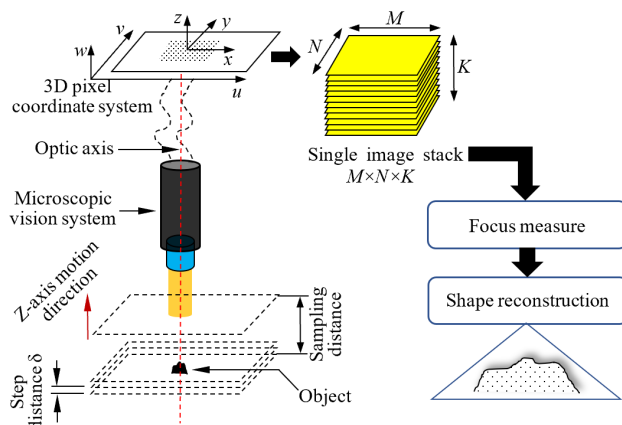


FIGURE 1. Traditional shape from focus in a single field field.

**Step 1:** The parameters such as step  $\delta$  and sampling distance are set, and the focusing surface of the objective lens is moved to some reference position (dashed box) below by the Z-axis motion system in preparation for image acquisition;

**Step 2:** Through the Z-axis motion system, the vision system is made to move a distance  $\delta$  along the direction of the optical axis, and the image is captured after stopping to obtain the first frame of the focus image. Then, the vision system is made to continue to move upward, and the Step 2 process is repeated until the distance of the vision system is out of the position of the sampling distance, and it enters Step 3.

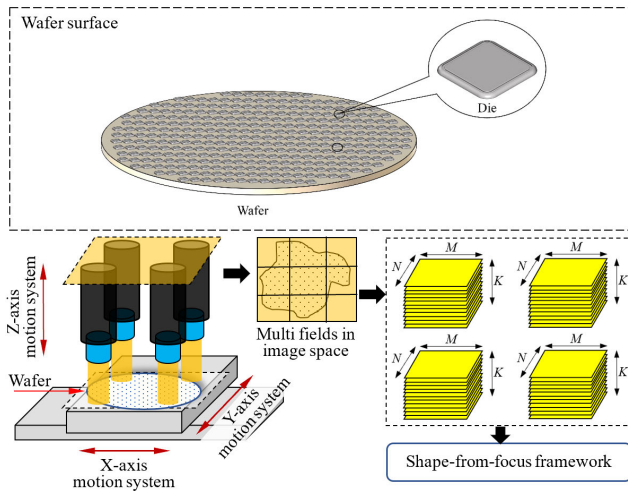
**Step 3:** The image acquisition process ends.

Through the above image acquisition process, an image sequence containing  $K$  image frames will be output. If the resolution of a single image is  $M \times N$  pixels, the number of pixels contained in the image sequence is  $M \times N \times K$  pixels. In an image sequence, the pixels in different images have spatial positional relationships, and a three-dimensional pixel space coordinate system  $o-uvw$  is created, where the  $u$ -axis and  $v$ -axis are also the coordinate axes in a single image, and the  $w$ -axis is parallel to the optical axis of the visual system so that the pixels in the different image frames form a three-dimensional pixel space. If given a pixel's lateral coordinate position  $(m, n)$ ,  $K$  pixels exist along the  $w$ -axis direction, and the  $k$ -th pixel's coordinate is  $(m, n, k)$ , these  $K$  pixels describe the out-of-focus process of a certain position in the lateral plane in the object space, the degree of focus is usually used to describe the focusing level of these  $K$  pixels, the pixel space is transformed into the FV, pixels corresponding to the maximal degree of focusing are regarded as the pixels that focus the most sharply, and the process of using their depth coordinate to restore the discrete shapes hidden in the FV is the shape recovery process.

In Fig. 2, the surface structure of a wafer can be seen, with a large number of micro-sized wafers lined up on it, blocks for cutting left between wafers. The transverse and longitudinal sizes of the wafers may be on the order of microns, tens of microns, and so on, depending on the requirements. Additionally, the microvision system constituted by microscopes is needed to take images properly when inspecting the surface attributes. The diameter of the wafer is large and can be 4, 6, 8, or 12 inches, and so on. The field of the microscopic vision system is limited, so it is not possible to capture images of the entire area with only a single field of view; it is necessary for the object to move or for the vision system to be able to capture images in more than one field. Based on the above requirements, a framework for SFF based on multiple fields is proposed, as shown in Fig. 2. The area to be photographed is divided according to the field range of the vision system and the area of the wafer surface. Partial overlap is needed between neighboring areas, the image sequences are acquired in each individual field, a collection of the image sequences is output, and the shape recovery is carried out based on the collection of the image sequences.

Based on the wafer surface characteristics and the measurement range demand, we designed a hardware system





**FIGURE 2.** Rules of shape from focus in multiple fields for wafer surface profile.

and a general algorithm framework for wafer surface shape recovery, as shown in Fig. 3. The hardware system consists of a microscopic vision system (camera, microscope and light source), a multi degree-of-freedom motion system (Z-axis translational stage, X&Y-axis translational stage, rotational stages A and B), and vibration damping equipment. The vision system is used to acquire the image sequence, the motion system is used to adjust the position and attitude of the vision system and the object, and the vibration damping equipment is used to eliminate the external vibration factors in the image acquisition process to ensure a smooth image acquisition environment. The multidegree-of-freedom motion system is controlled by a specialized motion control system. In addition, a pulse control system was designed to ensure fast image acquisition under multiple fields, which allows image sequences to be captured during camera movement.

Using the hardware system in Fig. 3, the area where the wafer is located can be divided into multiple fields, image sequences can be acquired in each field, and then these image sequences can be processed. The whole shape recovery process can be summarized into four steps (Step 1-Step 4): image stack acquisition, shape recovery and data optimization, multifield stitching, measurement and application. The above steps are analyzed and described in detail in Sections IV-B, IV-C and IV-D, respectively.

### B. METHOD FOR RAPID ACQUISITION OF WAFER IMAGE SEQUENCES

In SFF, the image sequence is fundamental to the implementation of the shape-from-focus method. It has been shown that there are three main methods for acquiring the image sequence: changing the focal length [23], the image distance [24] and the object distance [25]. Mutahira et al. [26] compared and analyzed three image sequence acquisition methods and showed that changing the focal length method

has significant parallax and scaling problems that affect reconstruction accuracy. This method is very challenging. Changing the image distance is difficult; the image plane requires precise alignment with the optical axis during movement, which can easily cause serious parallax problems. Changing the object distance is more commonly used and is easier to implement in terms of motion control. In this paper, image sequences are acquired by changing the object distance.

The image sequence quality has a very important impact on depth estimation accuracy and usefulness. The main factors affecting image sequence quality are the total number of images, the range of acquired samples, step  $\delta$ , magnification, the depth of field and the acquisition speed.

The step size and acquisition speed are particularly important, as too small a step size will result in errors when positioning clear pixels with maximum focus; too large a step size will result in partial loss of depth information. Muhammad and Choi [27] addressed this problem by proposing a sampling criterion for image sequences, which suggests that the sampling step should be proportional to the depth of field of the system, and proposed a formula for the sampling step.

Most of the research on the existing shape-from-focus methods focuses on the sequence recovery of a single field, and the acquisition method is mostly a step-by-step acquisition by setting up the step distance  $\delta$ , relying on the motion device to make the focusing surface of the objective lens move along the direction of the optical axis by one step, performing image acquisition after the motion device completely stops, continuing to move to the next position by the same step distance, and repeating the above steps for a cyclic operation. This method requires a pause after capturing each frame, and precise positioning is needed after each capture. This collection approach consumes considerable time, contradicting the proposed low-cost, high-performance characteristics. For wafers, the inspection size range is so large that a single field at 20X can no longer cover the range of a single wafer in a wafer. Therefore, it is necessary to divide multiple fields in the transverse plane of the area to be measured and acquire image sequences with equal steps for each field. However, the image sequences increase exponentially under multiple field conditions, and the time cost increases exponentially upon accumulation, which seriously affect the application of focused shape recovery in real scenes.

To address this problem, we propose a scheme for the fast acquisition and localization of large-area image sequences. First, the X, Y and Z axes of the hardware system shown in Fig. 3 are equipped with high-precision scales, which can output pulse signals and achieve micron-level positioning accuracy. The high-precision 3D motion control system provides precise panning motion and positional positioning, in which the Z-axis translational stage controls the camera's longitudinal movement along the direction of the optical axis. Additionally, the X- and Y-axis translational stages can realize the object's panning operation. Since a single field cannot cover the entire measured area, it is necessary to

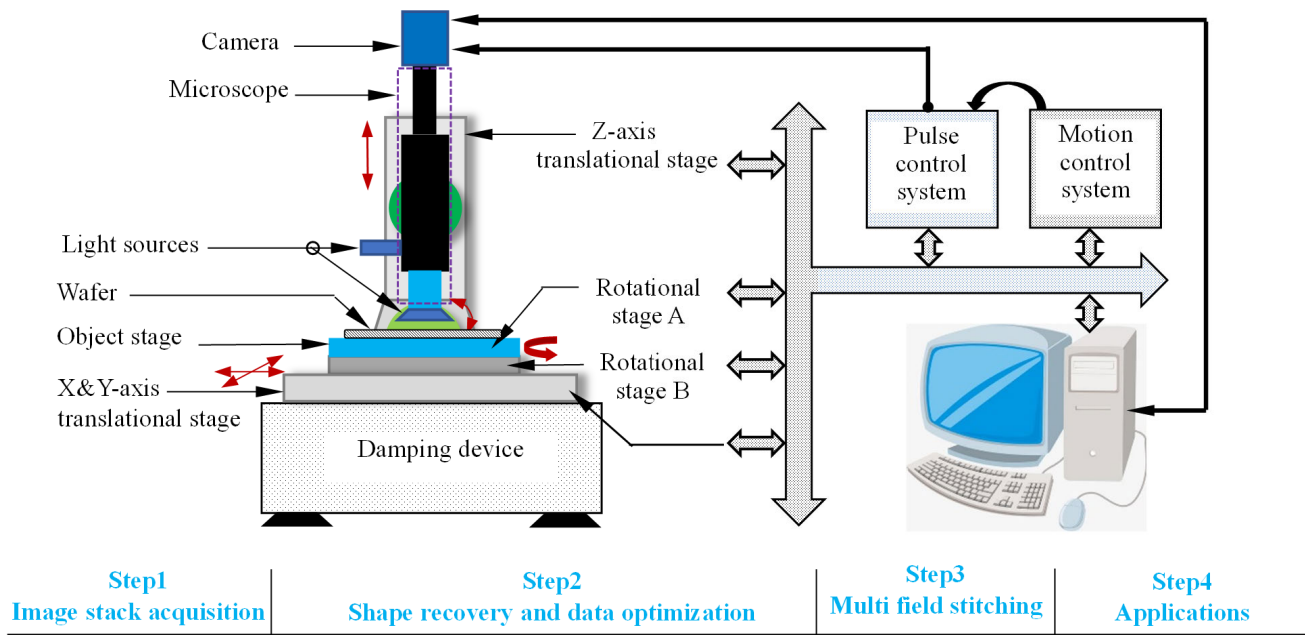


FIGURE 3. System setup and shape from the focus frame for wafer surface reconstruction.

calibrate the single-field range and divide the entire measured area into an array of image sequences with multiple rows and columns along the lateral direction. By numbering the multiple fields in advance, field switching can be realized quickly. The pulse counting device can read the number of pulses output from the scale in real time during the motion process, and the corresponding length of each pulse is calibrated in advance. The pulse counting device further processes the number of pulses read according to preset parameters to generate a signal for triggering the image sequence acquisition and sends it to the camera for real-time triggering. Triggering the camera through the grating ruler and pulse counting method to achieve image acquisition and position localization is very accurate. In addition, the acquisition strategy is not the traditional method to acquire image sequences only in the direction of the optical axis. We designed a three-axis pulsed image sequence acquisition scheme, which allows us to acquire image sequences not only along the X-axis direction but also along the Y-axis direction under multifield conditions. However, under multifield conditions, the time cost is inconsistent under different acquisition strategies due to factors such as the field-of-view range, step size, and number of frames to be acquired. To rapidly acquire sequences in multiple fields, we add the automatic discrimination and calculation of different parameters in the scheme to select the optimal acquisition strategy. Finally, we perform image sequence acquisition in equal steps according to the proposed scheme to obtain multiple multifield image sequences  $I_k^n(x, y)$  with dimensions  $M \times N \times K$ . where  $x \in \{1, 2, \dots, M\}$  and  $y \in \{1, 2, \dots, N\}$  denote the number of rows and columns of each image frame,  $k \in \{1, 2, \dots, K\}$  represents

the frame position where the image is located in the image sequence, and  $n$  denotes the position number of the multifield image sequence.

We propose a framework for rapidly acquiring image sequences of wafer samples in the surface shape detection environment of wafer samples ranging in height from micrometers to hundreds of micrometers. To achieve low cost and high precision, we use a system that meets the accuracy requirements of wafer samples ranging from micrometers to hundreds of micrometers. However, for the detection environment of the surface shape of nanoscale wafer samples, the displacement table used with a piezoelectric ceramic-driven translational stage is replaced to meet acquisition accuracy requirements. Therefore, this method framework has good compatibility and can meet the surface morphology detection function of wafer samples with heights ranging from micrometers to hundreds of micrometers and at the nanoscale according to actual needs.

### C. APPEARANCE RESTORATION METHODS

Image sequence acquisition is the shape basis of the focus technique, and focusing measurement and depth estimation are the two core elements of SFF. In Section IV-B, a multifield image sequence  $I_k^n(x, y)$  is acquired, and there are focused and out-of-focus pixel points in each frame of the image sequence. A well-focused image will have more high-frequency content. Thus, for each pixel point on the image sequence, a criterion is applied that allows for the efficient calculation of the focus quality, which is referred to as the FM. The calculated focus quality result is referred to as

the FV. This criterion is applied to amplify the focus value of well-focused high-frequency content and suppress the focus value of out-of-focus parts.

For an image sequence  $I_k^n(x, y)$ , we need to perform a focusing degree computation for each pixel point using the FM operator to transform it from pixel space to focusing degree space to obtain the FV and maintain the dimension consistent. It can be expressed as:

$$F_k^n(x, y) = FM \otimes I_k^n(x, y) \quad (5)$$

After calculating the FV of the image sequence, all the focusing values are counted for each pixel point in the focusing degree space along the direction of the optical axis, and the curve formed by all the focusing values is called the focusing degree curve; ideally, the focusing degree curve has a Gaussian distribution with a single peak, so the seeking peaks algorithm is generally utilized to obtain the maximum focusing value; its position is recorded as the depth data. Its formula can be expressed as:

$$D^n(x, y) = \operatorname{argmax} |F_k^n(x, y)| \quad (6)$$

The initial depth map can be obtained from the image sequence by the above formula to recover the discrete 3D shape. After interpolation, filtering and other operations to obtain the 3D surface shape of the object, splicing operations on the depth data using splicing techniques to obtain a complete 3D shape recovery, filtering techniques and depth data splicing are specifically introduced in the next section.

#### D. DEPTH DATA SPLICING AND FILTERING

As described in Section IV-C, the FM operator is utilized to calculate the focusing value. Then, the depth estimation is accomplished by finding the location of the maximum focusing value. This metric is based on the edge information; however, with the introduction of different noises in each section, the edge information is disturbed, and with the introduction of different types of noise, the depth maps obtained by a single noise reduction algorithm will be excessively smoothed or have a poor effect in removing the noise. To filter out this noise and address this problem, we first use the concept of Levy flight for noise filtering and then utilize statistical filtering for secondary noise reduction in the point cloud. The concept of Levy flight was first proposed by French mathematician Paul Levy. Levy flight mimics nature-based and physics-based phenomena [28], such as animals in nature, such as fruit flies that search for food randomly. These behaviors exhibit characteristics typical of Levy flight. Two features that must be specified to realize Levy flight are the step size and the wizard that guides the function realization. Mantegna proposed a simple method for determining step size  $Z$  [29], which can be expressed as:

$$Z = \frac{U}{|V|^{\frac{1}{\beta}}} \quad (7)$$

where  $\beta$  is the distribution index ( $0 < \beta \leq 2$ ),  $U$  and  $V$  are normally distributed variables with expectation 0 and standard deviation  $\sigma_u$  and 1, respectively, and  $\sigma_u$  can be expressed by the following equation:

$$\sigma_u = \left( \frac{\Gamma(1 + \beta) \cdot \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \cdot \beta \cdot 2^{\left(\frac{\beta-1}{2}\right)}} \right)^{1/\beta} \quad (8)$$

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt \quad (9)$$

Here, to achieve filter denoising optimization of a 3D point cloud, we define the step size as a matrix function as follows:

$$L_d = \alpha \cdot \operatorname{sign}(R) \cdot |z| \quad (10)$$

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \cdots & r_{m,n} \end{bmatrix} \quad (11)$$

where  $R$  is a matrix of random variables and  $\alpha$  is a scale parameter to control the Levy step size. For the initial depth map  $D^n$ , the depth map  $\widehat{D}_1^n$  after the first iteration of filtering and noise reduction using Levy flight can be expressed as:

$$\widehat{D}_1^n = D^n + L_d \quad (12)$$

However, due to the stochastic nature of the Levy flight method, to determine whether the current results meet the accuracy requirements, the mean square deviation function is used as its fitness function as a criterion for whether to update the current data. The fitness function can be expressed as:

$$\begin{cases} \chi_{ini} = \frac{1}{N} \sum \left[ \left\| D^n - \widehat{D}_t^n \right\|^2 \right] \\ \chi_{new} = \frac{1}{N} \sum \left[ \left\| D^n - \widehat{D}_{t+1}^n \right\|^2 \right] \end{cases} \quad (13)$$

where  $\chi_{ini}$  is the mean square error value after the  $t$ th iteration,  $\chi_{new}$  represents the mean square error value after the  $t+1$ st iteration,  $N$  is the size of the input initial depth map  $D^n$ , and  $\widehat{D}_t^n$  and  $\widehat{D}_{t+1}^n$  are the output depth maps after the  $t$ th and  $t+1$ st iterations. Throughout the iteration process, if the mean square deviation  $\chi_{new}$  of the data after  $t+1$  iterations is less than the mean square deviation  $\chi_{ini}$  after  $t$  iterations, we accept the current data; otherwise, we proceed to the next iteration. It can be expressed as:

$$\widehat{D}_{t+1}^n = \begin{cases} \widehat{D}_{t+1}^n & \chi_{new} < \chi_{ini} \\ \widehat{D}_t^n & \chi_{new} \geq \chi_{ini} \end{cases} \quad (14)$$

Then, the point cloud data are iteratively completed using the spatial distribution-based SOR denoising algorithm for each point's neighborhood statistics, eliminating the outliers that do not meet the conditions to complete the secondary noise reduction. We conduct a  $K$ -nearest neighborhood statistical analysis for each point, calculating the average distance from the point to its  $K$ -nearest neighbors, as well as the standard deviation of the distances. Subsequently, a threshold

$T$  is set, and points beyond this given threshold are filtered out and set to null.

In existing methods, it is common to set the nullified data to zero or a slightly larger value than zero. However, in a microscopic environment, even objects at small heights typically exhibit complex surface structures. Therefore, setting missing data to a fixed value appears to be unreasonable. To address this issue, we adopted a spline curve interpolation fitting method to interpolate missing data. Initially, the depth data are interpolated in the row direction using cubic spline curve fitting, supplementing missing data values. Then, the processed depth data are interpolated in the column direction using cubic spline curve fitting, which allows for a column wise interpolation to smooth the depth data after the previous supplementation, reducing the noise impact brought about by the interpolation fitting and enhancing the accuracy of the depth data. Finally, the ultimate depth map  $\check{D}$  is obtained.

Finally, since the depth data acquired in multiple fields of view are distributed in chunks, we need to perform depth data stitching on the obtained depth maps to obtain the 3D surface topography of the complete magnitude. However, the depth data in multiple fields come from different reference systems that are related in terms of position and attitude, which requires us to calibrate them uniformly and establish the transformation relationship between them. Two coordinate systems have been established in Section IV-A: the object-space coordinate system  $O$ - $XYZ$  and the image-space coordinate system  $o$ - $uvw$ . The correspondence between the two needs to be calibrated in advance to obtain the corresponding parameter pixel ratio  $\varepsilon$ . To eliminate the splicing traces between blocks, we use the progressive splicing strategy, where data from two neighboring fields of view need to be partially data fused during the splicing process. In general, the fusion area is set to 10% or 20%, with the maximum percentage not exceeding 50%. We first perform depth data splicing along the X-direction, and for the data in the fused region, we use a weighting method to determine the actual data for each pixel. The weighting coefficients are calculated by a linear method with the following expression:

$$\begin{cases} w_1(x) = \frac{x_R - x}{x_R - x_L} \\ w_2(x) = \frac{x - x_L}{x_R - x_L} \end{cases} \quad (15)$$

where  $x_R$  and  $x_L$  represent the horizontal coordinates where the right boundary of image A and the left boundary of image B are located, respectively, and  $x$  represents the horizontal coordinate where the current pixel is located. Let the depth data of image A and image B in the overlapping region be  $D_A(x)$  and  $D_B(x)$ ; then, the spliced data  $D(x)$  are:

$$D(x) = w_1(x) D_A(x) + w_2(x) D_B(x) \quad (16)$$

We perform an aggregation operation on the depth data of all the fields along the X-direction and then follow the same principle to perform a stitching operation in the Y-direction. The complete depth data are finally obtained.

Our algorithm outlines the key steps taken to reconstruct shape.

---

#### Algorithm Process Framework for Proposing Methods

---

- 1: **Input:** image stacks in multifields.
  - 2: **for**  $N = 1: n$  **do**
  - 3:     Calculate the focusing volume  $FM(i, j, k)$  using FM
  - 4:     Calculate initial depth map according to Eq. (6)
  - 5:     Filtering of depth data according to Eq. (12)-Eq. (14)
  - 6:     Iterative operation for step 4 using the fitness function
  - 7:     depth map  $D(x, y)$  & All-In-Focus image  $I(x, y)$
  - 8: **end**
  - 9: Multifield depth data and all-in-focus image stitching
  - 10: **Output:** Final depth map  $D_{fin}(x, y)$  & All-In-Focus image  $I_{AIF}(x, y)$
- 

## V. EXPERIMENTS AND DISCUSSIONS

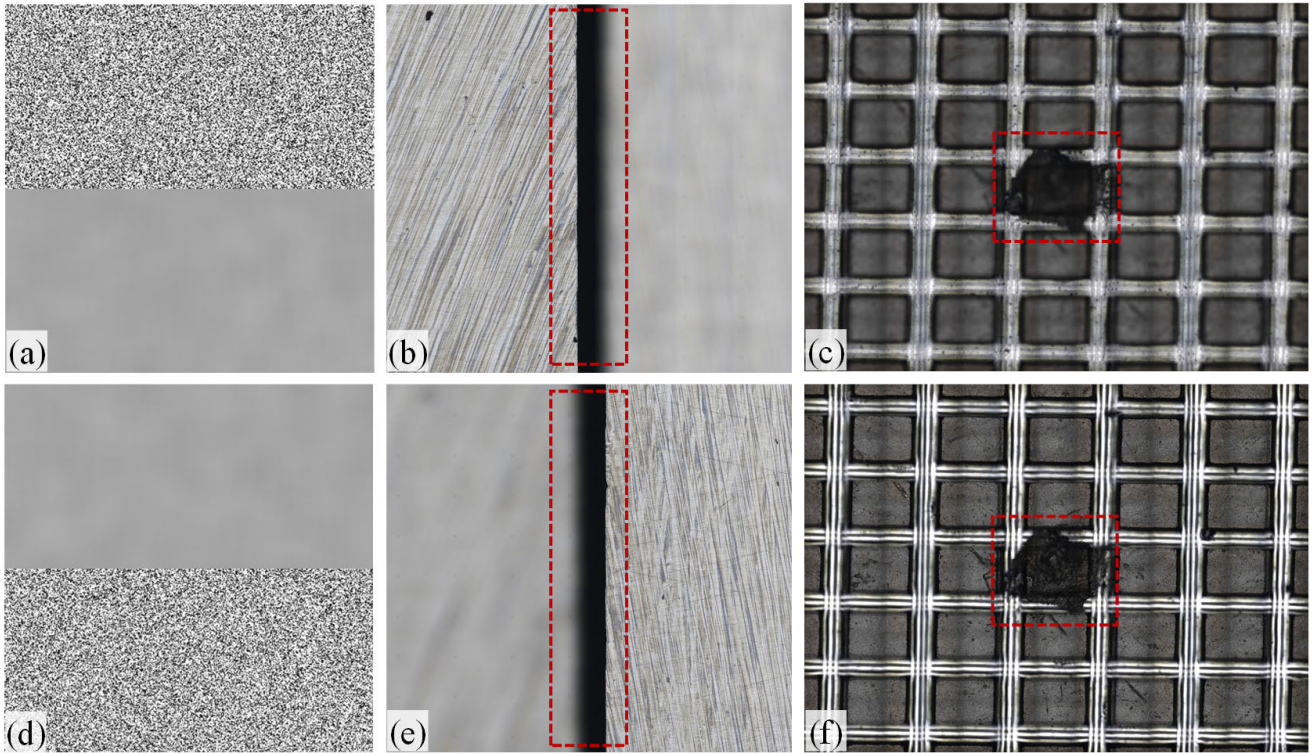
In this section, the experimental results are analyzed and discussed in detail in three parts. First, the detailed experimental setup and evaluation criteria are given. Second, the qualitative analysis of the proposed method is completed by the reconstruction effect of the samples and the cross-section curves of the wafer samples. Finally, the quantitative evaluation of the proposed method is completed by analyzing the reconstruction accuracy and noise robustness.

### A. EXPERIMENTAL SETUP

Experiments were conducted on a synthetic image sequence and two real image sequences to analyze the performance of the proposed method. The image sequence of the simulated ladder was synthesized using the simulation program, and both real object image sequences were taken under the microscopic vision system mentioned in Fig. 3. The system consists of an automatic zoom lens (Mvotem Optics MAZ7.0X), metallographic objective, CMOS industrial camera (Hikvision industrial camera) and a high-precision panning stage. The automatic zoom lens has a zoom ratio of 0.7X-4.5X; we chose a 10X metallurgical objective lens. The full-resolution image size collected by the industrial camera is  $1080 \times 1440$  pixels, the size of each pixel unit is  $3.75 \mu\text{m} \times 3.75 \mu\text{m}$ , and the frame rate can be up to 60 frames at full resolution. The real image sequences used in this paper were all acquired at 20X, the system field at 20X was only  $504.67289706 \mu\text{m} \times 672.89719608 \mu\text{m}$ , and all real image sequences were acquired using the fast image sequence acquisition method mentioned in Section IV-B.

Fig. 4 shows two sample images used at different frames. Among them, Fig. 4(a) and Fig. 4(d) are two frames of the simulated step image sequence, and Fig. 4(b) and Fig. 4(e) are two frames of the image sequence used for the height





**FIGURE 4.** Two sample images of synthesized and two real image sequences. (a) and (d) represent the two-frame images of the simulated step sequence. (b) and (e) represent the two-frame images of the volume block sample sequence. (c) and (f) represent the two-frame images of the wafer image sequence.

measurement image sequence, which consists of a combination of two blocks with a small difference in heights. The combination of two different height blocks can be thought of as two different height planar surfaces. Due to the chamfer of the gauge block at the boundary, there is a black gap area where the two gauge blocks are pieced together; however, it does not affect our subsequent height measurement and other operations. The image sequence has a lateral acquisition range of  $3000\ \mu\text{m} \times 3200\ \mu\text{m}$ , a height range of  $500\ \mu\text{m}$ , and a step size of  $4\ \mu\text{m}$ , so that the number of fields divided by the lateral range is 9 rows and 7 columns, respectively; 126 frames are captured for each field. Since the range of a single field is much smaller than the acquisition range, the stitched large-area image sequence is shown here.

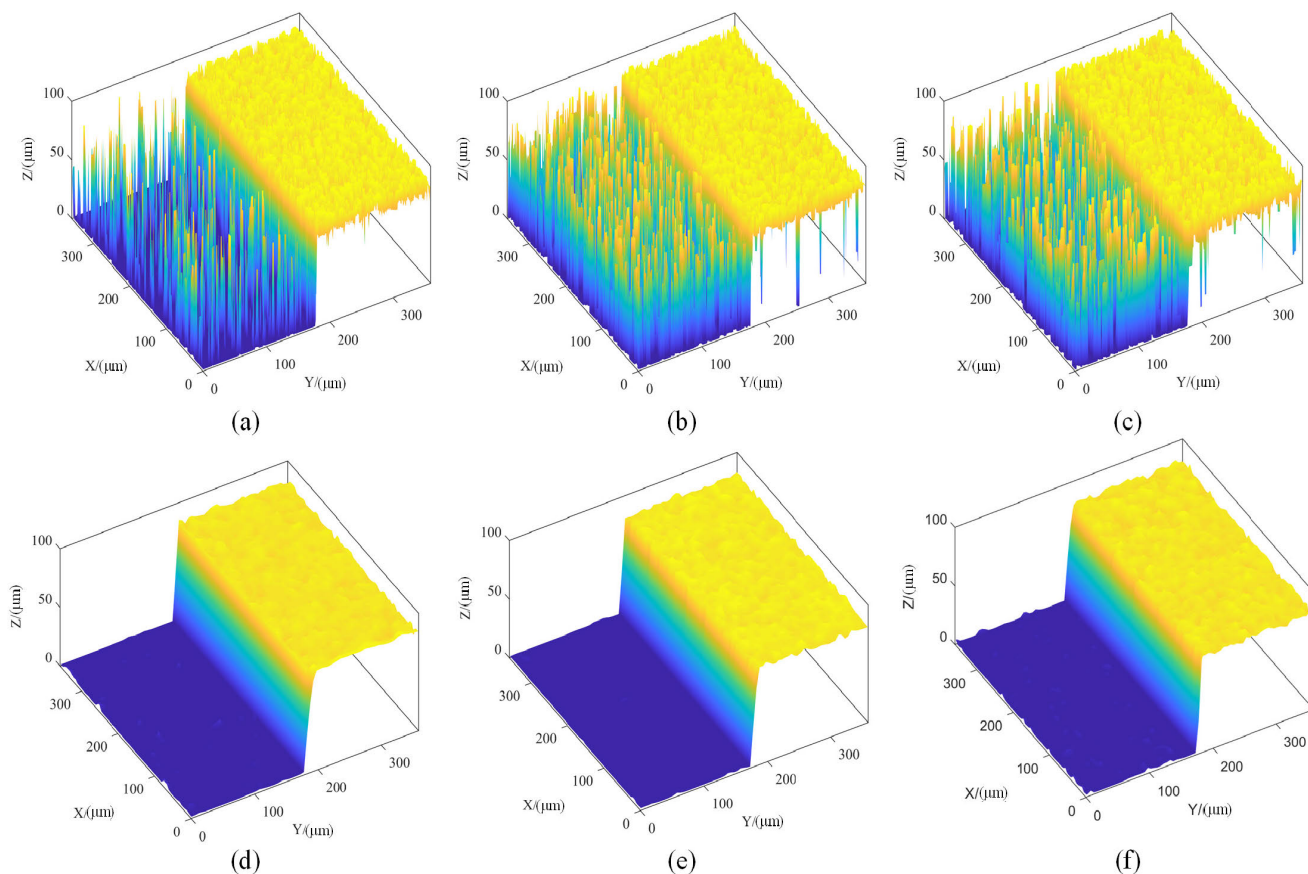
Fig. 4(c) and 4(f) are two frames of the wafer image sequence, which are also displayed in full view using the concatenated image sequence. The collection range of the wafer image sequence is  $6000\ \mu\text{m} \times 6000\ \mu\text{m}$ . The collection range is much larger than the field and cannot cover the overall object shape. Therefore, the image collection and processing functions of large field and microsmall-height objects are achieved through the multifield stitching function. The height range is  $400\ \mu\text{m}$ , the step size is set to  $4\ \mu\text{m}$ , and the number of fields divided in the X- and Y-directions for this image sequence is 16 and 13, respectively, with 101 image frames captured for each field. In the sample, we perform a gel sticking coating on one of the wafer particles for the differentiated display of the defect location.

The proposed method was applied to the SML, GLV, and TEN methods in experiments and compared with them to investigate the improvement performance. The effectiveness of the proposed method was evaluated through qualitative and quantitative analysis. In terms of quantitative evaluation, the accuracy and robustness of the proposed method were verified. The accuracy experiment quantitatively analyzed the comparison results by comparing the height difference between the known gauge blocks and the reconstructed measured gauge blocks. For synthetic objects, we used root mean square error (RMSE) and correlation (Corr) to quantitatively evaluate the results. Its definition is:

$$RMSE = \sqrt{\frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y |d(x, y) - \check{d}(x, y)|^2} \quad (17)$$

$$Corr = \frac{\sum_{x=1}^X \sum_{y=1}^Y (d(x, y) - \bar{d})(\check{d}(x, y) - \check{\bar{d}})}{\sqrt{\left(\sum_{x=1}^X \sum_{y=1}^Y (d(x, y) - \bar{d})^2\right) * \left(\sum_{x=1}^X \sum_{y=1}^Y (\check{d}(x, y) - \check{\bar{d}})^2\right)}} \quad (18)$$

where X and Y denote the depth map dimensions,  $d(x, y)$  denotes the estimated depth map, and  $\check{d}(x, y)$  denotes the true depth map. The mean values of the estimated depth map and the true depth map are denoted using  $\bar{d}$  and  $\check{\bar{d}}$ , respectively. The smaller the RMSE value is, the better the reconstruction



**FIGURE 5.** Depth maps computed using various methods for the step samples. (a)-(c) Depth maps obtained using the GLV, SML, and TEN methods. (d)-(f) Depth maps obtained by applying the method proposed in this paper on top of the GLV, SML, and TEN methods.

result; the closer the Corr value is to 1, the higher the depth estimation accuracy.

Due to the absence of true depth maps (GTs) in real object reconstruction, commonly used evaluation indicators such as RMSE, correlation (Coor), and peak signal-to-noise ratio (PSNR) cannot be used for the reconstruction of real objects. Although the above metrics cannot be used for image sequences of real objects, other metrics, such as smoothness (SSM), can be used [30]. Smoothness defines how smooth the reconstructed surface is and compares the smoothness of surfaces using the number of reconstructed surface paradigms, with a lower SSM value indicating a smoother reconstructed surface.

SSM

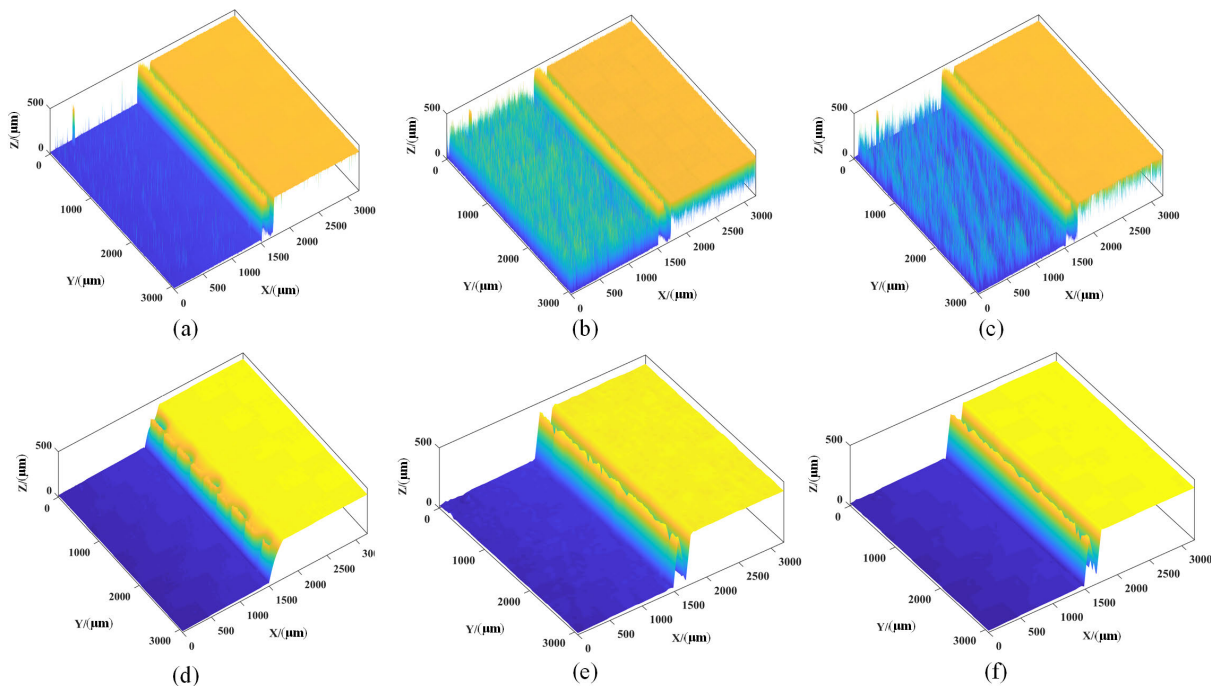
$$= E \left[ \sum_x \sum_y \left\{ \begin{aligned} &\left( \frac{d(x+1, y) - 2d(x, y) + d(x-1, y)}{d(x, y)} \right)^2 + \\ &2 \left( \frac{d(x+1, y+1) - d(x, y+1)}{-d(x+1, y) + d(x, y)} \right)^2 + \\ &\left( \frac{d(x, y+1) - 2d(x, y) + d(x, y-1)}{+d(x, y-1)} \right)^2 \end{aligned} \right\} \right] \quad (19)$$

### B. QUALITATIVE INORGANIC ANALYSIS

In this section, we visually compare the performance of our method with the SML, GLV and TEN methods by analyzing their reconstruction results using a set of simulated synthetic image sequences and two real image sequences. Fig. 5 shows the 3D reconstruction results of the simulated step image, where Fig. 5(a)-(c) shows the reconstruction results of the three methods, and Fig. 5(d)-(f) shows the reconstruction results obtained by using the proposed method based on the three methods. The reconstruction results obtained by the three methods have large surface noise and the surface shape is seriously affected. However, good reconstruction results can be obtained by utilizing the proposed method. The reconstruction results obtained by applying the proposed method using the GLV method have a smoother surface, and the shape reconstruction has a better effect.

In Figs. 6 and 7, the proposed method was compared and analyzed with the GLV, SML, and TEN methods on gauge blocks and wafer objects, respectively. The gauge block sample is formed by docking two standard gauge blocks of different heights. Due to the presence of chamfers on the edges of each gauge block, the chamfers cannot be clearly reconstructed. By observing the reconstruction results





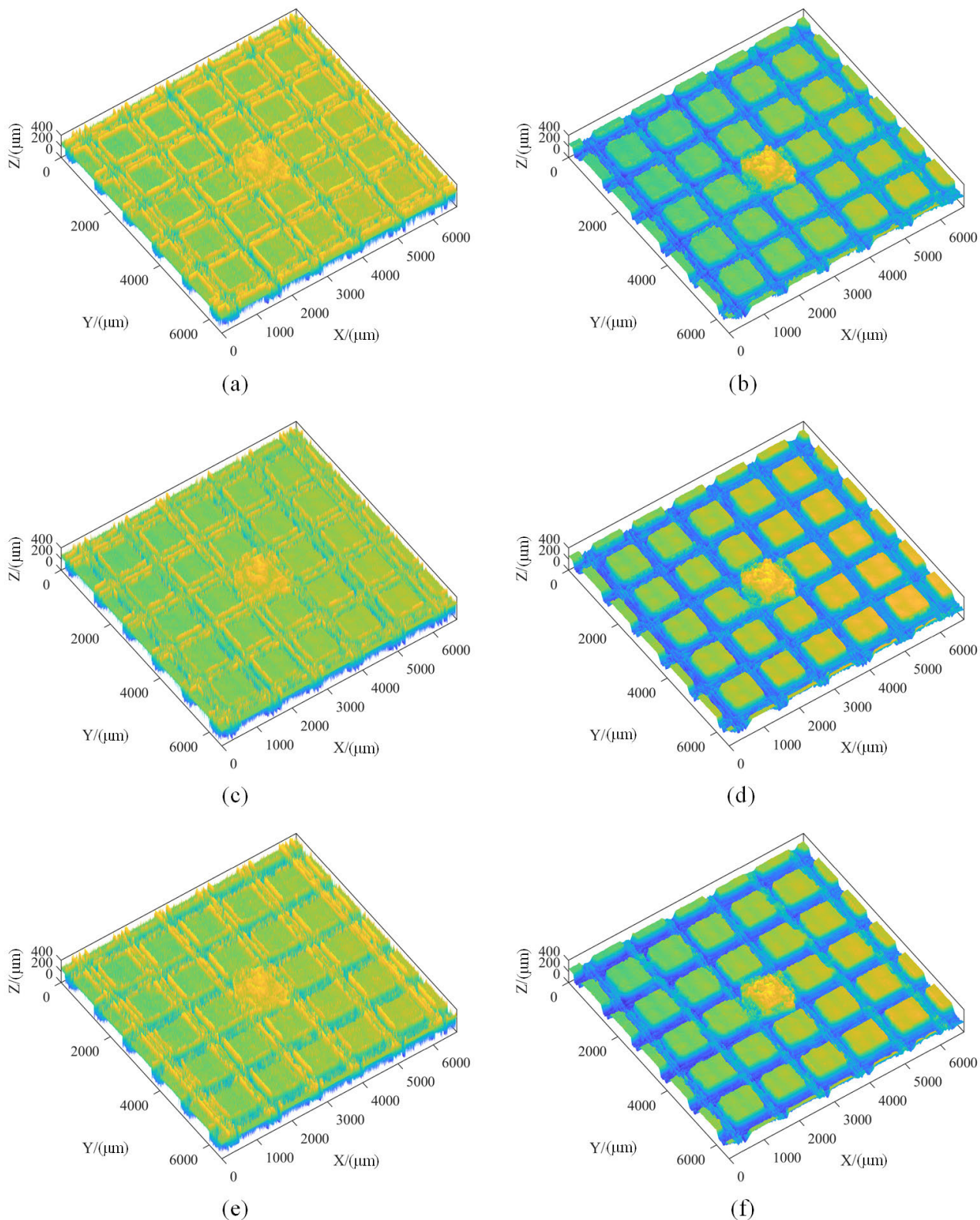
**FIGURE 6.** Depth maps computed using various methods for the volume block samples. (a)-(c) Depth maps obtained using the GLV, SML, and TEN methods. (d)-(f) Depth maps obtained by applying the method proposed in this paper on top of the GLV, SML, and TEN methods.

**TABLE 1.** Selected images in the partial field of the wafer sample image sequence.

Field A					
	Frame 1	Frame 11	Frame 21	Frame 31	Frame 41
	Frame 51	Frame 61	Frame 71	Frame 81	Frame 91
	Field B				
Field B	Frame 1	Frame 11	Frame 21	Frame 31	Frame 41
Field B					
Field B	Frame 51	Frame 61	Frame 71	Frame 81	Frame 91

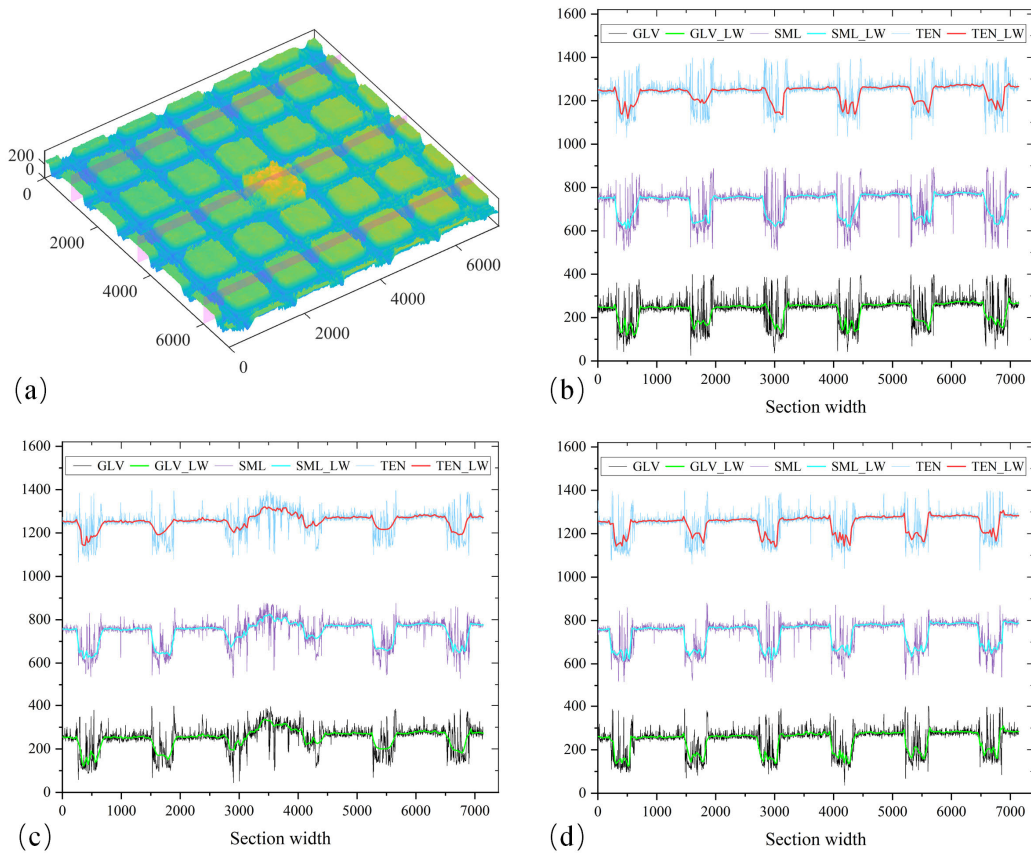
of the three methods, a large amount of noise was found on both sides of the surface at the junction of the gauge

block, especially in the chamfered part. None of the three methods achieved good reconstruction results, and the noise



**FIGURE 7.** Depth maps computed using various methods on wafer samples. (a)(c)(e) Depth maps obtained using the GLV, SML, and TEN methods, respectively. (b)(d)(f) Depth maps obtained by applying the method proposed in this paper on top of the GLV, SML, and TEN methods, respectively.





**FIGURE 8. Cross-section profile analysis of wafer samples reconstructed by different methods at different locations. (a) Schematic of the sample cross-section locations. (b) Cross-section curve analysis at Row 1,000. (c) Cross-section curve analysis at line 3,500. (d) Cross-section curve analysis at Row 6,000.**

completely covered the original shape. However, the proposed method was applied to three methods, achieving good reconstruction results on both sides of the gauge block boundary and having a good filtering effect on noise in the edge area, reconstructing the overall shape and structure of the object.

A grain of dies exists on the wafer and is regularly distributed; there are grooves between the dies, and the internal surface reflection coefficients of the recessed grooves are high; however, the surface of the die weakly reflects the light. The proposed method applies wafer sample defect detection, and some defects are artificially added to the wafer samples, as shown in Fig. 7, where a gelatinous substance is attached to one of the wafers.

Table 1 shows partial images from two fields in the wafer sample image sequence (see Supplement 1 for all images). Field A shows the wafer edges and concave grooves in the wafer sample. Analysis shows that the wafer surface has weak reflection, while the inner reflection ability of the concave grooves is strong. Additionally, there are no textured black edges on the wafer edges. Field B displays a partial image of the simulated defect location, indicating that clear images are generally distributed in a few frames.

The reconstruction results show that although the location of the colloid can also be seen in the reconstruction results of the three methods, there is a large amount of noise at the location of the trench between the dies, and the periphery of the die is caused by black edges, resulting in the no-texture phenomenon. The three methods produce dense noise at these locations. The GLV method exhibits significant depth errors in the region around the wafer, and the TEN and SML methods have the same problem but with less noise. The proposed method has a good filtering effect on this noise, better reconstruction results are obtained under all three methods, the edges of the structure are preserved, the defect regions are well reconstructed and the defect locations are clearly visible using the proposed method.

Fig. 8 shows the profile analysis of the wafer sample reconstruction results at different locations by different methods, taking the 1,000th, 3,500th, and 6,000th rows. Fig. 8(a) shows the specific locations of the sample profiles, and Fig. 8(b), (c) and (d) show the depth profile analysis results at the three locations. GLV, SML, and TEN used in the figure are the three classical methods, and GLV\_LW, SML\_LW, and TEN\_LW are the applications of the method proposed in this paper on the basis of the three methods. The results show

**TABLE 2.** Quantitative analysis of simulated step reconstruction results by different methods.

Result	GLV	GLV_LW	SML	SML_LW	TEN	TEN_LW
Corr	0.973	0.9844	0.9213	0.9863	0.9181	0.9869
RMSE	11.079	9.4675	17.3766	10.2835	17.7305	10.5159
PSNR	18.098	20.4295	14.1885	19.4275	14.0134	19.3713

**TABLE 3.** Results of SSM analysis of real object reconstruction by different methods.

Result	GLV	GLV_LW	SML	SML_LW	TEN	TEN_LW
Gauge blocks	2.7236	0.00011496	2.7468	0.00013596	2.0579	0.000086162
Wafers	703.7733	0.0570	1167	0.0619	770.0906	0.0405

**TABLE 4.** Evaluation of the accuracy of volume block sequences by different methods.

Result	GLV	GLV_LW	SML	SML_LW	TEN	TEN_LW
Heights	401.05	400.09	397.63	400.12	398.24	399.72
Errors	1.17	0.21	2.25	0.24	1.64	0.16
Flatness	3.69	1.45	6.71	2.06	5.14	1.42

**TABLE 5.** Summary of the time used by different methods for volume block sequences.

Result	GLV	GLV_LW	SML	SML_LW	TEN	TEN_LW
Acquisition time (s)	618	618	618	618	618	618
Single-frame processing Time (ms)	38	47	38	48	38	48
Splicing time (s)	16.52	17.89	16.72	16.82	16.36	17.51

**TABLE 6.** Summary of time spent on wafer sequencing by different methods.

Result	GLV	GLV_LW	SML	SML_LW	TEN	TEN_LW
Acquisition time (s)	1868	1868	1868	1868	1868	1868
Single-frame processing Time (ms)	40	53	40	52	39	52
Splicing time (s)	61.67	59.25	61.41	59.57	60.89	59.99

that all three methods have large noise fluctuations; however, after applying the proposed method, smooth cross-section depth curves are obtained. The cross-section depth curves at the 3,500 row position show that all three methods produce obvious noise fluctuations at the defect position; the proposed method can accurately reconstruct the morphology at the defect position, reflecting the accuracy of the proposed method (see Supplement 3 for detailed data).

### C. QUANTITATIVE ANALYSIS

In this section, the proposed algorithm is quantitatively analyzed and compared with three classical methods. Table 2 shows the quantitative analysis of the simulated step image reconstruction results applying RMSE, correlation (Coor), and PSNR by different methods. The proposed method has the smallest RMSE and the largest Coor and PSNR compared

to the three methods: GLV, SML, and TEN. After applying the proposed method to the TEN method, the correlation and the PSNR improved by 7.5% and 38.2%, respectively, and the RMSE was reduced by 40.7%, indicating that the method is very robust to noise.

Since real object reconstruction does not have a true depth map, we use the smoothness (SSM) metric for evaluation. Table 3 summarizes the SSM results of volume block sample and wafer sample reconstruction; the SSM values computed by the proposed method are the smallest among all the methods and are much smaller than those computed by the three methods: GLV, SML, and TEN.

For reconstruction accuracy assessment of the proposed method, we adopt the method proposed by Wang et al. to calculate the height difference of the block samples [31]. First, we select an  $800 \times 800$  area in each left and right region of

the block and then compare the calculated results with the standard results. The standard results were calculated from the test report issued by a professional testing organization; the height difference of the selected standard measuring block was  $399.88 \mu\text{m}$ . The height difference calculated by each method is shown in Table 4. The SML method has the largest deviation in the height calculation, which reaches  $2.25 \mu\text{m}$ ; the TEN and GLV methods are also above  $1 \mu\text{m}$ , while the proposed method applied to the three methods calculates the results close to the true height of the volume block. The maximum error relative to the true height is only  $0.24 \mu\text{m}$ . We calculate the flatness index of each method, in which the proposed method is smaller than the three classical methods, indicating that the proposed method is significantly better than the comparison methods and performs reconstruction performance well. This indicates that the reconstruction accuracy of the proposed method is significantly better than that of the comparison methods and performs reconstruction well.

Tables 5 and 6 summarize the time used for the volume block samples and wafer samples applying different methods, including the image sequence acquisition time, the single-frame image processing time, and the time used for the stitching operation. The analysis in the table shows that all the methods were tested using the same computer in the same environment to ensure calculation accuracy. The same sample sequences were processed by different methods. To ensure the computational speed, we use chunk processing and parallel computing for GLV, SML, TEN and the proposed method. Table 6 shows that the largest single-frame image processing time of the proposed methods is the GLV\_LW method, which has a processing time of 53 ms; the processing time of the GLV method is only 40 ms. Although we introduced computational volume and increased the computational cost, our method optimizes well in the reconstruction results and obtains good reconstruction quality with a small computational cost. In addition, the proposed method uses a shorter time to complete the depth data splicing operation with the all-in-focus image to obtain a larger morphological reconstruction amplitude and complete sample defect detection.

## VI. CONCLUSION

In this paper, a wafer surface topography recovery framework based on the principle of shape from focus is proposed, including large-area fast image acquisition, 3D surface shape recovery, and postdata processing. This surface reconstruction framework is versatile, and the acquired 3D shape can be used for various types of defect detection in the latter stage. The main contributions of the proposed method in this paper are summarized as follows:

- Traditional shape-from-focus methods are mostly based on single-field reconstruction, and the reconstruction range is limited. In this paper, we propose an early method to realize fast reconstruction of large-format wafer surfaces, which is very practical.
- This paper provides a low-cost and high-performance frame structure that can build a system using only an ordi-

nary optical microscope and motion control and can realize high-precision motion control utilizing the scale output pulse to control the acquisition of image sequences.

- A depth map denoising model is proposed that utilizes the principle of Levy flight to achieve noise filtering to improve the depth map and increase reconstruction accuracy.
- Tests were conducted on synthetic and real object images to compare and analyze the reconstruction results with SML, GLV, and TEN methods quantitatively and qualitatively and to compare and analyze the acquisition time, single-frame image processing time, and splicing time of real object image sequences.
- The experimental results show that the reconstruction results of the proposed method are better than those of the comparison methods; for example, for synthetic images, the proposed method has the smallest RMSE and the largest correlation and PSNRs. For real objects, its SSM is much smaller than that of the comparison method.

Although our proposed method performs well in processing wafer samples, the proposed method still has some limitations, as described below.

- Our proposed method involves iterative operations in the computational process with a higher computational cost compared to methods such as GLV, SML, and TEN. Although we used techniques such as CPU-based parallel computing to reduce its time cost, further computational efficiency enhancement is needed.
- For different samples, it is necessary to adjust and optimize some internal parameters, such as the step size of Levy flight noise reduction and the number of iterations, to obtain the best reconstruction effect. Therefore, adaptively adjusting the parameters of the noise reduction method will further improve the reconstruction accuracy and achieve a better noise reduction effect.

## REFERENCES

- [1] J. Y. Huang, Z. J. Guan, and W. M. Li, "A neural network-based method for rapid measurement of silicon wafer surface topography," *J. Funct. Mater. Devices*, vol. 29, no. 2, pp. 117–123, 2023.
- [2] T. L. Schmitz, A. Davies, C. J. Evans, and R. E. Parks, "Silicon wafer thickness variation measurements using the National Institute of Standards and Technology infrared interferometer," *Opt. Eng.*, vol. 42, no. 8, pp. 2281–2290, 2003.
- [3] S. Q. Liu, W. B. Yu, X. M. Zhao, L. Tao, and C. Lu, "Characterization of surface morphology of piezoelectric wafer material based on lift-off technique," *Piezoelectric Acousto-optics*, vol. 37, no. 3, pp. 512–518, 2015.
- [4] S.-A. Lee, H.-S. Jang, and B.-G. Lee, "Jitter elimination in shape recovery by using adaptive neural network filter," *Sensors*, vol. 19, no. 11, p. 2566, Jun. 2019.
- [5] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.*, vol. 46, no. 5, pp. 1415–1432, May 2013.
- [6] S. Salokhiddinov and S. Lee, "Iterative refinement of uniformly focused image set for accurate depth from focus," *Appl. Sci.*, vol. 10, no. 23, p. 8522, Nov. 2020.
- [7] M. T. Mahmood and I. H. Lee, "Shape from focus based on 3D structure tensor using optical microscopy," *Microsc. Res. Technique*, vol. 83, no. 1, pp. 48–55, Jan. 2020.
- [8] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

- [9] M. Dalibor and M. Karel, "Mathematical principles of object 3D reconstruction by shape-from-focus methods," *Mathematics*, vol. 9, no. 2253, pp. 1–20, 2021.
- [10] M. T. Mahmood, Y.-K. Choi, and S.-O. Shim, "Estimating shape from focus by Gaussian process regression," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2012, pp. 1345–1350.
- [11] H.-J. Kim, M. Mahmood, and T.-S. Choi, "An efficient neural network for shape from focus with weight passing method," *Appl. Sci.*, vol. 8, no. 9, p. 1648, Sep. 2018.
- [12] M. Moeller, M. Benning, C. Schönlieb, and D. Cremers, "Variational depth from focus reconstruction," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5369–5378, Dec. 2015.
- [13] U. Ali and M. T. Mahmood, "Combining depth maps through 3D weighted least squares in shape from focus," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Jan. 2019, pp. 1–4.
- [14] S. Salokhiddinov and S. Lee, "Deep spatial-focal network for depth from focus," *J. Imag. Sci. Technol.*, vol. 65, no. 4, Jul. 2021, Art. no. 040501.
- [15] G. Tan and Z. Wang, "Reachable set estimation of delayed Markovian jump neural networks based on an improved reciprocally convex inequality," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2737–2742, Jun. 2022.
- [16] J. Hu, G. Tan, and L. Liu, "A new result on  $H_\infty$  state estimation for delayed neural networks based on an extended reciprocally convex inequality," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, 2024.
- [17] W. Wu, X. Guo, Y. Chen, S. Wang, and J. Chen, "Deep embedding-attention-refinement for sparse-view CT reconstruction," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.
- [18] W. Wu, D. Hu, C. Niu, H. Yu, V. Vardhanabhuti, and G. Wang, "DRONE: Dual-domain residual-based optimization network for sparse-view CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 11, pp. 3002–3014, Nov. 2021.
- [19] J. M. Tenenbaum, "Accommodation in computer vision," M.S. thesis, Stanford Univ., Stanford, CA, USA, 1970.
- [20] F. S. Helmlí and S. Scherer, "Adaptive shape from focus with an error estimation in light microscopy," in *ISPA. Proc. 2nd Int. Symp. Image Signal Process. Analysis. Conjoint 23rd Int. Conf. Inf. Technol. Interface*, 2001, pp. 188–193.
- [21] J. Zheng and T. Zhang, "Wafer surface defect detection based on background subtraction and faster R-CNN," *Micromachines*, vol. 14, no. 5, p. 905, Apr. 2023.
- [22] L.-C. Chen, D.-H. Duong, and C.-S. Chen, "Optical 3-D profilometry for measuring semiconductor wafer surfaces with extremely variant reflectivities," *Appl. Sci.*, vol. 9, no. 10, p. 2060, May 2019.
- [23] S. Suwajanakorn, C. Hernandez, and S. M. Seitz, "Depth from focus with your mobile phone," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3497–3506.
- [24] S.-O. Shim, W. Aziz, A. Banjar, A. Alamri, and M. Alqarni, "Improving depth computation from robust focus approximation," *IEEE Access*, vol. 7, pp. 20144–20149, 2019.
- [25] U. Ali and M. T. Mahmood, "Energy minimization for image focus volume in shape from focus," *Pattern Recognit.*, vol. 126, Jun. 2022, Art. no. 108559.
- [26] H. Mutahira, V. Shin, M. S. Muhammad, and D. R. Shin, "Sampling-noise modeling & removal in shape from focus systems through Kalman filter," *IEEE Access*, vol. 9, pp. 102520–102541, 2021.
- [27] M. Muhammad and T.-S. Choi, "Sampling for shape from focus in optical microscopy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 564–573, Mar. 2012.
- [28] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 210–214.
- [29] E. H. Houssein, M. R. Saad, F. A. Hashim, H. Shaban, and M. Hassaballah, "Lévy flight distribution: A new metaheuristic algorithm for solving engineering optimization problems," *Eng. Appl. Artif. Intell.*, vol. 94, Sep. 2020, Art. no. 103731.
- [30] I.-H. Lee, M. Mahmood, and T.-S. Choi, "Robust depth estimation and image fusion based on optimal area selection," *Sensors*, vol. 13, no. 9, pp. 11636–11652, Sep. 2013.
- [31] C. Wang, X. Meng, G. Krolczyk, and H. Wei, "A cluster-based method for quantitative characterization and uncertainty evaluation of areal step height in structured surface metrology," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.



**SHAOHANG WANG** is currently pursuing the B.E. degree majoring in electronic and information science with the School of Mechanical, Electrical and Information Engineering, Shandong University. His current research interests include micro-manipulation robots, microscopic stereovision technology, image processing, and information technology.



**JIQIANG CHEN** received the master's degree in mechanical engineering from the Beijing University of Technology, in 2022, where he is currently pursuing the Ph.D. degree in mechanical engineering. His research interests include computer image processing and 3-D reconstruction of microscopic scenes.



**HUA SHI** is currently pursuing the B.E. degree majoring in intelligent manufacturing engineering with the College of Mechanical Engineering and Applied Electronics Technology, Beijing University of Technology. His current research interests include micro-manipulation robots, computer vision, and machine vision.



**MINGYE LI** received the B.E. degree majoring in industrial design from the College of Engineering, China Agricultural University. Her current research interests include image processing and information technology.

...