> Ms. No. TIM-24-03452 <
The following publication S. Gao, D. Li and C. Fai Cheung, "Multiframe Resolution-Enhanced Autostereoscopic System for On-Machine 3-D Surface Metrology," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-9, 2024, Art no. 5037909 is available at https://doi.org/10.1109/TIM.2024.3472890.

Multi-frame Resolution-enhanced Autostereoscopic System for On-machine Three-dimensional Surface Metrology

Sanshan Gao, *Da Li, and *Chi Fai Cheung, *Member, IEEE*

***Abstract*—This paper presents a multi-frame resolution-enhanced autostereoscopic system for the on-machine measurement of three-dimensional surfaces. It takes advantage of the vibration from the machine tool during the on-machine measurement process to acquire multiple frames of the target surface with offsets, thereby achieving resolution enhancement. A multi-frame resolution-enhanced deep-learning model is developed to generate resolution-enhanced raw elemental images which significantly improve the measurement resolution of the system. The performance of the system is evaluated by experiments and the results show that the spatial resolution of the measurement data is enhanced by four times with improved measurement accuracy.**

***Index Terms*—surface metrology, autostereoscopy, deep learning.**

I. INTRODUCTION

THE use of three-dimensional surfaces in the development of products to realize specifically designed optical and mechanical functions has become more widespread. Applications can be found in various industries such as biomedical [1], optics [2], aerospace [3], energy, etc. The increasing geometrical complexity of 3D surfaces creates considerable challenges in regard to their measurement, particularly for on-machine measurement. Surface dimensional measurement techniques encompass contact methods and noncontact methods. Contact profilometers utilize a stylus [4] that traverses the surface in vertical and lateral motions, thereby capturing the distance between points and the contact force exerted, ensuring accurate and precise measurements. To enhance resolution, especially for micro-structured surfaces, finer tips have been developed [5]. Zhang et al. [6] presented a smart sampling strategy for a touch-trigger on-machine probe to enhance the identification capability regarding defect areas. The strategy determines an optimum distribution of sample points and incorporates supplementary sampling in regions with defects. Zhu et al. [7] incorporated a tracking head based on scanning tunneling microscopy, a high-aspect ratio probe, and multi-axis moving stages in order to enhance the accuracy of on-machine measurements. Although various high-precision on-machine measurement systems have been proposed, their performance is influenced by the machine kinematic errors and susceptible to machine vibration [8]. Nonetheless, contact profilometers are limited by their time-consuming data acquisition process and the potential for surface damage due to the contact nature of the probe, particularly on soft or delicate materials. In contrast, non-contact profilometers, which are predominantly optical-based, have been the focus of extensive research, employing techniques such as interferometry [9], deflectometry [10], structured light [11], confocal microscopy [12], etc. The advantage of non-contact profilometers lies in their ability to measure surfaces without exerting additional force or causing any effects on the surfaces. Noncontact methods are more flexible to implement and require less time consumption and system complexity, especially for small measured parts with microstructures.

Autostereoscopic 3D surface metrology based on the light field theory is an emerging noncontact surface detection technology that utilizes a micro-lens (MLA) array that is incorporated into a single-lens imaging system to capture the raw 3D information of the measured surface in a single snapshot. This enables faster data acquisition for the on-machine measurement and real-time inspection. A real-time inspection using light field microscopy was achieved. This was made by transforming light-fields back into depth information, with the assistance of a view-channel-depth machine learning network [13]. A semantic segmentation light-field system was developed, by utilizing an end-to-end convolutional neural network [14] which efficiently extracts angular-spatial features from light field data. The system's ability to perform efficient semantic segmentation enhances the potential applications of light field technologies.

Manuscript received xx xx xxxx; revised xx xx xxxx; accepted xx xx xxxx. Date of publication xx xx xxxx; date of current version xx xx xxxx. The work described in this paper was mainly supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project No. R5047-22). The first two authors contributed equally to this work. (Corresponding author: Chi Fai Cheung and Da Li)

Sanshan Gao is with the State Key Laboratory of Ultra-Precision Machining Technology, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China (e-mail: sanshan.gao@connect.polyu.hk)

Da Li is with the State Key Laboratory of Ultra-Precision Machining Technology, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China, and also with the Institute of Modern Optics, Nankai University, Tianjin 300071, China (e-mail: da.li@nankai.edu.cn).

Chi Fai Cheung is with the State Key Laboratory of Ultra-Precision Machining Technology, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China (e-mail: benny.cheung@polyu.edu.hk).).

To improve the autostereoscopic measurement performance, a system-associated direct extraction of disparity information (DEDI) method [15] is used for 3D surface reconstruction, where disparity patterns based on the parallax information of the recorded data were presented to enhance the accuracy of depth estimation, providing a turnkey solution for the on-machine microstructure measurement of 3D surfaces. A calibration model, grounded in epipolar space theory, was presented [16] to ascertain the relationship between measured points and epipolar-space parameters for depth reconstruction. A method based on the point spread function [17] was developed to ascertain the spatial location of the MLA by analyzing the overlap of elemental images, thereby achieving more precise digital refocusing to enhance measurement accuracy. To augment the accuracy of disparity retrieval and improve the quality of surface reconstruction, a learning-based method [18] was proposed to achieve adaptive focus volume aggregation from digitally refocused data.

However, the resolution of this measurement system has been limited due to the division of the image sensor's pixels into multiple small areas by the numerous small apertures of the MLA. Additionally, these segmented small areas of pixels need to undergo a matching process and screening process before the final 3D point cloud of the target surface can be generated. In other words, the final resolution of the measurement system is directly influenced by the resolution of each segmented small area. The segmentation caused by the MLA undoubtedly has an adverse impact on the final resolution of the measurement system. Research aimed at enhancing the angular resolution of autostereoscopic data has been explored in [19, 20], including resolution enhancement methods for plenoptic cameras and a semi-supervised learning paradigm to improve data efficiency. Due to the superiority of deep learning in image processing, various learning models for resolution enhancement of blurry, noisy, and low-quality images have emerged. These models are implemented using data-driven, model-based, or unsupervised approaches. However, these methods may tend to produce finer artifacts to visually appeal to human perception, which could result in inaccurate details for precise measurements. As a result, the development of an effective method for accurate spatial-resolution recovery of autostereoscopic data is essential to further improve measurement performance.

This study introduces a multi-frame resolution-enhanced autostereoscopic system designed to enhance the resolution of measurement systems for on-machine measurement of 3D surfaces. The system capitalizes on the inevitable vibrations of machine tools, which can induce local blurring in images captured during the on-machine process, by utilizing these subtle displacements to acquire multiple frames of the target surface with slight offsets within a brief timeframe. After the enhancement process, the measurement data obtained are of high resolution, featuring clear and sharp details. This improvement enhances the precision and accuracy of the autostereoscopic measurements. The multi-frame resolution enhancement is achieved by utilizing the sub-pixel information contained in different frames with slight displacement, along with a deep learning-based resolution-enhanced network and training process. The processed resolution-enhanced image can reconstruct the 3D surface with significant improvements in both lateral and axial resolution. The performance of the proposed method and system is evaluated through experiments conducted on a sample with a micro-structured surface. A learning-based unfolding super-resolution network (USRNet) [21] is employed as a comparison benchmark in the experiment, owing to its effective enhancement of low-resolution images and its commendable generalizability. The proposed method is found to be able to enhance spatial resolution and improve measurement accuracy.

II. Multi-frame Resolution-enhanced Autostereoscopic Measurement

Fig. 1 is a schematic diagram of the system of multi-frame resolution-enhanced autostereoscopy for on-machine 3D surface measurement, including the recording and reconstruction processes. Different spatial locations of the elemental lenses in an MLA cause small differences in viewing angles in the elemental images received on the image sensor (known as disparities). The disparity information can be used to calculate the 3D information about the target surface and this is the reconstruction process. A quantitative expression of a specific point's disparity is determined based on the parameters of the system setup including the pixel size of the image sensor, the pitch of the MLA, the gap (the distance between the MLA and the image sensor), and the dimensional variation along the depth direction.

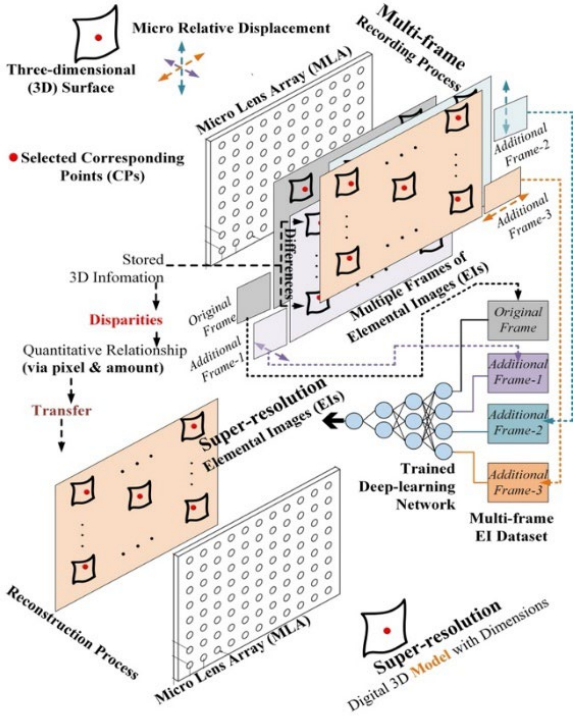

Fig. 1. Working principle of multi-frame resolution-enhanced autostereoscopic metrology for on-machine 3D surface measurement.

This quantitative disparity information, both in the lateral and depth directions, is transferred from the recording process to the reconstruction process. The corresponding points, which are the image points from different EIs, originate from an identical single object point in the object space (red points in Fig. 1) and satisfy the quantitative relationship between disparity information and system parameters. Corresponding points need to be accurately chosen based on the match of pixel information and its neighbourhood in EIs and form the 3D digital reconstruction at corresponding spatial locations. The reconstructed and object spaces are symmetrical both in the lateral and axial directions according to the reversibility of optical rays.

Since vibration from the machine tool between the target surfaces and the measurement system is unavoidable during the on-machine measurement process, such vibration causes a small movement at the micrometer scale towards the image sensors so that each of the measurement frames records a composition of different optical signals as shown in Fig. 1. Due to the finite sampling rate of the image sensor, the subtle displacement resulting from vibration may cause local blurring in the recorded images. Given that each frame represents a down-sampled result of the recorded scene, points corresponding to the same object point across different frames, along with their neighboring pixels, exhibit distinct grayscale values and pixel gradients. These local features and subpixel information can be used to compensate for the low-resolution signals and reconstruct high-resolution data. In this sense, high-resolution (HR) images can be reconstructed by analyzing and processing the different pixel representations of multiple frames.

As illustrated in Fig. 2, two different frames are recorded for the same object but with a slight displacement. The pixel distribution of the two frames, which refers to the pixel values of the target surfaces, differs when the object appears in different positions on the image sensor. After registration and fusion, the redundant pixel information is combined, forming a new pixel distribution in a sub-pixel space. The new pixel distribution is processed in the reconstruction process. A high-resolution frame with sharp edge information and details is generated from multiple frames.

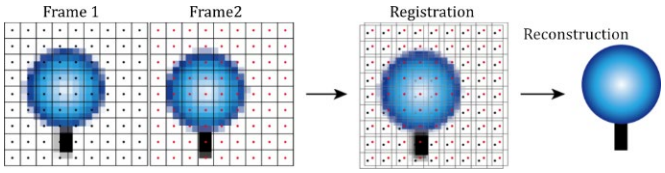

Fig. 2. Multi-frame resolution-enhancement process.

The two key issues to address in the multi-frame resolution-enhancement problem are registration and reconstruction. Conventional methods usually rely on a priori knowledge to extract features from the multiple frames and achieve registration based on these features. Fusion is accomplished through a series of designed kernels based on experiments or expert experiences. However, the measurement images often contain various types of noise, such as Gaussian noise, salt and pepper noise, and smudge noise, caused by factors like illumination, exposure, condition of the lenses, etc. As a result, conventional methods often struggle to extract effective features from the images and reconstruct high-resolution images with high robustness. It is inspiring to utilize deep learning to generate resolution-enhanced EIs by accurately registering and reconstructing high-resolution images. To this end, a multi-frame resolution-enhanced model based on a deep-learning network is developed to super-resolve low-resolution (LR) measurement image stacks into high-resolution images with denoising and clear details.

III. Multi-frame Resolution-enhanced Deep Learning Model

To generate resolution-enhanced EIs based on multiple frames of low-resolution EIs, a deep learning network is developed. A supervised training process is used to generate resolution-enhanced images based on image data collected under various conditions, light intensity, recording device, etc.

A. Model framework

As shown in Fig. 3, the proposed multi-frame resolution-enhanced deep learning model consists of four components: a single-frame resolution-enhanced network, a registration network, an auxiliary-frame resolution-enhanced network, and a series of convolutional layers for post-processing. Fig. 3 shows a schematic diagram of the model. The captured multiple frames are split into a base frame, and auxiliary frames. The output of the high-resolution image has the same geometrical position as the base frame and the auxiliary frames are used to provide redundant sub-pixel information. The base frame is first up-sampled using Bilinear interpolation so that the spatial dimension is increased to a desired value. The up-sampled base frame is then input to the single-frame resolution-enhanced network, where the input frame is converted into a stack of high-dimensional single-frame features through the processing of the convolutional layers and activation functions. These features are used for subsequent processing. All the frames, including the base frame and the auxiliary frames, are input into the registration network for alignment. Although the displacement detection and registration can also be achieved by traditional methods such as Scale Invariant Feature Transform (SIFT) [22], these methods are more sensitive to the noises that are unavoidable during on-machine measurement due to various illumination, vibration, and machining environment. In addition, the registration network can realize an end-to-end training and inference fashion, eliminating the need for extra preprocessing of the raw measurement data.

Displacement between frames may lead to blur and image artifacts during the convolutional operation, as points corresponding to the same object may possess different coordinates across multiple frames. Different views of the same object point may not align within a single convolutional kernel, resulting in image artifacts and blurred regions, as shown in Fig.

4. However, through translating, these points can be effectively aligned within a unified kernel window, leading to the noticeable elimination of image artifacts. Consequently, an affine transformation between these frames is essential to mitigate the effects during convolution. Since these frames only undergo slight displacement, it is assumed that only translation occurs. According to the affine transformation matrix, the registration process is formulated as:

$$\begin{bmatrix} \tilde{x}_{Ai} \\ \tilde{y}_{Ai} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \theta_x \\ 0 & 1 & \theta_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{Ai} \\ y_{Ai} \\ 1 \end{bmatrix} \tag{1}$$

where $I(x_{Ai}, y_{Ai})$ is one of the auxiliary frames and $I(\tilde{x}_{Ai}, \tilde{y}_{Ai})$ is the corresponding registered frame which has been aligned with the base frame. θ_x and θ_y are the translation parameters in the x and y directions, respectively. The output of the registration network is the translation parameters (θ_x, θ_y). The translation distribution predicted by the registration network aligns with the pixel displacement in the training data. Random translations can be introduced into the training data to simulate various vibration amplitudes for different devices.

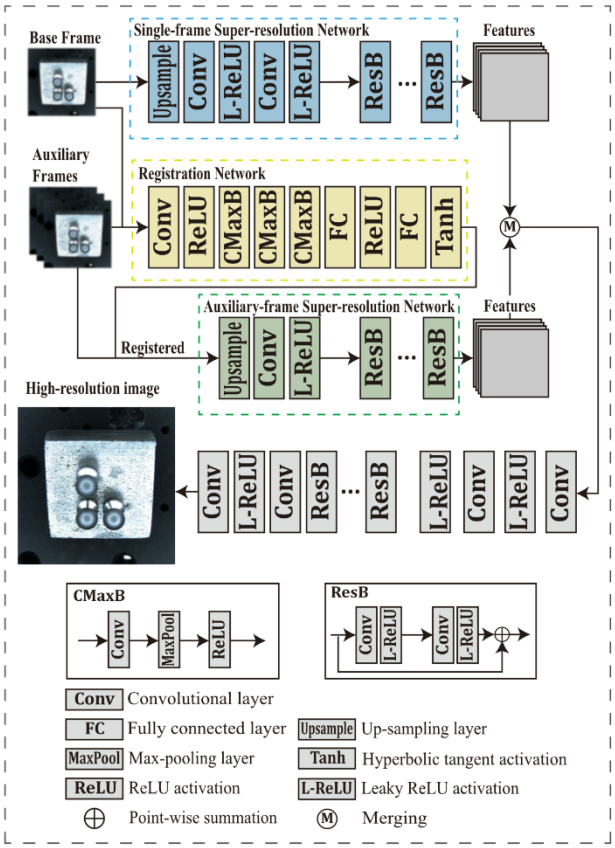

Fig. 3. Multi-frame resolution-enhanced deep-learning model.

To effectively reduce the dimensionality of the input frames while retaining key information, max-pooling filters are used in the registration network. Fully connected layers are followed in the convolutional layers to realize the prediction of the translation parameters. The Tanh (hyperbolic tangent) activation function is used to compress the output translation parameters within the range of [-1, 1]. The output auxiliary frames are aligned with the base frame through registration. Similar to the single-frame super-resolution route mentioned earlier, the aligned auxiliary frames are up-sampled and input into the auxiliary-frame resolution-enhanced network. Consequently, these auxiliary frames are converted into a stack of multi-frame features. These two stacks of features, namely the single-frame features and the multi-frame features, are merged and then input into the post-processing convolutional layers. After the post-processing layers, a high-resolution image is reconstructed. All the aforementioned sub-networks utilize a residual connection architecture (ResB in Fig. 3) to prevent gradient vanishing.

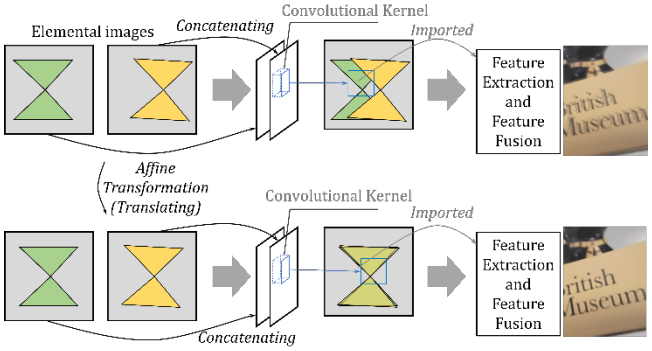

Fig. 4. Impact of image translating to the convolution.

In the proposed convolutional layers, a kernel size of 3 is employed to reduce the number of network weights. Images are downscaled to a 64×64 resolution prior to input into the registration network, ensuring consistent predictions of the translating parameters. Given that the registration network's output is a relative translation value, the reduction in image dimensions does not impact the efficacy of the desired affine transformation. Except for the registration network, which uses Rectified Linear Unit (ReLU) and Tanh activation functions, all the other sub-networks employ Leaky ReLU as their activation functions. Research [23, 24] has shown the superiority of Leaky ReLU for super-resolution applications.

B. Model training

The behaviour of the neural network is invariably influenced by the distributions of the input and ground truth; thus, denoising is facilitated by introducing noise into the raw measurement data, thereby transforming the data distribution into a noisy one. Subsequently, the network is trained to estimate the maximum a posteriori probability for the retrieval of high-resolution information from the data with noise. To this end, a supervised training process is used based on image data collected under various conditions, light intensity, recording device, etc. The training data are first down-sampled as the input, and the raw data are used as ground truth. To achieve a clearer reconstruction, noises are added on the input data to

simulate the noises in realistic environments. Hence, the objective of the proposed multi-frame resolution-enhanced network is not only to recover the high-resolution information but also to achieve denoising.

The loss function is comprised of three parts: reconstruction loss, gradient loss, and perceptual loss. Reconstruction loss measures the errors between the reconstructed high-resolution images and the corresponding ground truth using mean absolute errors as the criterion. To preserve the edge information in the reconstructed images, the gradient loss compares the gradients of the ground truth and the reconstructed images in the horizontal and vertical directions. The resulting errors are used to calculate the gradient loss. The perceptual loss [25] quantifies the feature distance between the ground truth and the reconstructed images. This is done by extracting features using a pre-trained VGG network [26], which is a widely-used trained network developed by the Visual Geometry Group at the University of Oxford. The total loss function is formulated as:

$$L = \sum \left(\left| I^{HR} - \hat{I} \right| + \left| \nabla_x I^{HR} - \nabla_x \hat{I} \right| + \left| \nabla_y I^{HR} - \nabla_y \hat{I} \right| + \left| \phi\left(I^{HR} \right) - \phi\left(\hat{I} \right) \right|^2 \right) \tag{2}$$

where I^{HR} is the high-resolution images reconstructed by the proposed network, $\hat{I}$ is the corresponding ground truth, and $\phi(\cdot)$ denotes the VGG network. In addition, the training data are augmented by rotation, flipping, and random cropping to realize thorough training of the proposed network.

C. Implementation details

In this work, the measurement images are super-resolved and up-scaled 4-fold. The number of total input frames is 4. The single-image resolution-enhanced network and the auxiliary-frame resolution-enhanced network both contained two residual blocks. The post-processing layers contained two residual blocks.

The training data are collected by a 2D imaging system and a Lytro Illum commercial light-field camera. Since a single EI captured by the proposed measurement system has a limited number of pixels, the data obtained by the Lytro Illum camera and the 2D system with higher resolution can provide much richer pixels to achieve more effective training of the resolution-enhanced model that learns the mapping function from the low-resolution images to high-resolution images. Multiple scenes that include various samples such as sphere surfaces, machining parts, bonding wires, and other objects containing complex surfaces are captured under different illumination conditions for the construction of the training dataset. Each scene contains 4 frames with slight displacement. For more efficient learning using the limited measurement data, the up-scale factor is set to 2 during training. The well-trained model is employed to enhance the resolution by 4 times during inference. Data augmentation is conducted by rotating the training data by 45, 90, 135, 180, 225, 270, and 315 degrees, flipping them from left to right or from up to down, and cropping the data into random-size patches.

Due to the autostereoscopic system's constraints in illumination and recording conditions, the raw data utilized as ground truth for network training often encompass noise and blur, which obstruct the network's ability to super-resolve sharper and clearer details. Moreover, acquiring high-quality, high-resolution autostereoscopic data is challenging. In response to this, the training process incorporates various noise types, including Gaussian noise, local mosaic processing, and local blurring, into the recorded data, enabling the network to learn the recovery of finer details from these noised data. This noise is applied randomly across different regions and at varying levels, further augmenting the finite dataset. However, introducing additional noise alters the original distribution of the autostereoscopic data, potentially resulting in the super-resolved data containing extraneous image artifacts. To mitigate this, finetuning is conducted to adjust the network weights on the original dataset without introducing noise. In other words, the finetuning dataset contains a reduced number of training samples, attributable to the omission of augmentation, but maintains the same distribution as the raw autostereoscopic data. The perceptual loss's penalty coefficient is set to 0.01 throughout finetuning, directing the network's focus toward the pixel-level discrepancies between the recovered results and the ground truth to minimize the emergence of unintended artifacts. A comparative analysis presented in Fig. 5 involves the network trained on augmented noisy data without finetuning, the network directly trained on the original dataset, and the finetuned network. Although training directly on the dataset without noise augmentation can yield similar or even higher PSNR and SSIM metrics, the resulting images tend to exhibit blurred edges and may even display a sawtooth effect. The model trained on noisy data without finetuning is prone to generating more artifacts and noise due to the misidentification of noise as salient features. In contrast, the images produced after finetuning possess sharper and clearer edges, preserving the original structure of the surfaces.

The input patch size is set to 128. The model is implemented using PyTorch and trained using NVIDIA RTX 2080 GPUs. The Adam optimizer [27] is used during backpropagation, and the initial learning rate is set to 10^{-4}. The learning rate decays every 10 training epochs.

D. Surface reconstruction

The depth estimation process is based on the direct extraction of disparity information method [15], disparity patterns, and shape from focus via digital refocusing [28], as shown in Fig. 6. Digital refocusing is first performed using the recorded autostereoscopic data, allowing for the acquisition of a stack of refocused images. The corresponding points should be focused at a specific depth plane, which is equivalent to finding focus regions in the refocused image stack. A focus measure operator is used to detect focus points in every refocused image, resulting in the obtainment of a focus volume. After smoothing and denoising the focus volume, a preliminary

depth map is estimated using the winner-take-all strategy. An all-in-focus image can be obtained based on the preliminary estimation. To further refine the estimated depth, guided filtering is applied to the preliminary estimation based on the all-in-focus image. Outlier points caused by incorrect estimation, such as small locally convex or concave regions in the depth map, are further removed using pre-defined thresholds based on the assumption of continuous surfaces. As a result, desired depth maps, point clouds, and the corresponding all-in-focus images can be acquired from the low-resolution (LR) or high-resolution (HR) autostereoscopic measurement data.

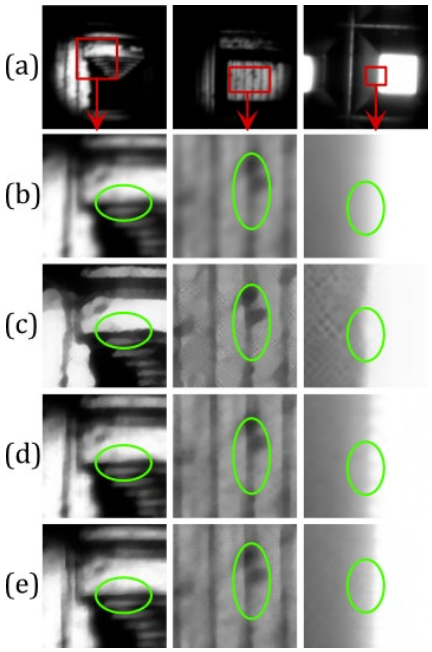

Fig. 5. An ablation study for the evaluation of the finetuning improvement. (a) is the reference scenes super-resolved by 4 folds, and (b) ~ (e) are local enlargements extracted from the results of the Bilinear interpolation, the model trained on the noised dataset without finetuning, the model trained directly on the original dataset, and the model initially trained on the noised dataset then finetuned on the original dataset, respectively.

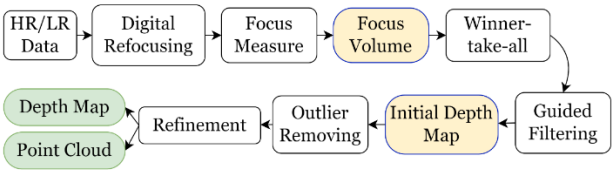

Fig. 6. Framework of the surface reconstruction process from the autostereoscopic measurement data.

IV. Experiments

A. System setup for on-machine measurement

As shown in Fig. 7, a prototype of the multi-frame resolution-enhanced autostereoscopic 3D surface measurement system is built to perform on-machine 3D surface measurement. The whole system is mounted on the motion stage of a Moore Nanotech 350FG ultra-precision machine. The overall field of view (FOV) of the measurement system is 625 μm diagonally based on the offline calibration of the system in terms of the overall magnification of the objective lens and zoom lens, and the actual used image sensor size.

To evaluate the accuracy and resolution of the autostereoscopic system, a series of measurement experiments is conducted on a 3D micro-structured sample. The sample is mounted on the air bearing work spindle of the machine.

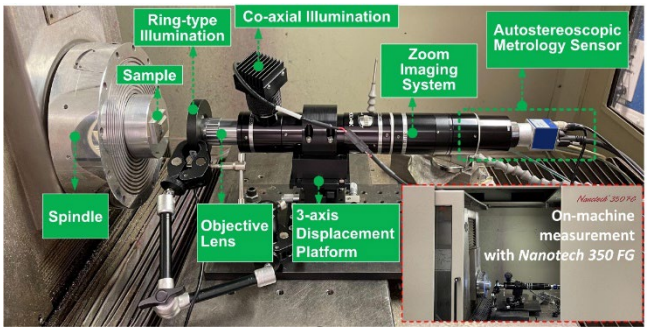

Fig. 7. On-machine measurement through a multi-frame resolution-enhanced autostereoscopic 3D surface measurement system.

Multiple frames of the EIs of the sample with offsets of pixels are captured during the on-machine measurement process. The offsets among the multiple collected frames are analyzed based on pixel values and greyscale projection. The results show that the sub-pixel level offsets between the different frames happened during the on-machine measurement which qualifies for the proposed multi-frame resolution-enhanced method.

To investigate the frame jitter caused by the vibration of the machine tool, the SIFT descriptor is used to determine the pixel-level shifting among multiple continuous frames captured during the measurement. The first frame is used as the reference, and the other frames are matched with the reference frame to detect small displacements. Key points of the reference frame and the other frames are detected and computed using the SIFT descriptor and detector. The matching is achieved using the FLANN (Fast Library for Approximate Nearest Neighbors) method. Since the SIFT detector can achieve detection on a sub-pixel scale, the sub-pixel-level distances between the matched points can be determined, thus identifying the frame jitter. To decrease the effects resulting from inaccurate matching, a total of 500 groups of matched points from the reference frame and the detected frame are used for analysis. To eliminate the effects resulting from noise, the same displacement detection is performed on noised data, which are generated by adding extra Gaussian noise to the reference frame. The average grayscale difference per pixel between the noised data and the reference data is on the same scale as the grayscale difference among the multi-frame data, ensuring the validity of the comparison. The difference is 0.950 per pixel between the noised data and the reference, and 0.931 per pixel between the multiple frames. The

results of the multi-frame data and the noised data are shown in Fig. 8, where a total of 5 frames, excluding the reference frame, are involved. The red lines represent the pixel displacement of the noised data, and the blue lines represent the pixel displacement of the multi-frames. The results demonstrate that sub-pixel displacement occurs during the on-machine measuring process. Hence, the on-machine measurement data adhere to the aforementioned multi-frame super-resolution assumption and can achieve resolution enhancement on a sub-pixel level.

B. Experimental analysis

Fig. 9 provides a quantitative comparison between the learning-based super-resolution method USRNet [21] and the proposed method. The original images, which serve as ground truth, are downscaled by a factor of two for the resolution enhancement experiments. The figure displays the super-resolved outputs alongside their respective error maps. The Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) serve as the evaluation metrics. Five samples recorded by the autostereoscopic system are analyzed, with the average metrics for all data presented in Table 1. Additionally, data captured by a two-dimensional vision device under vibration are utilized to validate the proposed method's effectiveness. The results clearly indicate that the proposed method outperforms in terms of higher PSNR and SSIM scores across all tested scenes.

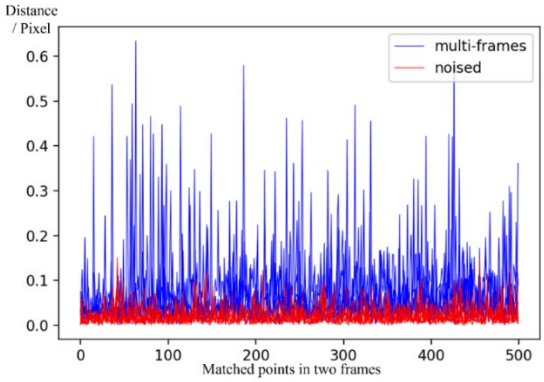

Fig. 8. Jitter analysis of the multiple measurement frames captured during the on-machine process (blue lines) and the noised data generated by imposing Gaussian noises (red lines).

Since the goal of resolution enhancement is to strengthen the feature points on the measured surfaces for more precise matching, evaluating the outcomes solely based on PSNR and SSIM is insufficient. A qualitative comparison among Bilinear, USRNet, and the proposed method is illustrated in Fig. 10, where the input data are expanded by a factor of four. Fig. 10 also includes local enlargements of the scenes, demonstrating that the proposed method achieves superior visual sharpness for high-frequency information. It is clear that the proposed method excels in enhancing high-frequency signals, typically the edges or key points of the measured sample, which are crucial for depth estimation in the shape-from-focus process. Enhancing these high-frequency signals aids in the accuracy of focus measurement on the refocused image stack, ensuring that corresponding points are detected at the correct depth plane.

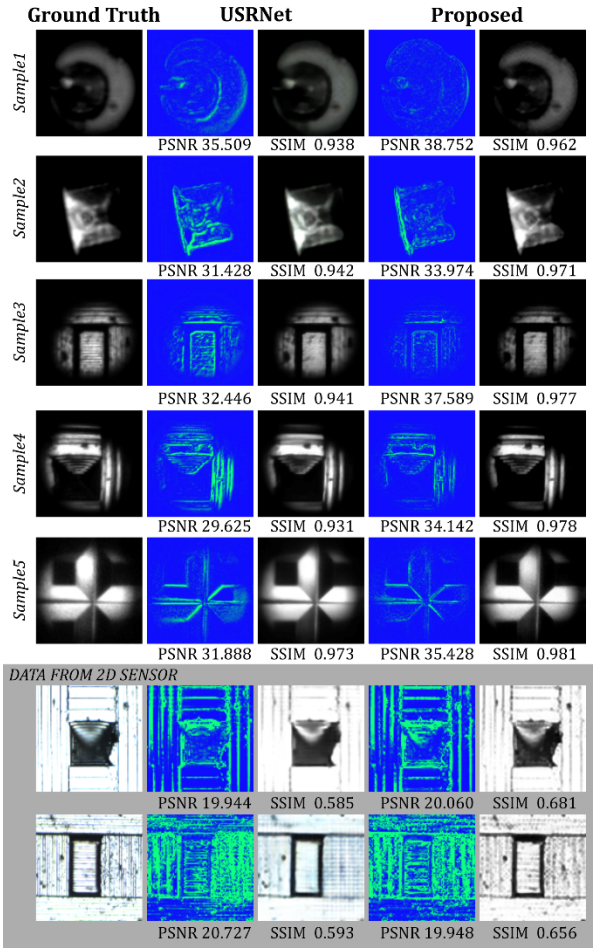

Fig. 9. Quantitative comparison on resolution-enhancement results by USRNet and the proposed. Error maps are exhibited to assess the discrepancies between the super-resolved outcomes and the ground truth.

TABLE I
QUANTITATIVE ANALYSIS (PSNR/SSIM) ACROSS FIVE SAMPLES RECORDED BY THE MEASURING SYSTEM

	USRNet	**The Proposed**
Sample1	35.740/0.953	38.931/0.970
Sample2	32.849/0.956	35.554/0.977
Sample3	32.296/0.924	36.742/0.973
Sample4	29.972/0.940	33.697/0.976
Sample5	33.301/0.975	36.722/0.981

Based on the autostereoscopy theory, digital refocusing is able to reconstruct a series of image slices with various focus depths so that the height of the measured sample is able to be detected. In terms of the detection of the focus region so as to determine the desired depth information, a Sobel filter is used as the focus measure operator. A curve of the focus levels is

obtained, and the peak value of the curve can be detected. Fig. 11 presents the digitally refocused images from low-resolution data alongside those obtained from the high-resolution data acquired by the proposed method, permitting a qualitative evaluation of the enhancements in the refocused images. The corresponding edge detection results, which are used for the identification of focus points, are displayed next to the local enlargements of the refocused images, vividly demonstrating the improvements in the focus detection process achieved by the high-resolution data.

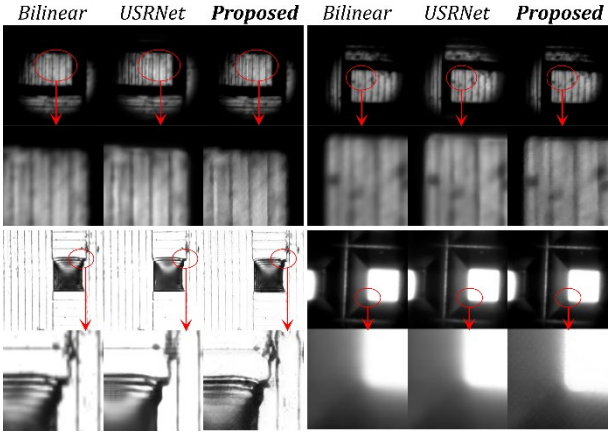

Fig. 10. Qualitative comparison on resolution-enhancement results using various methods, including Bilinear, USRNet, and the proposed. Local Enlargements are presented additional to evaluate the enhancement for the details

Fig. 11 also includes local enlargements of the scenes, demonstrating that the proposed method achieves superior visual sharpness for high-frequency information. It is clear that the proposed method excels in enhancing high-frequency signals, typically the edges or key points of the measured sample, which are crucial for depth estimation in the shape-from-focus process. Enhancing these high-frequency signals aids in the accuracy of focus measurement on the refocused image stack, ensuring that corresponding points are detected at the correct depth plane.

Fig. 12 shows a comparison between the conventional single-frame method and the proposed multi-frame resolution-enhanced deep-learning method through the all-in-focus image generated during shape from focus, the depth estimation results, and the point clouds, which vividly verify the resolution enhancement of the proposed method. A measurement result from a commercial measurement product – Zygo Nexview Optical profiler is presented as the reference. The resolution of each EI both in lateral and axial directions is enhanced by 4 folds. Disparity information is extracted from the stack of refocused images. For example, the focusing of the top and bottom surfaces in the refocused images shown in Fig. 12 (a) and (b) facilitates the determination of the depth of points located on these two surfaces. Subsequent to this, depth reconstruction and refinement are executed in accordance with the process shown in Fig. 6. For evaluative purposes, depth maps derived from both low-resolution and high-resolution data, along with their respective three-dimensional point clouds, are exhibited. Beyond the increased density of points afforded by multi-frame high-resolution data, the defects present in the low-resolution data are mitigated by the proposed multi-frame enhancement technique.

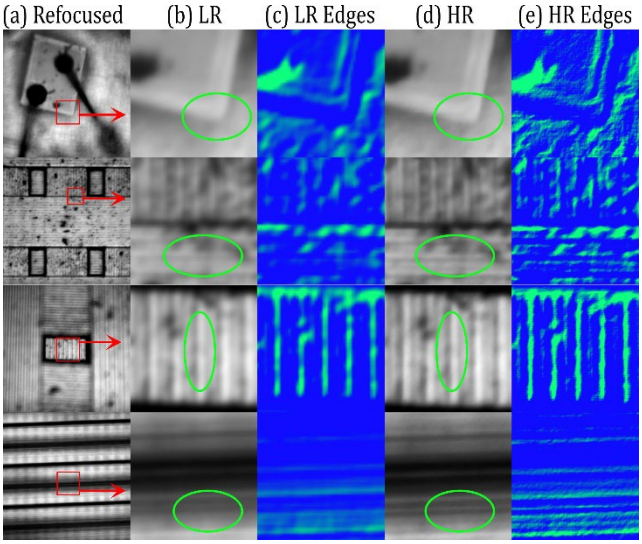

Fig. 11. Digital refocused results from low-resolution (LR) and high-resolution (LR) data, alongside their edge detection results, which ascertain the accuracy of the shape-from-focus method. (a) Refocused images. (b) Refocused images from the LR data. (c) Edge detection results of (b). (d) Refocused images from the multi-frame HR data. (e) Edge detection results of (d).

Fig. 13 shows the error maps analyzed by the iterative closest point (ICP) method which compares measured data acquired via repeated measurements. The repeatability of the proposed method displays better performance in terms of the standard deviation of 10 repeated measurements. However, it is still found that the deviation at the edges of the surface profile is much larger in both the single-frame and multi-frame systems. This could be resulted from the capability of the focus measurement operator which is sensitive to high-frequency signals. These signals may not only be edges and key points but could also be noise. Hence, further investigations and research for robust and adaptive focus measurement operators can further benefit the improvement of depth estimation accuracy for autostereoscopic measurement.

V. CONCLUSION

In this paper, the development of a multi-frame resolution-enhanced autostereoscopic system for on-machine 3D surface measurement is presented. The system takes advantage of the machine vibration together with a multi-frame resolution-enhanced deep-learning model to acquire multiple frames of the target surface profile with offsets to enhance the resolution and accuracy of on-machine 3D surface measurement. The results of performance evaluation show that the proposed system achieves higher measurement accuracy than the conventional

single-frame system in repeated measurements. The proposed method also improves the intensity of point cloud data with the additional improvement of measurement accuracy and robustness. Future work is suggested to concentrate on the focus detection method, which typically plays a crucial role in the surface reconstruction process utilizing shape-from-focus techniques. Additionally, devising a more robust method for directly retrieving depth information from autostereoscopic signals may also be a potential direction.

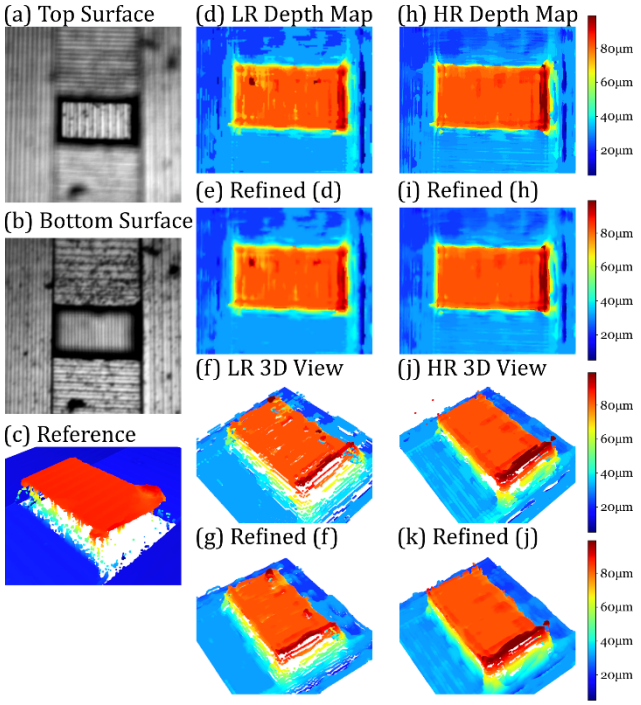

Fig. 12. Comparison of surface reconstruction results from single-frame (SF) system and multi-frame (MF) system. (a) A refocused image with the top surface in focus. (b) A refocused image with the bottom surface in focus. (c) The reference surface profile. (d) The depth map reconstructed by the SF system using low-resolution data. (e) The refined result of (d). (f) The point cloud generated by the SF system. (g) The refined result of (f). (h) The depth map reconstructed by the MF system using high-resolution data. (i) The refined result of (h). (j) The point cloud produced by the MF system. (k) The refined result of (j).

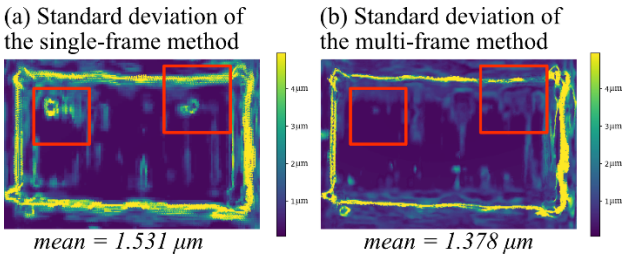

Fig. 13. Standard deviation of repeated measurements using (a) the traditional single-frame method and (b) proposed multi-frame resolution-enhanced method.

ACKNOWLEDGMENT

The work described in this paper was mainly supported by a grant from the Research Grants Council (Project No. R5047-22) of the Hong Kong Special Administrative Region, China).

REFERENCES

1. Tetsuka, H. and S.R. Shin, *Materials and technical innovations in 3D printing in biomedical applications.* Journal of materials chemistry B, 2020. **8**(15): p. 2930-2950.
2. Hinman, S.S., K.S. McKeating, and Q. Cheng, *Plasmonic sensing with 3D printed optics.* Analytical chemistry, 2017. **89**(23): p. 12626-12630.
3. Civcisa, G. and T. Leemet. *3D surface roughness parameters of nanostructured coatings with application in the aerospace industry*. in *Applied Mechanics and Materials*. 2015. Trans Tech Publ.
4. Lei, L., et al., *A 3D micro tactile sensor for dimensional metrology of micro structure with nanometer precision.* 2014. **48**: p. 155-161.
5. Bauza, M., et al., *Surface profilometry of high aspect ratio features.* 2011. **271**(3-4): p. 519-522.
6. Zhang, Y., et al., *Intelligent sampling strategy for freeform surfaces based on on-machine measurement through real-time interactive communication.* IEEE Transactions on Instrumentation and Measurement, 2023.
7. Zhu, W.-L., et al., *On-machine measurement of a slow slide servo diamond-machined 3D microstructure with a curved substrate.* Measurement Science and Technology, 2015. **26**(7): p. 075003.
8. Gao, W., et al., *On-machine and in-process surface metrology for precision manufacturing.* Cirp Annals-Manufacturing Technology, 2019. **68**(2): p. 843-866.
9. Hao, Q., et al., *Virtual interferometer calibration method of a non-null interferometer for freeform surface measurements.* 2016. **55**(35): p. 9992-10001.
10. Liu, Y., et al., *Full-field 3D shape measurement of discontinuous specular objects by direct phase measuring deflectometry.* 2017. **7**(1): p. 1-8.
11. Song, Z., et al., *A high dynamic range structured light means for the 3D measurement of specular surface.* 2017. **95**: p. 8-16.
12. Fu, S., et al., *In-situ measurement of surface roughness using chromatic confocal sensor.* 2020. **94**: p. 780-784.
13. Wang, Z., et al., *Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning.* Nature methods, 2021. **18**(5): p. 551-556.
14. Jia, C., et al., *Semantic segmentation with light field imaging and convolutional neural networks.* IEEE Transactions on Instrumentation and Measurement, 2021. **70**: p. 1-14.
15. Li, D., et al., *Disparity pattern-based autostereoscopic 3D metrology system for in situ measurement of*

microstructured surfaces. Optics letters, 2015. **40**(22): p. 5271-5274.
16. Zhou, P., et al., *Light field calibration and 3D shape measurement based on epipolar-space.* Optics Express, 2019. **27**(7): p. 10171-10184.
17. Kong, L. and P.J.I.J.o.E.M. Zhou, *A light field measurement system through PSF estimation by a morphology-based method.* 2021. **3**(4): p. 045201.
18. Gao, S. and C.F. Cheung, *Autostereoscopic 3D Measurement Based on Adaptive Focus Volume Aggregation.* Sensors, 2023. **23**(23): p. 9419.
19. Gao, S., C.F. Cheung, and D. Li, *Semi-supervised angular super-resolution method for autostereoscopic 3D surface measurement.* Optics Letters, 2024. **49**(4): p. 858-861.
20. Nava, F.P.r. *Super-resolution in plenoptic cameras by the integration of depth from focus and stereo*. in *2010 Proceedings of 19th International Conference on Computer Communications and Networks*. 2010.
21. Zhang, K., L.V. Gool, and R. Timofte. *Deep unfolding network for image super-resolution*. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
22. Lindeberg, T., *Scale invariant feature transform.* 2012.
23. Xu, Q., et al., *Effective Face Detector Based on YOLOv5 and Superresolution Reconstruction.* Computational and Mathematical Methods in Medicine, 2021. **2021**.
24. Lai, W.-S., et al. *Deep laplacian pyramid networks for fast and accurate super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
25. Johnson, J., A. Alahi, and L. Fei-Fei. *Perceptual losses for real-time style transfer and super-resolution*. in *European conference on computer vision*. 2016. Springer.
26. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv e-prints, 2014: p. arXiv:1409.1556.
27. Kingma, D.P. and J. Ba, *Adam: A Method for Stochastic Optimization.* arXiv e-prints, 2014: p. arXiv:1412.6980.
28. Li, D., et al., *Autostereoscopy-based three-dimensional on-machine measuring system for micro-structured surfaces.* Optics express, 2014. **22**(21): p. 25635-25650.