NanoNeRF: Robot-assisted Nanoscale 360° reconstruction with neural radiance field under scanning electron microscope

Xiang Fu, Yifan Xu, Hu Su*, Song Liu*, *Member, IEEE*

Abstract— The pursuit of 3D reconstruction from 2D images for nanomanipulation under scanning electron microscopy stands as a critical research endeavor. Previous methods either necessitates additional lighting which is difficult in standard SEM devices or relies on feature matching with low resolution and precision, further constraining reconstruction performance. In this paper, we propose a novel robot-assisted nanoscale 360° reconstruction approach, which simplifies SEM setups and maximizes the utilization of robot motion and feedback. By harnessing a nanorobotic system, we capture 360° multi-view images automatically with precise mapping information and camera postures. Sequentially, neural radiance field reconstruct the pixel-wise structure and synthesizing images from diverse perspectives. Experimental results using two real datasets demonstrates our approach's efficacy, achieving PSNR of 28.1 and SSIM of 0.93 for nanotube reconstruction, and PSNR of 32.8 and SSIM of 0.98 for AFM cantilever reconstruction. These results validate the reliability and robustness of our proposed robot-assisted reconstruction method.

I. INTRODUCTION

In recent years, nanorobotic manipulation under scanning electron microscopy (SEM) has emerged as a pivotal technique in nanomaterial characterization [1-4], nanoscale device fabrication [5], and the construction of nanoelectromechanical systems (NEMS) [6], showcasing its significant potential across various fields such as materials science, semiconductor research, and cell biology [7]. SEM offers invaluable insights into nanoscale topographies by employing a focused electron beam to scan the sample surface and capturing secondary and backscattered electrons resulting from beam-sample interactions. However, its limitation lies in providing primarily 2D images, which compromises depth information crucial for nanorobotic manipulation tasks. The lack of depth perception hampers complex 3D trajectory planning, visual tracking of posture during rotation, and 3D feedback, essential for intricate tasks like nano-assembly, material surface characterization, and cell manipulation.

Over recent years, various 3D reconstruction methods for SEM have been proposed, with a growing focus on computer vision techniques, classified in single-view [8-9] and multi-view approaches [10-16]. Single-view methods, such as photometric stereo techniques, analyze a set of 2D images from a single view with varying light directions. For example, Paluszynski et al. [8] and Xu Liu et al. [9] proposed photometric stereo methods for reconstructing depth
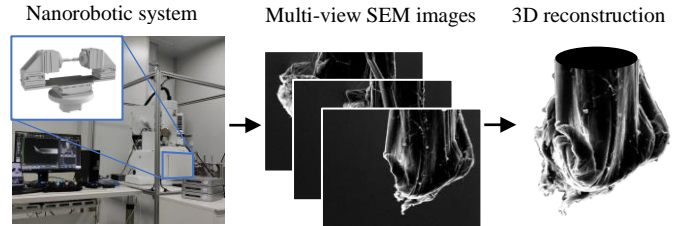
Fig. 1. Our approach utilizes a nanorobot to achieve in-situ rotation control for multi-view imaging under SEM. Corresponding camera postures are calculated, training NeRF for 3D reconstruction with the SEM images.

information, particularly suited for smooth samples and fiber-to-chip coupling, respectively. While these methods eliminate the need for additional camera views, their reliance on a single viewpoint limits their ability to achieve complete reconstructions. Moreover, these methods require additional lighting, making it challenge to capture images under various illumination directions in standard SEM setups.

Multi-view vision, on the other hand, utilizes feature point matching across multiple images to reconstruct the sample's 3D surface under SEM. For instance, methods proposed by Eulitz et al. [10], Zolotukhin et al. [11], and Samak et al. [12] focus on reconstructing surface roughness. Ding Weili et al. [13] introduced linear motion from nanorobots to provide epipolar-plane images for depth information restoration. Although multi-view methods alleviate the need for additional lighting, their reconstruction accuracy is contingent upon precise alignment and feature matching, which may be susceptible to inherent noise in SEM images. Furthermore, these methods typically reconstruct objects within views on a single side, limiting acquisition of complete structure.

To address these challenges, this paper proposes a robot-assisted 360° reconstruction approach utilizing neural radiance fields (NeRF), as illustrated in Fig 1. A nanorobot is established, implementing rotary manipulation to enable automated acquisition of 360° multi-view images through in-situ rotation. During automated image acquisition, camera postures are obtained referring to robot feedback. NeRF, a pixel-wise reconstruction technique, offers both structural and visible intensity information, facilitating visualization for characterization studies. We integrate NeRF with mapping between image coordinates and 3D physical space, calibrated by the nanorobot's three linear and one rotary nanopositioners. This novel method introduces robot-assisted deep learning techniques into 3D reconstruction under SEM, providing comprehensive 3D structural information and synthesizing images from novel perspectives.

II. NANOROBOTIC SYSTEM

A. Nanorobotic System Configuration

As illustrated in Fig. 2, a canonical nanorobotic system comprises an SEM camera and four nanopositioners,
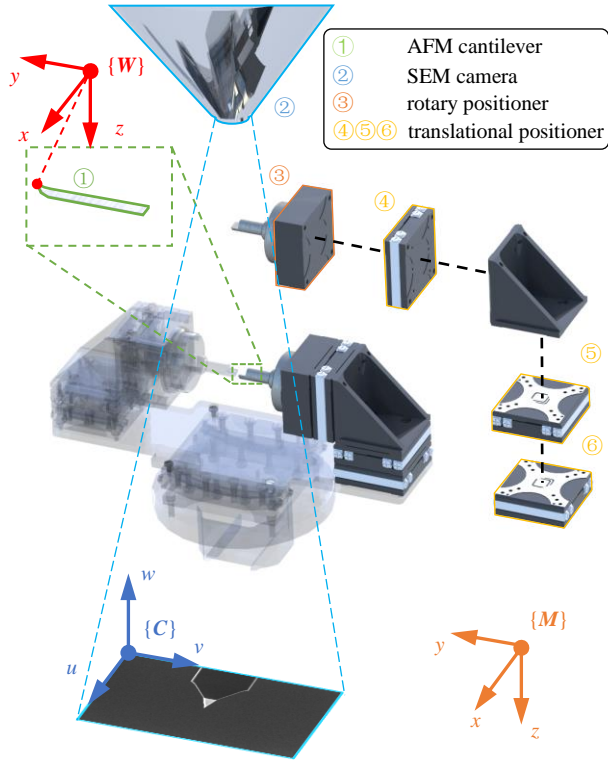
Fig. 2. Configuration of the nanorobotic system and coordinates system. The system contains 2 rotation units and the non-transparent unit is used in this experiment. Each unit is composed of three translational DOFs and one rotary DOF.

providing three translational degrees of freedom (DoFs) and one rotary DoF. The nanorobot is situated within the vacuum chamber of SEM, with the three translational nanopositioners orthogonally mounted and one rotary nanopositioner utilized at the end of the robotic arm. The translational nanopositioners provide a precision of 1 nm, while the rotary nanopositioner offers an angle precision of 1 m°, ensuring highly reliable feedback in in-situ rotation control for multi-view imaging. The observed sample is affixed at a cantilever mounted on the rotary nanopositioner. The SEM camera captures 2D images of the workspace from a single top view.

In Fig. 2, three coordinate systems are also portrayed. The 3D motion space $\{M\}$ is defined on the base of robotics system, with x-, y-, z-axes aligned with the orientations of the translational DoFs. The image space $\{C\}$ is established on the image plane of SEM camera, with u and v-axes representing two axes of SEM image and w-axis representing the optical axis which is orthogonal to u-v plane. The world space $\{W\}$ is defined on a tracked spatial point of the sample, with x-, y-, z-axes initially aligned with directions in $\{M\}$. As the sample is rotated, the origin and directions of $\{W\}$ alter in $\{M\}$ with the sample's rotation.

Under SEM, the sample is rotated and the three precise translational DoFs remain its in-situ rotation. Throughout the in-situ rotation control by nanorobot, the sample maintains its in-situ without moving out of field of view (FoV), and SEM obtains a series of multi-view images, using fast scanning mode to avoid image shifting. The corresponding camera postures are determined by referencing the feedback from the nanopositioners.

B. Kinematic Model for Nanorobot

During rotation, unexpected displacement of sample occurs due to the misalignment between the sample and the rotation axis. In order to maintain in-situ of the sample for multi-view imaging, the kinematic model for nanorobot on the sample is established as follows.

$$
\begin{cases}
X_t = R(f, \Delta\theta)X_{t-1} + u_t + \left(I - R(f, \Delta\theta)\right)\sum_{n=0}^{t-1} u_n \\
X_t = \begin{bmatrix} P_x^t & P_y^t & P_z^t \end{bmatrix}^T, u_t = \begin{bmatrix} u_x^t & u_y^t & u_z^t \end{bmatrix}
\end{cases}
\tag{1}
$$

where, at t^{th} time step, the state X_t and control input u_t separately represent the spatial position of the observed object (cantilever), the translation motion of the nano-positioners, and the influence brought by the additional random noise. $\Delta\theta$ is the rotation angle for each step which is constant during in-situ rotating control. $R(f, \theta)$ is the standard rotation matrix around the rotation axis f about rotation angle θ, which can be expressed as:

$$
R = \begin{bmatrix}
f_x f_x v\theta + c\theta & f_y f_x v\theta - f_z s\theta & f_z f_x v\theta + f_y s\theta \\
f_x f_y v\theta + f_z s\theta & f_y f_y v\theta + s\theta & f_z f_y v\theta - f_x s\theta \\
f_x f_z v\theta - f_y s\theta & f_y f_z v\theta + f_x s\theta & f_z f_z v\theta + s\theta
\end{bmatrix}
\tag{2}
$$

where $f = [f_x, f_y, f_z]^T$ is the normalized orientation vector of the rotation axis f, $c\theta$ and $s\theta$ are the cosine and sine of the rotation angle θ, respectively; $v\theta$ denotes $vers\theta$, defined as $(1 - \cos\theta)$.

III. RECONSTRUCTION APPROACH

A. Overview

Fig. 3 illustrates the overview of our proposed approach for reconstruction under SEM. A nanorobotic system is established under SEM, comprising three translational nanopositioners and one rotary nanopositioner. The pre-calibrated Jacobian matrix facilitates mapping between image system $\{C\}$ and world space $\{W\}$. During multi-view image acquisition, the rotary nanopositioner rotates the observed sample, providing different views captured from SEM. Simultaneously, the three linear nanopositioners are employed to compensate for displacement induced by rotation, necessitating precise rotating control to maintain in-situ for imaging. Consequently, a series of sequential multi-view images is automatically captured in fast scanning mode, with an averaging of 20 images for each single angle to denoise. Following image acquisition, the captured images are subsequently post-processed, providing a clear dataset. Meanwhile, the corresponding camera postures are computed, as discussed in Section II. C.

Incorporating the imaging process modeled encompassing emission and absorption phenomena, each pixel of an image from a particular view can be rendered by referencing the 3D neural function module, elaborated upon in Section II. D. For every ray corresponding to a pixel, multiple volumes are sampled along the ray, with the neural module mapping the spatial position and viewing direction to the volume's density and intensity which gauges the volume's capacity for emitting electrons by interaction. The camera projection is executed via affine transform, particularly at high magnification of SEM.
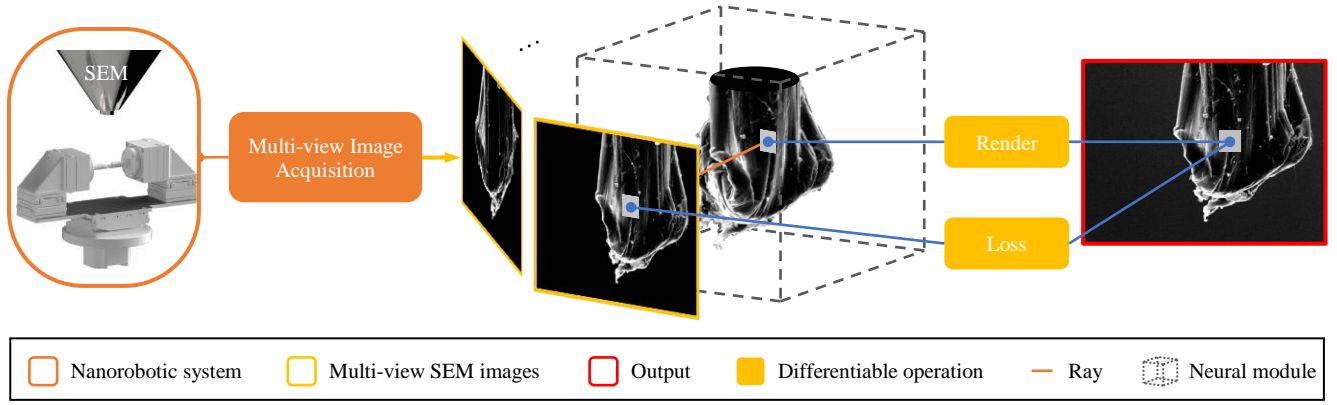
Fig. 3. Overview of our reconstruction approach. Nanorobotic system firstly provides multi-view SEM images and corresponding camera postures. Then, referring to the camera posture, a ray is generated and the intensity of the pixel is rendered, compared with the pixel of the given image to calculate the loss and train the neural module.

B. Robot-assisted Image Acquisition

To capture multi-view images under SEM, the nanorobot executes rotation motion. However, owing to the inherent spatial distance between the observed sample and the rotation axis, even small rotation angles can result in significant unexpected displacement of the sample. To obtain sequential multi-view SEM images, the nanorobot employs several control steps to maintain sample's in-situ. A spatial point on sample is anchored to enable in-situ control, continuously tracked using SEM feedback as a feature point during in-situ rotation of the sample.

As depicted in Fig. 4 (b), a spatial point on sample serves as the anchor, establishing the origin of the world space $\{W\}$. This anchor point is continuously tracked using SEM feedback as a feature point throughout the rotation process. During rotation, SEM camera orbits around the anchor point in the robot motion space $\{M\}$, yielding multiple imaging views. To ensure the anchor point's in-situ, an output x_d is designed as the expected spatial position. Utilizing the 2D information from SEM vision feedback, a Kalman filter-based observer is deployed to estimate the spatial position $\hat{X}$ for the anchor point on sample. The initial estimated spatial position X_0 is pre-calibrated. Subsequently, a PID controller is applied, and the compensatory input u_t for the three linear nanopositioners in the PID system can be expressed as

$$u_t = K_P \cdot e_t + K_I \cdot \sum_{n=1}^{t} e_n + K_D \cdot \left(e_t - e_{t-1} \right), e_t = \hat{X} - x_d \quad (3)$$

where e_t is the spatial displacement of the tracked point on sample at the tth step, and K_P, K_I, K_D are the parameters of the PID controller.

After the rough positioning of the sample using the PID controller, the sample is still subject to an image error of several micrometers. Consider camera projection as affine transform, fine positioning is then employed to reduce image error without depth error. Multiple control steps are executed until the image error is reduced to less than 2 pixels, equal to 400 nm at magnification of 1000. During each step of fine positioning, the compensation u_f can be expressed as follows.

$$u_f = J^{-1} \begin{bmatrix} \Delta u & \Delta v & 0 \end{bmatrix}^T \quad (4)$$

where Δu and Δv are the displacement on image plane.

Following the fine positioning for the anchor, images are captured, albeit with inherent noise and background interference. Extensive efforts have been devoted to denoising SEM images, with non-local means (NLM) method [17] as a promising approach, preserving image details while efficiently reducing noise. Hence, we employ NLM to denoise the images. Subsequently, to mitigate background interference
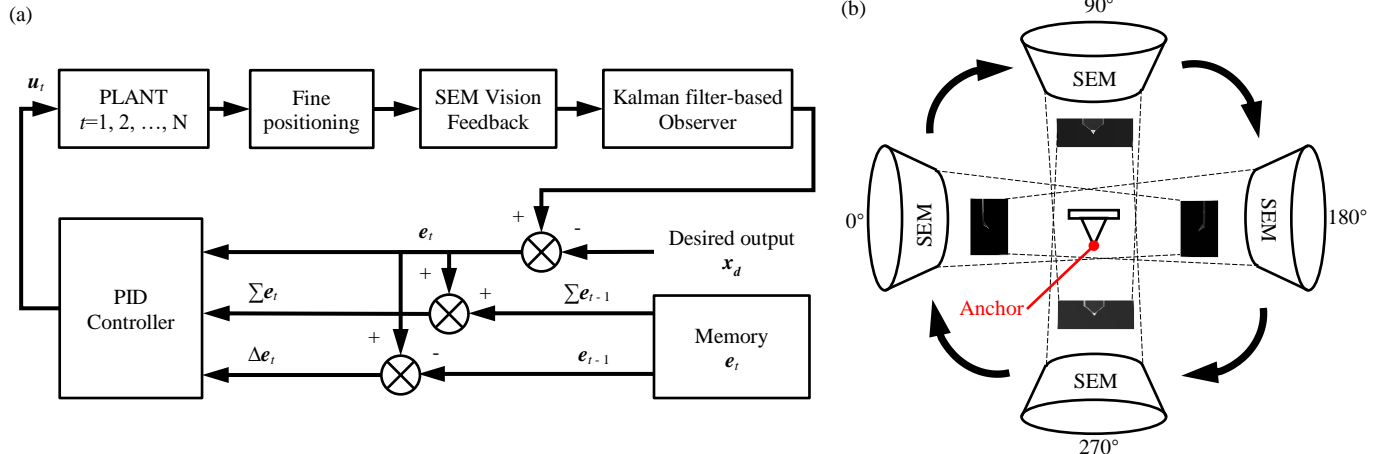

Fig. 4. (a) Overview of the in-situ rotating control. (b) Side-view diagram of SEM's capturing multi-view images. Take an AFM cantilever whose tip is set to be the anchor point as an instant, the sample is rotated by 360° to obtain images of full angle.

in the imaging model, we nullify the grayscale values in the background. Initially, coarse background segmentation is performed using the Canny operator to detect the edge and separate the images into foreground and background regions. The remaining background region is then manually removed. As a result, sequential multi-view images comprising only the foreground without noise are obtained.

C. Camera Pose Alignment

After capturing the series of sequential SEM images, the reconstruction process necessitates precise camera posture for each image. The camera posture delineates the transformation relationship between the camera system $\{C\}$ to the world space $\{W\}$. The spatial coordinates for neural model are established in $\{W\}$, while the posture furnishes spatial information for generating ray corresponding to each pixel during rendering. Throughout sample's 360° rotation for multi-view imaging, the transformation matrix H from the world space $\{W\}$ to camera system $\{C\}$ can be expressed as

$$H(\theta) = \begin{bmatrix} R(f,\theta) & R(f,\theta)T_0 \\ 0 & 1 \end{bmatrix} \quad (5)$$

where T_0 is the initial camera's spatial position.

The origin of camera system is determined on the anchored point on the image, which is extracted from the denoised SEM image. For instance, if the detected point's position on the image plane is (u_0, v_0) and the pixel position in $\{C\}$ for ray generation is denoted as $(i, j) = (u_1 - u_0, v_1 - v_0)$, where (u_1, v_1) is the pixel position on the image plane.

D. Rendering

In a scanning electron microscope, images are captured by directing an electron beam emitted from its internal electron gun across sample, and detecting secondary and backscattered electrons generated from interactions between the beam and the sample. The intensity of the emitted electrons is influenced by the topography and composition of the sample, as well as the interactions with primary electrons from the beam. Therefore, both emission and absorption occur during the imaging process, leading to emission-absorption model.

In an emission-absorption model [18], the differential formula of the intensity of the emitted electrons activated by a beam ray $r(t)$ can be expressed as

$$\frac{dE(r(t))}{dt} = \sigma(r(t))C(r(t),d) - \sigma(r(t))E(r(t)) \quad (6)$$

where $E(r(t))$ denotes the intensity of detected emitted electrons for the beam ray $r(t)$, σ is the volume density affecting both emission and absorption, C measures emitted electrons by interaction, which is related to position and direction. The ray of the electron beam $r(t) = o + t \cdot d$, where o is the origin of the ray, d is the direction and t is the distance.

The solution of the equation (6) is as follows.

$$E(r(t)) = \int_{t_n}^{t_f} T(t)\sigma(r(t))C(r(t),d)dt, T(t) = \exp(-\int_{t_n}^{t} \sigma(r(s))ds) \quad (7)$$

In practice, the equation is written in discrete form and adopts hierarchical volume sampling [19].

At high magnification under SEM, the projection model for imaging is an affine transformation, which differs from the

standard camera in macro photography. Consequently, for a given image, the rays corresponding to each pixel exhibit the same direction but vary in orientation. Suppose the transformation matrix H from the 3D motion space $\{M\}$ to the image space $\{C\}$ is given, the origin o and direction d for each ray starting from each pixel can be calculated as

$$d = H * \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T \quad (8)$$

$$o = i \cdot (JH*)^{-1} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T + j \cdot (JH*)^{-1} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T + T \quad (9)$$

where (i, j) is the pixel position on the image plane as the origin is the tracked feature point; J is the Jacobian matrix, transforming from 3D motion space to image space; $H*$ indicates the upper-left 3×3 submatrix in matrix H; and T is the forth column vector in H, depicting the translation from the camera system to the motion space. The Jacobian matrix is pre-calibrated through nanorobot's linear motion.

With imaging model and camera projection established, each pixel can be rendered using a 3D neural module that maps the position and viewing direction to the volume's density and intensity. The neural module is structured by a multilayer perceptron (MLP). Following the training of the 3D neural module, the sample can be reconstructed both structurally and visually.

IV. EXPERIMENT AND RESULTS

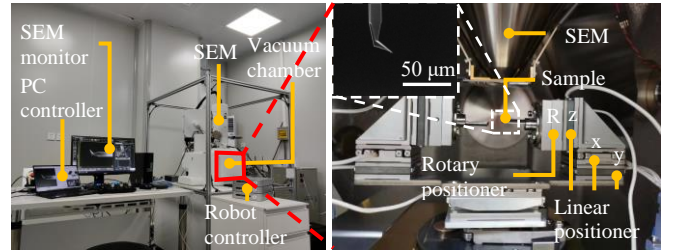### A. Automated Image acquisition

Fig. 5. Experimental nanorobotic system platform equipped with ECSx3030, ECR3030 nanopositioners (Attocube Systems AG), and JEOL JSM-IT500HR/LA InTouchScope™ SEM. AFM cantilever and carbon tube are used as robot end-effector for reconstruction validation.

Fig. 5 displays the established nanorobotic system. The scanning electron microscope used for automated image acquisition is JEOL JSM-IT500HR/LA, imaging samples in high vacuum mode by detecting secondary electrons, under 10 kV acceleration voltage. 7 frames are captured per second (frame/s) with image size of 640×480. An AFM cantilever and a carbon nanotube were placed on the end-effector of the nanorobot and reconstructed in this experiment as shown in Fig. 6. The AFM cantilever used in this experiment is NANOSENSORS ATEC-FM, and the nanotube has a diameter of 60 μm, with a cut fracture on its tip. The translational nanopositioners mounted in the nanorobotic system are Attocube piezoelectric ECSx3030, providing repeatability precision of 50 nm. The rotary nanopositioner is an Attocube piezoelectric ECR3030, with the rotation axis installed close to y-axis in $\{M\}$.

For automated image acquisition, in-situ rotation was implemented with several parameters calibrated. Orientation of the rotation axis f is calibrated as $f = [-0.01, 0.998, 0.02]^T$.
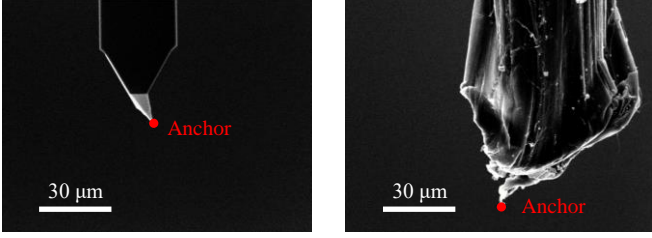
Fig. 6. Observed samples and the anchored point. Left is an AFM cantilever and right is a carbon nanotube.

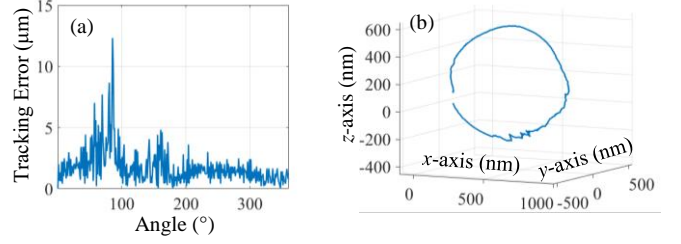Fig. 7. (a) Tracking error on image plane during rotation with PID controller. (b) Real path of linear nanopositioners during in-situ rotation.

The parameters in PID controller are set to be 0.6, 0, 0.1. The Jacobian matrix J is calibrated as

$$J = \begin{bmatrix} 2.20e-5 & 5.02e-3 & -6.25e-5 \\ 5.01e-3 & -8.66e-5 & -5.00e-6 \\ -1.54e-5 & -1.58e-5 & 4.97e-3 \end{bmatrix} \text{ pixel/nm} \quad (10)$$

In order to validate the PID controller's performance, an in-situ rotation experiment was conducted, where the misalignment is 400 μm. As shown in Fig. 7, the average error was 1.73 μm. In practice of imaging acquisition, the initial anchor's positions for AFM cantilever and carbon nanotube are calibrated to be $[396123, 0, 214680]^T$ nm and $[214680, 0, 214680]^T$ nm and SEM captured images at magnification of 1000. PID controller provided coarse positioning and the fine positioning ensured the tracked anchor point in-situ within 2 pixels. Considering tracking error, the anchor point was tracked again after denoising images to align camera posture.

B. Reconstruction results

1) Datasets

Here we present results of the reconstruction approach in two real datasets separately of AFM cantilever and a nanotube tip, collected for each one rotation angle (360 SEM images for each dataset). 60 images were split for training. The images of AFM cantilever were cropped in 300×300 pixels and the images of a nanotube were 640×480 pixels, as shown in Fig. 8. Background segmentation of the two datasets was implemented, with Canny operator detecting the edge and dividing the image into fore and background. The background was subtracted, as the volume density was zero.

For camera posture alignment, T_0 in Section IV. B was

given as $[0, 0, -10]^T$ as one unit equal to 10 μm. Rendering boundary t_f and t_n was properly set referring to the estimated volume space.

2) Metric

We consider synthesizing images of novel views to validate the performance of reconstruction under SEM. Peak Signal Ratio (PSNR), Structural Similarity Index (SSIM) and Mean Squared Error (MSE) are applied.

3) Training details

In terms of training MLP, we use an ADAM optimizer with different learning rates for different datasets and loss functions. For dataset of AFM cantilever, the learning rate is set to be 5e-5 and 5e-7 as loss function of L1 and MSE. For dataset of nanotube, the neural module was trained with a learning rate of 5e-4. We use 256 neurons in the hidden layers of the 3D reconstruction module, and apply a batch of 1024 rays for each sampling 64 coordinates in the coarse volume and 128 additional coordinates in the fine volume. The networks were trained for 500,000 and 600,000 iterations separately for AFM cantilever and CNT, and validation error was reported every 50000 iterations. The optimization for each dataset typically takes about 400,000~600,000 iterations to converge on a single NVIDIA 4070Ti Super (14 hours).

4) Reconstruction outcome

Fig. 8 displays the qualitative results of the experiments on the two real datasets. For the AFM cantilever dataset, the reconstruction approach achieved a PSNR of 32.82, SSIM of 0.9788 and MSE of 6.733e-4. Regarding the nanotube dataset, the reconstruction approach attained a PSNR of 28.13, SSIM of 0.9167 and MSE of 0.0017.

An ablation study was conducted to analyze the impact of loss function, as presented in Table I. For the dataset with high texture complexity, both loss functions exhibited similar performance. However, for the dataset characterized by low
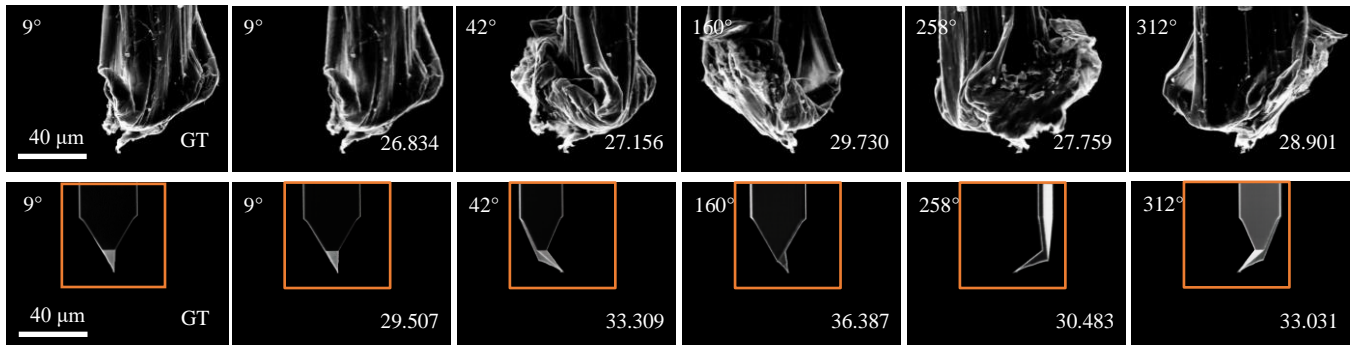

Fig. 8. Synthesized results and ground truth at different rotation angle for the two datasets. The presented images are respectively at angle of 9°, 42°, 160°, 258°, and 312°. The first row is on dataset of a nanotube. The second row is on dataset of an AFM cantilever. The rectangles with orange border are the actual rendered region of 300×300 pixels. All the data were captured at magnification of 1000. Peak Signal Ratio (PSNR) for each outcome is presented at the lower right corner and the corresponding rotation angle is at the upper left corner.
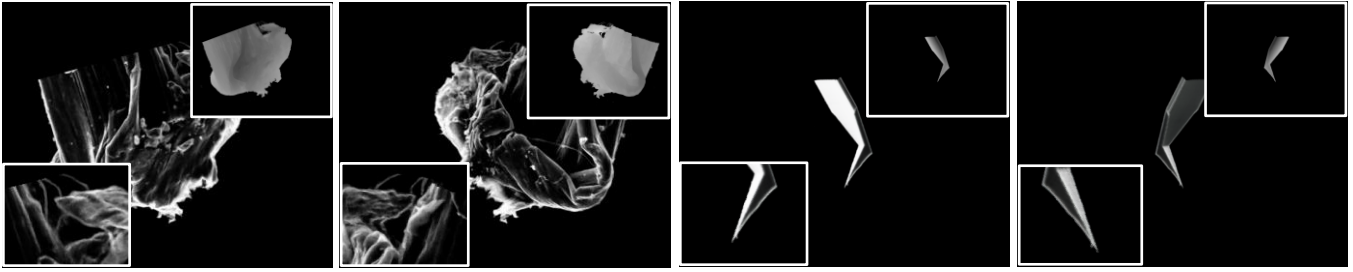
Fig. 9. Synthesized results from novel views. For each image, the left lower rectangle presents the details and the right upper rectangle displays depth map.

TABLE I. EXPERIMENTAL RESULTS ON RECONSTRUCTION

Sample	Loss		PSNR↑	SSIM↑	MSE↓
	L_{l1}	L_{mse}			
AFM		√	**32.82**	**0.9788**	**6.733e-4**
AFM	√		/	/	/
CNT		√	**28.13**	0.9167	**0.0017**
CNT	√		27.60	**0.9275**	0.0019

texture complexity and a high proportion of zero intensity (black) pixels, L1 loss function proved ineffective in training the neural module. This failure occurs because black pixels on the AFM cantilever were prone to being interpreted as a volume density of zero.

Due to the limited from single rotary DoF of nanorobot, the multi-view images were captured with views rotated around y-axis in $\{M\}$, resulting in an incomplete dataset with limited perspectives. Apart from those views, we synthesized images and depth maps from novel views, as presented in Fig. 9. The details are well reconstructed, such as on the fracture.

V. CONCLUSION

A robot-assisted reconstruction approach for SEM images is proposed, with nanorobot automatically acquiring multi-view images and mapping between image coordinates and physical space. Neural radiance field is employed, reconstructing the structural information and synthesizing images from novel views. The experiments present a result of PSNR 28.13 for a nanotube and PSNR 32.82 for an AFM cantilever, validating the efficiency. Images from other novel views were also synthesized, reflecting the robustness.

REFERENCES

[1] J. Qu et al., "An SEM-Based Nanomanipulation System for Multi-Physical Characterization of Single InGaN/GaN Nanowires," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, 2020, pp. 2866-2871.

[2] J. Qu et al., "An SEM-Based Nanomanipulation System for Multiphysical Characterization of Single InGaN/GaN Nanowires," in IEEE *Transactions on Automation Science and Engineering*, vol. 20, no. 1, pp. 233-243, Jan. 2023.

[3] Huiyang Ding, and Li Ma, "Visual Servoing-Based Nanorobotic System for Automated Electrical Characterization of Nanotubes inside SEM," in *Sensors*, vol. 18, no. 4, pp. 1137, 2018.

[4] N. Yu et al., "Characterization of the Resistance and Force of a Carbon Nanotube/Metal Side Contact by Nanomanipulation", in *Scanning*, vol. 2017, pp. 123–135.

[5] P. Wang, H. Lu, and Y. Shen, "Flexible 3-D Helix Fabrication by In-Situ SEM Micromanipulation System," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 7, pp. 5565-5574, July 2020.

[6] A.A.G. Requicha, "Nanorobots, NEMS and Nanoassembly," *Proceedings of the IEEE*, vol. 91, no. 11, pp. 1922–1933, 2003.

[7] W. Shang, H. Lu, Y. Yang and Y. Shen, "7-DoFs Rotation-Thrust Microrobotic Control for Low-Invasive Cell Pierce via Impedance Compensation," in *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 6, pp. 5095-5106, Dec. 2022.

[8] J. Paluszyński, W. Slowko, "Surface Reconstruction with the Photometric Method in SEM," in *Vacuum*, vol. 78, pp.533-537, 2005.

[9] X. Liu, H. Hu, W. Jiang and X. Sun, "Object-oriented modified photometric stereo algorithm for 3D imaging in fiber-to-chip coupling," *The 9th International Conference on Group IV Photonics (GFP)*, San Diego, CA, USA, 2012, pp. 255-257.

[10] M. Eulitz, G. Reiss. "3D Reconstruction of SEM Images by Use of Optical Photogrammetry Software," in *Journal of structural biology*, vol. 191, no. 5, pp. 190-196, 2015.

[11] A. A. Zolotukhin, I. V. Safonov, and K. A. Kryzhanovskii, "3D Reconstruction for a Scanning Electron Microscope," in *Pattern recognition and image analysis*, vol. 23, pp. 168-174,2013.

[12] D. F. A. R. D. Samak, A. Fischer, and D. Rittel, "3D Reconstruction and Visualization of Microstructure Surfaces from 2D Images," in *CIRP annals*, vol. 56, no. 1, pp. 149-152, 2007.

[13] W. Ding et al., "Automatic 3D Reconstruction of SEM Images based on Nano-robotic Manipulation and Epipolar Plane Images," in *Ultramicroscopy*, vol. 200, pp. 149-159, 2019.

[14] Ahmad P. Tafti et al., "3DSEM++: Adaptive and Intelligent 3D SEM Surface Reconstruction," in *Micron*, vol. 87, pp. 33-45, 2016.

[15] A. V. Krishna et al., "Surface Topography Characterization using 3D Stereoscopic Reconstruction of SEM Images," in *Surface Topography: Metrology and Properties*, vol. 6, no.2, pp. 024006, 2018.

[16] A. Kudryavtsev, "3D Reconstruction in Scanning Electron Microscope: from Image Acquisition to Dense Point Cloud," *Université Bourgogne Franche-Comté*, 2017.

[17] A. Buades, B. Coll, J. M. Morel, "Non-local Means Denoising," in *Image Processing On Line*, vol. 1, pp. 208-212, 2011.

[18] J. T. Kajiya, and B. P. Von Herzen, "Ray Tracing Volume Densities," in *ACM SIGGRAPH computer graphics*, vol. 18, no. 3, pp. 165-174, 1984.

[19] B. Mildenhall et al, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *Communications of the ACM*, vol. 65, no. 1, pp. 99-106, 2021.