Multi-Scale Visual Servoing Framework for Optical Microscopy Based on SIFT Matching

Yameng Zhang, *Graduate Student Member, IEEE*, Ao Xu, Yuhan Chen, *Member, IEEE*, Max Q.-H. Meng, *Fellow, IEEE*, and Li Liu, *Member, IEEE*

***Abstract*—This letter introduces an innovative multi-scale visual servoing framework for optical microscopy, engineered to automatically reposition the microscope for high-magnification target view across multiple magnifications, thereby facilitating repetitive and accurate histologic biopsies. The framework encompasses an active microscope-camera system equipped with both auto-calibration and multi-scale visual servoing capabilities. The auto-calibration technique addresses the challenges posed by the limited depth of field and pattern requirements of the microscope-camera system, and determines its intrinsic and hand-eye parameters through a two-step algorithm. The calibration data is then utilized to execute a SIFT matching-based visual servoing control at progressively increasing magnifications, using only a single high-magnification target view as a reference, ultimately enabling rapid and precise repositioning of the microscope. Experimental results demonstrate the precision and stability of the auto-calibration method, as well as the robustness of the visual servoing method against occlusion, blur, and low illumination.**

***Index Terms*—Optical microscope, auto-calibration, multi-scale visual servoing, SIFT matching.**

I. Introduction

HISTOLOGIC biopsies are medical procedures that extract cells or tissues from the body for examination, often utilized in the diagnosis of cancer and other diseases. Despite their invasive nature, biopsies are essential for accurate diagnosis and treatment planning. Once a biopsy specimen is collected, it is analyzed *ex vivo*, typically involving examination under an optical microscope. In histologic biopsies, specific tissue generally has certain standard views that enable physicians to make rapid and accurate diagnoses. As such, tuning the microscope to find these standard views is a crucial task. Additionally, physicians may perform medical treatment on specimens and replace them on the stage to evaluate their evolution over time. Therefore, it is necessary to tune the microscope to bring the standard view back into its field of view (FOV) for comparative studies [1]. The objective of this work is to develop a framework that can quickly and accurately reposition a microscope for a specific target view, particularly at high magnifications, thereby enhancing the efficiency of histologic biopsies.

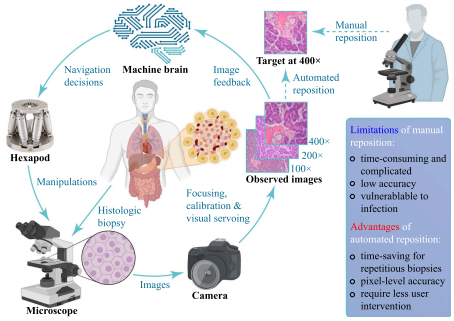

Fig. 1. Microscope repositioning for target view at $400\times$ magnification: A specimen is positioned under a microscope and imaged using a camera. The image is transmitted to a computer for analysis, where navigation decisions are made based on focusing, calibration, and visual servoing algorithms. These decisions direct a Hexapod robot to drive the microscope to the target view. This figure was created with BioRender.com.

As shown in Fig. 1, a microscope repositioning framework for a target view at $400\times$ magnification is presented. The caption provides a brief description of the repositioning process. Multi-scale Visual Servoing (VS) control is used as the core of the framework to align the microscopic views at $100\times$, $200\times$, and $400\times$ with the $400\times$ target view in sequence. This approach adjusts from coarse to fine, with an expanded FOV, thus ensuring a broader convergence range than direct VS control [2] at $400\times$. Compared to manual repositioning, the proposed method outperforms in terms of time efficiency, simplicity, accuracy, and minimal intervention to the specimen.

The implementation of multi-scale VS control necessitates the acquisition of a high-quality image and the precise determination of intrinsic and hand-eye parameters [3], [4]. This, in turn, requires the execution of prerequisite autofocusing and system

Manuscript received 4 June 2023; accepted 27 September 2023. Date of publication 18 October 2023; date of current version 30 October 2023. This letter was recommended for publication by Associate Editor L. Xin and Editor X. Liu upon evaluation of the reviewers' comments. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1312400, in part by the Hong Kong Research Grants Council (RGC) Theme-based Research Scheme (TRS) under Grant T42-409/18-R, and in part by the Hong Kong RGC General Research Fund (GRF) under Grants 14204321 and 14220622. *(Corresponding author: Li Liu.)*

Yameng Zhang and Ao Xu are with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, SAR, China (e-mail: zhangyameng@link.cuhk.edu.hk; xuao575@gmail.com).

Yuhan Chen is with the Shenzhen Key Laboratory of Robotics Perception and Intelligence, Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: chenyh7@sustech.edu.cn).

Max Q.-H. Meng is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, SAR, China, also with the Shenzhen Key Laboratory of Robotics Perception and Intelligence, Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen 518055, China, and also with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada (e-mail: max.meng@ieee.org).

Li Liu is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, SAR, China, and also with the Shenzhen Key Laboratory of Robotics Perception and Intelligence, Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: liliu@cuhk.edu.hk, liu6@sustech.edu.cn).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2023.3325688, provided by the authors.

Digital Object Identifier 10.1109/LRA.2023.3325688

calibration procedures, respectively [5]. To date, extensive research has been conducted on camera calibration for macro-scale applications, resulting in robust calibration procedures [6], [7]. However, micro-scale applications have distinct qualities that set them apart from regular camera calibration. The microscope, based on optical principles, has a limited depth of field (DOF) that requires the specimen to be positioned within a narrow range perpendicular to the optical system axis to capture a sharp image. This characteristic hinders the implementation of conventional multiplane-based calibration methods, as the image quality may significantly deteriorate when the calibration pattern is placed in different orientations. Tsai's algorithm [7], for example, requires the use of a calibration pattern tilted at a minimum of 30° in relation to the image plane, rendering the algorithm singular in the parallel case. Near-parallel approximations have been made in [8] to address this issue. However, pre-calibrating the focal length is required by the algorithm, which is not practical in microscopy applications. Zhang's algorithm [6], [9] employs several planar calibration patterns and the concept of homography transformation. While parallel-image calibration is feasible, it reduces the number of solvable intrinsic parameters. In addition to the optical constraint, the fabrication of reliable calibration patterns that contain micro-markers, which serve as distinguishable reference points, poses a significant challenge due to their complexity and high cost. Besides, calibration patterns require meticulous installation and removal whenever magnification levels are altered, which limits operational flexibility.

Another significant challenge for microscope repositioning is the multi-scale VS method. Our goal is to align the microscopic views at $100\times$, $200\times$, and $400\times$ with the $400\times$ target view in sequence. Conventional image-based visual servoing (IBVS) methods, such as Photo-VS [10], PCA-VS [11], and DCT-VS [12], are ineffective in this scenario due to inconsistencies in image gradients. Felton et al. [13], [14] have developed VS methods in autoencoder latent space, which show potential for addressing this challenge. However, their methods rely on deep metric learning, which is computationally demanding and challenging to fit various specimens. Li et al. [15], [16] have developed a zooming-free VS method for nanorobot end-effector navigation under a microscope. Nonetheless, their method focuses solely on the coordinate of a single end-effector tip rather than the entire image, rendering it unsuitable for our applications.

This study aims to address the challenges associated with microscope repositioning by proposing a multi-scale VS framework for optical microscopy. The main contributions are summarized as follows:

1) An active microscope-camera system was built, integrating a parallel robot for actuation and a camera as an optical sensor, enabling automatic repositioning for the target view.
2) An auto-calibration method was devised to determine the intrinsic and hand-eye parameters of the microscope-camera system, utilizing the SIFT matching technique and transferring a highly nonlinear optimization problem to linear regression problems, enhancing calibration stability. This method is calibration pattern-free.
3) A SIFT matching-based VS method was developed to enable the microscope-camera system to be repositioned for a target view across various magnifications, while also being robust against disturbances. This method uses only a single target view as a reference.

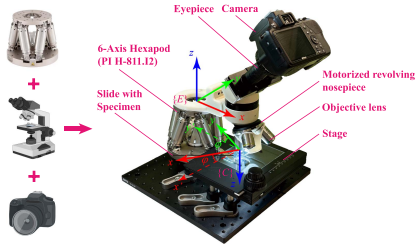

Fig. 2. Configuration of microscope-camera system.

The remainder of this letter is organized as follows: Section II presents the system configuration and robot tasks. Section III establishes the preliminaries of our method. Section IV elaborates on the proposed auto-calibration and multi-scale VS methods. Section V evaluates our approaches through experiments. Finally, Section VI concludes the letter.

II. System Description and Robot Tasks

Fig. 2 illustrates the configuration of our microscope-camera system. Unlike conventional microscope systems, our design incorporates a camera attached to the eyepiece, allowing for the extraction of microscopic images through an image capture card. Moreover, the microscope body is mounted on a 6-axis miniature Hexapod (PI H-811.I2), which employs a parallel-kinematic design to provide six degrees of freedom. The Hexapod offers a travel range of (± 17 mm, ± 16 mm, ± 6.5 mm, $\pm 10^\circ$, $\pm 10^\circ$, $\pm 21^\circ$) in the directions of $(X, Y, Z, \theta_X, \theta_Y, \theta_Z)$, with unidirectional repeatability of ($\pm 0.15\,\mu$m, $\pm 0.15\,\mu$m, $\pm 0.06\,\mu$m, $\pm 2\,\mu$rad, $\pm 2\,\mu$rad, $\pm 3\,\mu$rad). This level of precision enables the microscope to locate intricate features of specimens at high magnifications. To facilitate the automation of our optical microscope system, we have integrated a stepper motor into the revolving nosepiece, enabling automatic switching of objective lenses. The eyepiece magnifies at $10\times$, while the objective lenses offers magnifications of $10\times$, $20\times$, and $40\times$. By rotating the revolving nosepiece, the overall magnifications of the microscope system can be switched to $100\times$, $200\times$, and $400\times$. Fig. 2 also presents two frames: the microscope frame $\{C\}$ and the robot end-effector frame $\{E\}$. In the context of optical microscopy, it is crucial to maintain parallel alignment between the object and image planes. To achieve this, the Hexapod is meticulously programmed to align the xy plane of frame $\{C\}$ with the object plane in a parallel orientation. However, a potential z-axis rotational displacement between frame $\{C\}$ and $\{E\}$ exists, which can be quantified by φ.

The primary objective of the robot is to accurately reposition the microscope for target view at a high magnification of $400\times$. This objective entails three subtasks: autofocusing to achieve the sharpest image by adjusting the distance between the objective lens and the specimen, system calibration to determine the intrinsic and hand-eye parameters for precise control of the microscope, and VS control to reliably locate and observe the target view.

III. Preliminaries

A. Perspective Projection

Calibration is the process of finding the parameters that define the mapping between 3D coordinates of a point, denoted as $\mathbf{X} = (X, Y, Z)$, in the current microscope frame $\{C\}$, and its

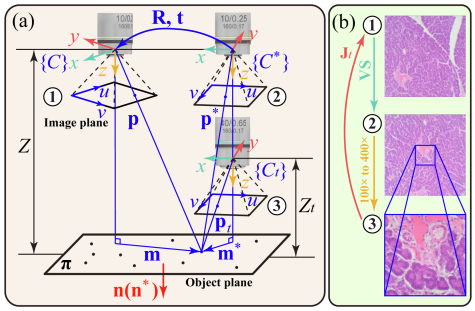

Fig. 3. Notations and geometric model for VS task: (a) Positions 1 and 2 represent the initial and target positions of the 100× objective lenses, while position 3 represents the target position of the 400× objective lens. Positions 1 and 2 are co-planar, and positions 2 and 3 are co-axis. (b) Images 1, 2, and 3 are captured at positions 1, 2, and 3, respectively. VS control aligns the 100× view with the 400× target view at the center using the projective homography matrix $\mathbf{J}_t$ of image 1 *wrt.* image 3.

projection $\mathbf{u} = (u, v)$ onto the image plane. By introducing notations $\mathbf{p} = (u, v, 1)$ and $\mathbf{m} = (X, Y, 1)$, this mapping can be represented as follows:

$$\mathbf{p} = \begin{bmatrix} k_x & s & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{m} = \mathbf{K}\mathbf{m}. \tag{1}$$

$\mathbf{K}$ is the intrinsic matrix, with k_x and k_y representing the pixel/meter ratio, s representing the orthogonality between the image frame axis, and (u_0, v_0) denoting the principal point coordinate in the image plane. Notably, we can write the perspective projection in (1) since Z is typically constant at a given magnification.

In our approach, we adopt a simplified method by setting the origin approximately at the center of the image and assume the skew parameter s is zero [17]. Additionally, we do not account for radial or tangential distortion, as microscope distortion is generally minimal. Thus, the calibration process only involves determining the values of k_x and k_y as the unknowns.

B. Hand-Eye Matrix

The hand-eye matrix characterizes the spatial relationship between the microscope frame $\{C\}$ and the robot end-effector frame $\{E\}$, which are rigidly connected [18]. The hand-eye matrix, denoted as ${}^{E}\mathbf{T}_C \in \mathbb{SE}(3)$, is expressed as:

$${}^{E}\mathbf{T}_C = \begin{bmatrix} {}^{E}\mathbf{R}_C & {}^{E}\mathbf{t}_C \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}. \tag{2}$$

The matrix ${}^{E}\mathbf{T}_C$ comprises a rotational matrix ${}^{E}\mathbf{R}_C \in \mathbb{SO}(3)$, which describes the orientation of frame $\{C\}$ relative to $\{E\}$, and a translational vector ${}^{E}\mathbf{t}_C = \boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3) \in \mathbb{R}^3$, which denotes the position of frame $\{C\}$ relative to $\{E\}$ in terms of frame $\{E\}$. As described in Section II, ${}^{E}\mathbf{R}_C$ can be provided directly by:

$${}^{E}\mathbf{R}_C = \begin{bmatrix} -\sin\varphi & -\cos\varphi & 0 \\ -\cos\varphi & \sin\varphi & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{3}$$

By means of ${}^{E}\mathbf{T}_C$, a relationship between the displacements of the microscope and robot end-effector can be established. Let $\Delta\mathbf{x}_c = (x_c, y_c, z_c, \theta_{cx}, \theta_{cy}, \theta_{cz})$ and $\Delta\mathbf{x}_e = (x_e, y_e, z_e, \theta_{ex}, \theta_{ey}, \theta_{ez})$ denote the 6-D displacement vectors of the microscope and the robot end-effector in their own frames $\{C\}$ and $\{E\}$, respectively. It is asserted that

$$\begin{cases} \theta_{cx} = \theta_{cy} = \theta_{ex} = \theta_{ey} = 0, \\ z_c = z_e = 0, \end{cases} \tag{4}$$

since the object and image planes should remain parallel and the DOF is fixed at a given magnification. This allows us to establish the following relationship:

$$\begin{cases} x_c = (\eta_2 - y_e)\cos\varphi - \eta_2\cos(\varphi + \theta_{ez}) \\ \quad +(\eta_1 - x_e)\sin\varphi - \eta_1\sin(\varphi + \theta_{ez}), \\ y_c = (\eta_1 - x_e)\cos\varphi - \eta_1\cos(\varphi + \theta_{ez}) \\ \quad +(y_e - \eta_2)\sin\varphi + \eta_2\sin(\varphi + \theta_{ez}), \\ \theta_{cz} = -\theta_{ez}. \end{cases} \tag{5}$$

C. Euclidean Homography and Projective Homography

Homography characterizes the projective geometry of two cameras (or a single camera at two different views) and a world plane [17]. Euclidean homography and projective homography are two important types of homography, which are defined and illustrated in Fig. 3.

Fig. 3(a) shows three defined objective lens positions, labelled 1, 2, and 3. Positions 1 and 2 represent the initial and target positions of the 100× objective lens, while position 3 represents the target position of the 400× objective lens. Positions 1 and 2 share an identical DOF of Z, while position 3 has a shorter DOF of Z_t. Fig. 3(b) shows the observed images at positions 1, 2, and 3. The VS control serves to actuate the microscope so that features of interest in image 3 are centered in the current view, as shown in image 2, based on the continuously computed projective homography matrix $\mathbf{J}_t$ of image 1 *wrt.* image 3. Additionally, if no positional error exists in the microscope's revolving nosepiece, positions 2 and 3 are coaxial. As a result, position 3 can be achieved with just a vertical translation from position 2, meaning that Euclidean features $\mathbf{m}^*$ observed at positions 2 and 3 are identical. It is noteworthy that the current magnification 100× in Fig. 3 can also be 200× or 400× depending on the VS stages.

Suppose each feature $\mathbf{m}$ expressed in frame $\{C\}$ (or $\mathbf{m}^*$ expressed in frame $\{C^*\}$) belonging to the object plane $\boldsymbol{\pi}$. $\mathbf{n}$ and $\mathbf{n}^*$ are the normal vectors of $\boldsymbol{\pi}$ expressed in frames $\{C\}$ and $\{C^*\}$, respectively. Then, the following relationship holds:

$$\mathbf{m}^* = \left(\mathbf{R} + \mathbf{t}\mathbf{n}^{*\top}\right)\mathbf{m}, \tag{6}$$

where $\mathbf{R} \in \mathbb{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$ represent the rotational matrix and translational vector of frame $\{C\}$ *wrt.* $\{C^*\}$, respectively. The Euclidean homography matrix $\mathbf{H}$ can be defined as:

$$\mathbf{H} = \mathbf{R} + \mathbf{t}\mathbf{n}^{*\top} = \begin{bmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{bmatrix}, \tag{7}$$

where θ denote the rotational displacement around the z-axis, and x and y represent the translational displacements along the x and y axes, respectively, between frames $\{C\}$ and $\{C^*\}$.

Based on (1), (6), and (7), we obtain the relationship:

$$\mathbf{p}_t = \mathbf{K}_t\mathbf{H}\mathbf{K}^{-1}\mathbf{p}, \tag{8}$$

where $\mathbf{K}_t$ and $\mathbf{K}$ denote the intrinsic matrices of lenses 3 and 1, respectively. The projective homography matrix $\mathbf{J}_t$ can be defined as:

$$\mathbf{J}_t = \mathbf{K}_t\mathbf{H}\mathbf{K}^{-1}. \tag{9}$$

To compute $\mathbf{J}_t$, an affine transformation estimation algorithm should be applied.

TABLE I
STATE-OF-THE-ART FEATURE DETECTORS AND DESCRIPTORS

Categories	Descriptors	Detectors	Speed	Accuracy
Handcrafted feature-based methods	SIFT [22]	DoG	Slow	**Very high**
	AKAZE [23]	Hessian	Medium	Medium
	ORB [24]	FAST [25]	**Very fast**	Low
	BRISK [26]	FAST [25]	Fast	High
Learning-based methods	SuperGlue [27]	SuperPoint [28]	Attain high accuracy, but require training on specialized datasets.	
	LoFTR [20]	Detector-free		
	DKM [29]	Detector-free		
	ASpanFormer [19]	Detector-free		

D. Affine Transformation Estimation

The estimation of an affine transformation can be achieved through a combination of feature detector, descriptor, matcher, and affine parameter calculator. Table I presents state-of-the-art algorithms for feature detectors and descriptors, categorized as handcrafted feature-based and learning-based methods. Learning-based methods excel at finding correspondences in texture-less areas and repetitive patterns without relying on keypoint detection [19], [20]. However, they require extensive training on specialized datasets, which is computationally demanding and less adaptable to different types of biopsy specimens. In contrast, handcrafted feature-based methods are simpler to implement and more versatile across various specimen types, making them more suitable for our purposes. For feature matching, both the Fast Library for Approximate Nearest Neighbors (FLANN) and Brute Force (BF) methods are commonly used, with the BF method generally providing higher matching values [21]. Among handcrafted feature-based methods, SIFT is the most accurate feature descriptor for variations in scale, rotation, and illumination. Therefore, the SIFT/BF matcher is chosen to estimate the affine transformation, which involves computing a projective homography matrix $\mathbf{J}_t$ that establishes a correspondence between a source image and a target image. The affine transformation from a source image keypoint $\mathbf{p} = (u_s, v_s, 1)$ to a target image keypoint $\mathbf{p}_t = (u_t, v_t, 1)$ is represented as:

$$\mathbf{p}_t = \begin{bmatrix} a_1 & a_2 & b_1 \\ a_3 & a_4 & b_2 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix} \mathbf{p} = \mathbf{J}_t \mathbf{p}. \tag{10}$$

The determination of affine transformation parameters necessitates a minimum of three matches [6]. To account for outliers and ensure reliable parameter estimation, we employ the Random Sample Consensus (RANSAC) algorithm [30].

IV. METHODS

In this section, we introduce a two-step auto-calibration method for the microscope-camera system that enables the determination of intrinsic and hand-eye parameters without the need for microfabricated calibration patterns. Additionally, we present a multi-scale VS method for accurately positioning the microscope and acquiring the target view across increasing magnifications.

A. Auto-Calibration of Optical Microscope

Considering the robot end-effector's movement of x_e, y_e, and θ_{ez} at a specific microscope magnification with an intrinsic matrix of $\mathbf{K}$, the projective homography matrix $\mathbf{J}$ linking the observed images before and after the movement can be formulated as:

$$\mathbf{J} = \mathbf{K}\mathbf{H}\mathbf{K}^{-1}. \tag{11}$$

As outlined in Section III-D, the SIFT and RANSAC algorithms can be employed to determine the projective homography parameter $\mathbf{b}$. Combining (1), (5), (7), (10), and (11), we obtain (12) shown at the bottom of next page.

The objective of calibration is to determine the intrinsic parameters k_x and k_y, as well as the hand-eye parameters η_1, η_2, and φ. One intuitive approach to solve for these parameters involves collecting a series of Δx_e and $\mathbf{b}$ pairs as inputs and outputs, respectively, and then adopting nonlinear optimization algorithms *wrt.* these target parameters. However, the high degree of nonlinearity in (12) can potentially cause numerical instability. To address this concern, we propose a two-step calibration technique that separates these target parameters into two distinct sets, namely, k_x, k_y, and φ, as well as η_1 and η_2. By leveraging linear regression techniques, the optimization problem can be solved with an enhanced level of stability. The calibration approach is detailed as follows:

Step 1: to determine k_x, k_y, and φ. By setting θ_{ez} to zero, which imposes a constraint on the z-axis rotation of the robot end-effector, the variables η_1 and η_2 in (12) can be eliminated. Consequently, (12) can be reformulated as:

$$\begin{bmatrix} x_e & y_e \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} -k_x \sin\varphi & -k_y \cos\varphi \\ -k_x \cos\varphi & k_y \sin\varphi \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ \vdots & \vdots \end{bmatrix}. \tag{13}$$

(13) exhibits a single match, but it is possible to incorporate multiple matches by adding one row to both the first and final matrices for each match. To obtain a solution, a minimum of two matches is required, although additional matches can enhance the accuracy and stability of the estimated parameters.

We have devised a strategy wherein the robot end-effector follows eight pre-defined horizontal translational routes, as depicted in Fig. 4(a), with the displacement vector (x_e, y_e) recorded for each movement. The SIFT and RANSAC algorithms are subsequently employed to determine (b_1, b_2) in each motion step, and the collected data is systematically organized into (13). The traversal routes are chosen in order to achieve uniformly distributed positions and encompass eight distinct moving directions for the robot end-effector. The resulting linear system can be represented as:

$$\mathbf{X}\mathbf{V} = \mathbf{Z}. \tag{14}$$

$\mathbf{V} \in \mathbb{R}^{2\times 2}$ can be estimated using the least-squares method:

$$\widehat{\mathbf{V}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Z}. \tag{15}$$

The parameters k_x, k_y, and φ can then be determined by solving the optimization problem:

$$\begin{aligned} \min_{k_x, k_y, \varphi} \quad & \|\mathbf{V}(k_x, k_y, \varphi) - \widehat{\mathbf{V}}\|_F^2 \\ \text{subject to} \quad & k_x < 0, \quad k_y < 0, \quad |\varphi| \le \frac{\pi}{2}. \end{aligned} \tag{16}$$

While solving (16) analytically is challenging, numerical methods can effectively address this problem. However, inappropriate initial values can lead to numerical instability. To mitigate this issue, we can determine initial values for k_x, k_y, and φ by excluding one value from $\widehat{\mathbf{V}}$ and solving for these parameters using the remaining values. These initial values, while not optimal, are close to the optimal values and can enhance the reliability of the solution.

Step 2: to determine the hand-eye matrix parameters η_1, η_2. After determining the values of k_x, k_y, and φ, the parameters η_1 and η_2 can be derived by setting $x_e = 0$ and $y_e = 0$. This

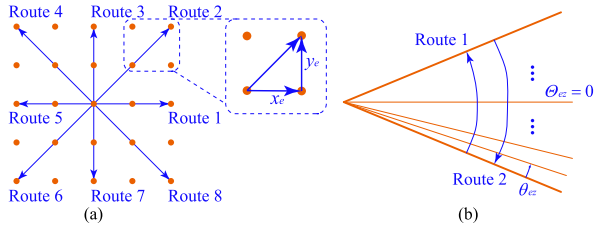

Fig. 4. Traversal routes of the robot end-effector: (a) eight horizontal translational routes and (b) two z-axis rotational routes.

constraint restricts the horizontal movement of the robot end-effector while allowing rotation around the z-axis. As a result, (12) can be reformulated as follows:

$$\mathbf{U}\mathbf{y} = \mathbf{z}, \tag{17}$$

where

$$\mathbf{U} = \begin{bmatrix} k_x[\sin\varphi - \sin(\varphi+\theta_{ez})] & k_x[\cos\varphi - \cos(\varphi+\theta_{ez})] \\ k_y[\cos\varphi - \cos(\varphi+\theta_{ez})] & k_y[\sin(\varphi+\theta_{ez}) - \sin\varphi] \\ \vdots & \vdots \end{bmatrix},$$

$$\mathbf{y} = \begin{bmatrix} \eta_1 & \eta_2 \end{bmatrix}^\top,$$

$$\mathbf{z} = \begin{bmatrix} b_1 + \dfrac{k_x v_0 \sin\theta_{ez}}{k_y} - u_0(1-\cos\theta_{ez}) \\ b_2 - \dfrac{k_y u_0 \sin\theta_{ez}}{k_x} - v_0(1-\cos\theta_{ez}) \\ \vdots \end{bmatrix}.$$

While a single match is sufficient for determining the solution of $\mathbf{y}$, incorporating additional matches can enhance the reliability of the estimated parameters. To this end, the robot end-effector is programmed to rotate around its z-axis, as shown in Fig. 4(b), while recording the corresponding rotational angle θ_{ez}. In each motion step, the SIFT and RANSAC algorithms are used to obtain $\mathbf{b}$, and this data is organized into (17). Using the least-squares method, the parameter $\mathbf{y}$ can be estimated by:

$$\widehat{\mathbf{y}} = \left(\mathbf{U}^\top\mathbf{U}\right)^{-1}\mathbf{U}^\top\mathbf{z}, \tag{18}$$

where $\widehat{\mathbf{y}}$ can be the optimal solution of (η_1, η_2).

B. Multi-Scale VS Control

The objective of multi-scale VS is to control the motion of the microscope using camera image data, so that features of interest in the target view are centered and aligned with the current view across increasing levels of magnification. Algorithm 1 outlines the general process of multi-scale VS. Prior to conducting VS control, it is necessary to compute the intrinsic matrix at the target scale, denoted as $\mathbf{K}_t$. This involves switching to $400\times$ magnification by rotating the revolving nosepiece and performing an autofocusing process to adjust the distance between the specimen and objective lens for optimal image sharpness. $\mathbf{K}_t$ is then determined using the auto-calibration techniques described in Section IV-A.

After determining $\mathbf{K}_t$, VS control is conducted across three different magnifications: $100\times$, $200\times$, and $400\times$. For each

Algorithm 1: Multi-Scale VS Control.

Input: Grayscale target image $\mathbf{I}^*$ at $400\times$
// Calibrate at $400\times$ and obtain $\mathbf{K}_t$
Switch to $400\times$ by rotating the revolving nosepiece;
$Z_t \leftarrow$ Autofocus($400\times$);
$\mathbf{K}_t \leftarrow$ Auto-calibration(Z_t);
// Visual servoing control
for *each magnification* $\in [100\times, 200\times, 400\times]$ **do**
  Switch to $magnification$ by rotating the revolving nosepiece;
  $Z \leftarrow$ Autofocus($magnification$);
  $\mathbf{K}, {}^E\mathbf{T}_C \leftarrow$ Auto-calibration(Z);
  repeat
    Capture grayscale image $\mathbf{I}$ from camera;
    Compute $\mathbf{J}_t$ using SIFT and RANSAC;
    Compute $\widetilde{\mathbf{H}}$: $\begin{bmatrix}\widetilde{\mathbf{R}} & \widetilde{\mathbf{t}} \\ \mathbf{0} & 1\end{bmatrix} \leftarrow \mathbf{K}_t^{-1}\mathbf{J}_t\mathbf{K}$;
    Perform SVD: $\mathbf{U}\mathbf{S}\mathbf{V}^\top \leftarrow \widetilde{\mathbf{R}}$;
    Compute $\mathbf{e}$: $(x, y) \leftarrow \widetilde{\mathbf{t}}$, $\theta \leftarrow \mathbf{U}\mathbf{V}^\top$;
    Compute $\mathbf{v}_c$ using Eq. (22);
    Compute robot end-effector velocity $\mathbf{v}_e$ using Eq. (5) and ${}^E\mathbf{T}_C$ and actuate the robot;
  until $\mathbf{e} < \mathbf{e}_{thr}$;
end

magnification, the revolving nosepiece is rotated to switch to the desired magnification and an autofocusing process is performed to acquire the sharpest image. The intrinsic and hand-eye parameters, represented by $\mathbf{K}$ and ${}^E\mathbf{T}_C$, are determined using auto-calibration techniques. A VS cyclic process is then initiated until positional errors converge to minimal values. Further details on our proposed method are provided below:

1) Autofocusing: Developing a fast and precise method to achieve the sharpest image is the objective of the autofocusing process [5], [31]. In our selected method, we first define a candidate value range for the z-axis position of the microscope, within which an optimal position z_m exists that yields the sharpest image. The robot end-effector is then driven to achieve each z-axis position within the range and the sharpness of the acquired image is computed by convolving it with a Laplacian mask to calculate the second derivative and subsequently averaging the absolute values of the convolution results [32]. The optimal z-axis position z_m corresponding to maximum sharpness is then determined and the robot is moved to this position. The Laplacian mask is chosen for edge detection because it is a well-established and easy-to-implement approach using OpenCV, allowing for quick and reliable measurement of image sharpness.

2) System Calibration: The auto-calibration method is detailed in Section IV-A. The calibration motion steps vary depending on the level of magnification, with smaller steps being used for higher magnifications. The selection of an appropriate calibration motion step is based on ensuring the effective functioning of the SIFT and RANSAC algorithms.

3) VS Control: To achieve alignment between the current and target views, a position-based visual servoing (PBVS) algorithm is utilized. This algorithm is formulated as a minimization problem, where the error $\mathbf{e}(\mathbf{r})$ is defined as follows:

$$\mathbf{e}(\mathbf{r}) = \mathbf{s}(\mathbf{r}) - \mathbf{s}^*, \tag{19}$$

where $\mathbf{s}(\mathbf{r})$ denotes the positional features extracted at the current microscope pose $\mathbf{r}$, while $\mathbf{s}^*$ represents the features at the target microscope pose $\mathbf{r}^*$.

Our PBVS scheme is designed by using $\mathbf{s} = (x, y, \theta) \in \mathbb{R}^3$, where x, y, and θ are defined in (7). In this case, we have $\mathbf{s}^* = \mathbf{0}$,

$$\begin{cases} b_1 = u_0(1-\cos\theta_{ez}) - k_x\left[\dfrac{v_0\sin\theta_{ez}}{k_y} + (-\eta_2 + y_e)\cos\varphi + \eta_2\cos(\varphi+\theta_{ez}) + (-\eta_1 + x_e)\sin\varphi + \eta_1\sin(\varphi+\theta_{ez})\right] \\ b_2 = v_0(1-\cos\theta_{ez}) + k_y\left[\dfrac{u_0\sin\theta_{ez}}{k_x} + (\eta_1 - x_e)\cos\varphi - \eta_1\cos(\varphi+\theta_{ez}) + (-\eta_2 + y_e)\sin\varphi + \eta_2\sin(\varphi+\theta_{ez})\right] \end{cases}. \tag{12}$$

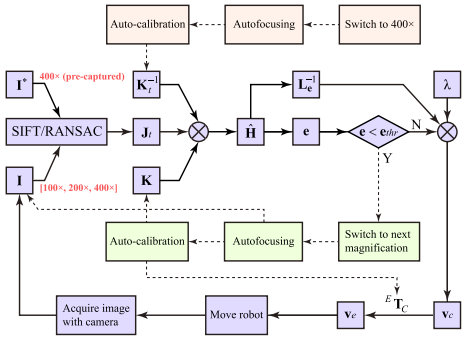

Fig. 5. Multi-Scale VS control architecture: The procedure begins with the preparation of $\mathbf{K}_t$ via autofocusing and auto-calibration at $400\times$ magnification. VS control is then executed across $100\times$, $200\times$, and $400\times$ magnifications to align the current view $\mathbf{I}$ with the pre-captured $400\times$ target view $\mathbf{I}^*$. Each magnification level involves initial autofocusing and auto-calibration to acquire $\mathbf{K}$ and ${}^E\mathbf{T}_C$. Subsequently, SIFT and RANSAC algorithms are utilized to compute $\mathbf{J}_t$ of $\mathbf{I}$ *wrt.* $\mathbf{I}^*$. This is followed by the computation of $\widehat{\mathbf{H}}$, which aids in obtaining $\mathbf{e}$ and $\mathbf{L}_\mathbf{e}$, eventually leading to the computation of $\mathbf{v}_c$. The microscope is actuated with a velocity of $\mathbf{v}_c$ by inferring the robot end-effector velocity $\mathbf{v}_e$ with ${}^E\mathbf{T}_C$. The VS control escalates to a higher magnification when $\mathbf{e}$ falls within a threshold $\mathbf{e}_{thr}$.

$\mathbf{e} = \mathbf{s}$, and the interaction matrix $\mathbf{L}_\mathbf{e}$ related to $\mathbf{e}$:

$$\mathbf{L}_\mathbf{e} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{20}$$

which establishes a relationship between the microscope velocity $\mathbf{v}_c$ and the temporal variation of $\mathbf{e}$:

$$\dot{\mathbf{e}} = \mathbf{L}_\mathbf{e}\mathbf{v}_c, \tag{21}$$

where $\mathbf{v}_c = (v_{cx}, v_{cy}, \omega_z) \in \mathbb{R}^3$. v_{cx} and v_{cy} denote the microscope translational velocity along its x and y axes, respectively. ω_z represents the microscope rotational velocity around its z-axis. To achieve an exponential decoupled decrease of $\mathbf{e}$ (i.e., $\dot{\mathbf{e}} = -\lambda\mathbf{e}$), we derive the VS control law as:

$$\mathbf{v}_c = -\lambda {\mathbf{L}_\mathbf{e}}^{-1}\mathbf{e}, \tag{22}$$

where $\lambda \in \mathbb{R}$ is a gain parameter, which directly influences the response speed and stability of the system. A higher value of λ results in faster convergence of the error to zero, leading to a more rapid system response. However, a higher λ may also cause undesired oscillations or even instability in the system due to the amplification of feature noises. Hence, it is crucial to carefully tune λ to achieve a balance between system response speed and stability. The architecture of our PBVS scheme is shown in Fig. 5. The control law operates within a closed-loop system, where $\mathbf{v}_c$ is computed based on each new image acquired by the camera. The iteration process proceeds until $\mathbf{e}$ falls below a predetermined threshold $\mathbf{e}_{thr} \in \mathbb{R}^3$, indicating that the microscope has reached its desired position. In cases where $\mathbf{e}$ exceeds the threshold $\mathbf{e}_{thr}$, we employ the SIFT and RANSAC algorithms to compute $\mathbf{J}_t$ that links the current image to the target image. We subsequently estimate the Euclidean homography matrix $\widehat{\mathbf{H}}$ using step 13 in Algorithm 1. Due to disturbances and errors, $\widetilde{\mathbf{R}} \in \mathbb{R}^{2\times2}$ is not a perfect orthogonal matrix. However, we can determine an optimal orthogonal matrix $\widehat{\mathbf{R}} \in \mathbb{SO}(2)$ that approximates $\widetilde{\mathbf{R}}$ by solving the optimization problem as follows:

$$\min_{\widehat{\mathbf{R}}} \|\widehat{\mathbf{R}} - \widetilde{\mathbf{R}}\|_F^2 \quad \text{subject to } \widehat{\mathbf{R}}^\top\widehat{\mathbf{R}} = \mathbf{I}. \tag{23}$$

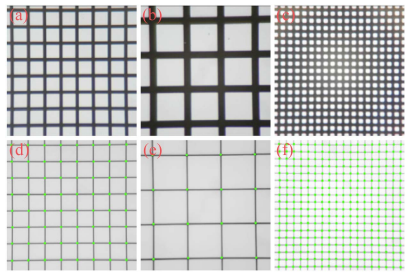

Fig. 6. Calibration patterns and extracted feature points: (a)–(c) show calibration patterns with $100 \times 100\,\mu\text{m}^2$, $100 \times 100\,\mu\text{m}^2$, and $10 \times 10\,\mu\text{m}^2$ grids observed at $100\times$, $200\times$, and $400\times$ magnifications, respectively. (d)–(f) display extracted feature points after image dilation and Harris corner detection, corresponding to (a)–(c), respectively. Feature points are matched with their actual positions to obtain intrinsic and hand-eye parameters [6], [33]. Calibrating each magnification requires three distinct pattern images captured at different poses, which is complex and time-consuming.

The analytical solution to the problem has been proven to be $\widehat{\mathbf{R}} = \mathbf{U}\mathbf{V}^\top$, provided that the singular value decomposition (SVD) of $\widetilde{\mathbf{R}}$ is $\mathbf{U}\mathbf{S}\mathbf{V}^\top$ [6]. Thus, we can determine $\mathbf{e}$ using $\widetilde{\mathbf{t}}$ and $\widehat{\mathbf{R}}$, and subsequently compute the microscope velocity $\mathbf{v}_c$ using (22). Furthermore, employing (5) and ${}^E\mathbf{T}_C$ allows us to infer the desired velocity of the robot end-effector $\mathbf{v}_e$. The robot is continuously actuated towards the target position until $\mathbf{e}$ falls within the threshold $\mathbf{e}_{thr}$.

V. Experimental Assessments

In this section, we assessed the efficacy of our proposed calibration and VS methods using a porcine pancreatic section. The experimental setup is presented in our supplementary video. The results and evaluations are presented as follows:

A. Precision and Stability of Auto-Calibration

The first experiment aimed to establish the precision and stability of our auto-calibration method. We followed the calibration procedure outlined in Section IV-A and computed the intrinsic and hand-eye parameters for objective lenses with different magnifications: $100\times$, $200\times$, and $400\times$. The horizontal moving step for both x_e and y_e was assigned to be $100\,\mu\text{m}$, $50\,\mu\text{m}$, and $20\,\mu\text{m}$, respectively, while the rotational step for θ_{ez} was set to be $0.05°$, $0.02°$, and $0.01°$, correspondingly. A total of 50 rounds of calibration were executed, and the statistical results of the calibrated parameters are presented in Table II. The standard deviations of these parameters are found to be minuscule, affirming the stability of our proposed calibration method. Despite the standard deviation of η_2 being similar to its mean, it doesn't imply unreliability as the designated value for η_2 is very close to zero. To validate the precision of our calibration method, we provided a control group relying on calibration patterns to compute the intrinsic and hand-eye parameters. The details on calibration patterns are presented in Fig. 6, and the calibration results are provided in Table II. It is evident that both the intrinsic and hand-eye parameters are highly consistent for the two calibration methods.

Calibrated parameter discrepancies are likely due to fabrication and assembly errors. For instance, variations in the hand-eye parameters may arise from errors in the fabrication and assembly of the nosepiece. The intrinsic parameters may also be affected, as these parameters are coupled in the calibration process.

TABLE II
PARAMETER STATISTICAL RESULTS AFTER 50 ROUNDS OF CALIBRATION

Methods	Magnifications	k_x	k_y	φ (deg)	η_1 (mm)	η_2 (mm)
Zhang's [6] & Tsai's [33]	$\times100$	-447636	-450301	0.4589	125.1713	0.3845
	$\times200$	-935292	-941844	0.4162	125.1291	0.4235
	$\times400$	-1949773	-1961684	0.4943	124.7092	0.0984
Ours	$\times100$	-448672 ± 367	-451767 ± 414	0.4665 ± 0.0198	125.4031 ± 0.1072	0.0663 ± 0.0511
	$\times200$	-938325 ± 1351	-942616 ± 822	0.3868 ± 0.0401	125.1043 ± 0.1570	0.6236 ± 0.1355
	$\times400$	-1953155 ± 3047	-1981235 ± 2908	0.5722 ± 0.0597	124.9100 ± 0.2071	0.1166 ± 0.1683
Convergence range	$\times100$	$-4.76e5 \sim -4.22e5$	$-4.79e5 \sim -4.25e5$	$-8.9 \sim 8.9$	$116.2 \sim 134.4$	$-9.1 \sim 9.1$
	$\times200$	$-9.95e5 \sim -8.82e5$	$-9.99e5 \sim -8.86e5$	$-4.1 \sim 4.1$	$120.7 \sim 129.5$	$-4.4 \sim 4.4$
	$\times400$	$-2.07e6 \sim -1.84e6$	$-2.10e6 \sim -1.86e6$	$-1.9 \sim 1.9$	$122.7 \sim 126.9$	$-2.1 \sim 2.1$

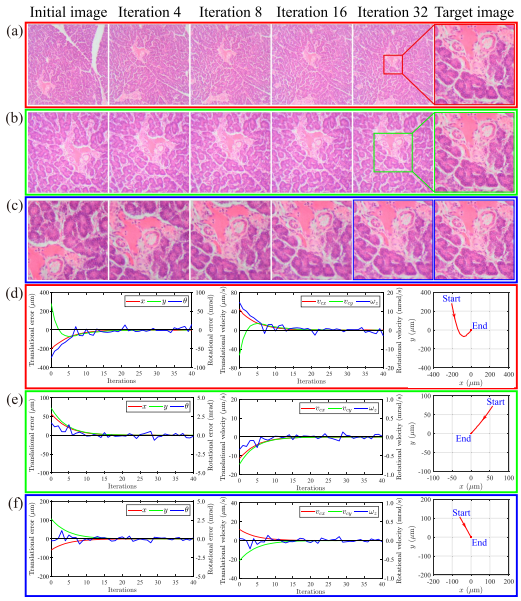

Fig. 7. Multi-Scale VS: (a)–(c) depict experimental images at magnification levels of (a) $100\times$, (b) $200\times$, and (c) $400\times$. The first column displays the initial images $\mathbf{I}$. The second to fifth columns show the images at iteration steps of 4, 8, 16, and 32, respectively. The last column presents the target image $\mathbf{I}^*$ at $400\times$. Variations of $\mathbf{e}$ and $\mathbf{v}_c$ *wrt.* iteration steps, as well as the microscope trajectories, at different magnification levels are depicted in (d)–(f) for (d) $100\times$, (e) $200\times$, and (f) $400\times$.

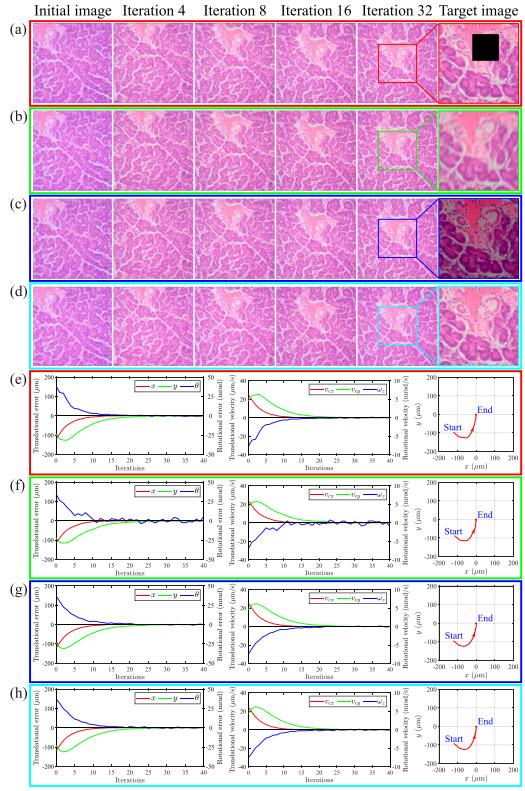

Fig. 8. Robustness evaluation of VS: (a)–(c) depict experimental images with various types of target image disturbance: (a) a target image with a $120 \times 120\,\text{pixel}^2$ block mask, (b) a target image blurred by fast movement or inaccurate autofocusing, and (c) a target image with subtracted intensity of 80. (d) displays a disturbance-free target image as a control group. The first column displays the initial images $\mathbf{I}$. The second to fifth columns show the images at iteration steps of 4, 8, 16, and 32, respectively. The last column presents the target images $\mathbf{I}^*$ at $400\times$. Variations of $\mathbf{e}$ and $\mathbf{v}_c$ *wrt.* iteration steps, as well as the microscope trajectories, are depicted in (e)–(h), corresponding to (a)–(d), respectively.

Additionally, vibrations during experimentation may introduce image blur, which can affect the precision of our calibration results. To assess the impact of these errors, we have provided a convergence range for all parameters in Table II. Within this range, the VS control can effectively align the source and target images. These ranges were determined through experimentation. Our calibration methods have proven to be precise enough for stable VS control, as the minor errors in calibration parameters do not cause our system to fall outside of the convergence range.

B. Assessments of Multi-Scale VS

In the second experiment, we evaluated the performance of the multi-scale VS control outlined in Algorithm 1. The experiment involved using a pre-captured target image at $400\times$ magnification, and consecutively aligning the current views at $100\times$, $200\times$, and $400\times$ with it. A gain parameter of $\lambda = 0.2$ was applied uniformly across all magnifications. This value was experimentally determined to provide a satisfactory balance between system response speed and stability. The process and data were recorded during the experiment and are presented in Fig. 7, where Fig. 7(a)–(c) show the experimental images captured during the VS procedure. It is observed that the initial images $\mathbf{I}$ gradually aligned with the target image $\mathbf{I}^*$ and eventually features of interest in the target image were centered in the current views. Fig. 7(d)–(f) depict the variations of $\mathbf{e}$ and $\mathbf{v}_c$ *wrt.* iteration steps, as well as the microscope trajectories. These plots demonstrate the exponential decrease of both $\mathbf{e}$ and $\mathbf{v}_c$, indicating successful completion of the VS control. To further verify the effectiveness of our multi-scale VS method, we conducted experiments on additional specimens, including a frog liver section and a mesophytic root section, with results presented in our supplementary video. We also performed repetitive repositioning tasks on three specimens: porcine pancreatic, frog liver, and mesophytic root sections, each tested five times. Autofocusing and auto-calibration procedures were omitted after the first test on each specimen due to minimal environmental changes. The multi-scale VS method averaged a repositioning time of 30.7 s, significantly less than the 78.2 s required for

manual repositioning at $400\times$ magnification. This demonstrates the enhanced efficiency of histologic biopsies using our method.

C. Robustness of VS

In the third experiment, we evaluated the robustness of our VS method against external disturbances. The experiment followed the procedures outlined in Algorithm 1 at a magnification of $200\times$. The outcomes of this robustness assessment are presented in Fig. 8, where Fig. 8(a)–(c) exhibit experimental images featuring various types of disturbances. The first examination involved masking out a 120×120 pixel^2 rectangular area from the target image, to simulate situations where occlusions are present in either the target or current image. The second examination incorporated a blurred target image to simulate the conditions of rapid movement or inaccurate autofocusing. In the third examination, we decreased the target image intensity by 80, to mimic illumination variations between the target and current images. Fig. 8(d) presents a control group devoid of any disturbances. The corresponding $\mathbf{e}$ and $\mathbf{v}_c$, along with the microscope trajectories, are presented in Fig. 8(e)–(h), showing that both $\mathbf{e}$ and $\mathbf{v}_c$ exhibited an exponential decrease until achieving a favorable alignment between the target and current images. Notably, Fig. 8(f) exhibits higher fluctuations in $\mathbf{e}$ and $\mathbf{v}_c$ due to the impact of blur on the image gradient distribution, resulting in errors in the affine transformation estimation. In contrast, a substantial portion of gradient distribution information was retained in the presence of occlusions and illumination variations. As a result, the affine transformation can still be reliably estimated using the RANSAC algorithm. Overall, these experimental results reveal that our VS method exhibits remarkable robustness against disturbances.

VI. Conclusion

This letter has presented several innovative contributions to the advancement of active microscope-camera systems. Specifically, we have developed an auto-calibration method that accurately determines the intrinsic and hand-eye parameters of the microscope-camera system across different magnifications, providing stable and calibration pattern-free results. Moreover, we have introduced a multi-scale VS method that uses a single high-magnification target view as a reference, enabling rapid repositioning of the microscope across various magnifications. Experimental evaluations demonstrate the precision and stability of our auto-calibration method and the robustness of our multi-scale VS approach against occlusion, blur, and low illumination. These findings show significant potential for applications in microscopy and related fields.

References

[1] B. Dahroug, B. Tamadazte, and N. Andreff, "PCA-based visual servoing using optical coherence tomography," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 3430–3437, Apr. 2020.

[2] F. Chaumette and S. Hutchinson, "Visual servo control. I. basic approaches," *IEEE Robot. Automat. Mag.*, vol. 13, no. 4, pp. 82–90, Dec. 2006.

[3] S. Sarabandi, J. M. Porta, and F. Thomas, "Hand-eye calibration made easy through a closed-form two-stage method," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 3679–3686, Apr. 2022.

[4] J. Jiang, X. Luo, Q. Luo, L. Qiao, and M. Li, "An overview of hand-eye calibration," *Int. J. Adv. Manuf. Technol.*, vol. 119, no. 1-2, pp. 77–97, 2022.

[5] X. Sha, H. Sun, Y. Zhao, W. Li, and W. J. Li, "A review on microscopic visual servoing for micromanipulation systems: Applications in micromanufacturing, biological injection, and nanosensor assembly," *Micromachines*, vol. 10, no. 12, 2019, Art. no. 843.

[6] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[7] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Automat.*, vol. 3, no. 4, pp. 323–344, Aug. 1987.

[8] Y. Zhou and B. J. Nelson, "Calibration of a parametric model of an optical microscope," *Opt. Eng.*, vol. 38, no. 12, pp. 1989–1995, 1999.

[9] M. Ammi, V. Fremont, and A. Ferreira, "Automatic camera-based microscope calibration for a telemicromanipulation system using a virtual pattern," *IEEE Trans. Robot.*, vol. 25, no. 1, pp. 184–191, Feb. 2009.

[10] C. Collewet and E. Marchand, "Photometric visual servoing," *IEEE Trans. Robot.*, vol. 27, no. 4, pp. 828–834, Aug. 2011.

[11] E. Marchand, "Subspace-based direct visual servoing," *IEEE Robot. Automat. Lett.*, vol. 4, no. 3, pp. 2699–2706, Jul. 2019.

[12] E. Marchand, "Direct visual servoing in the frequency domain," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 620–627, Apr. 2020.

[13] S. Felton, P. Brault, E. Fromont, and E. Marchand, "Visual servoing in autoencoder latent space," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 3234–3241, Apr. 2022.

[14] S. Felton, E. Fromont, and E. Marchand, "Deep metric learning for visual servoing: When pose and image meet in latent space," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 741–747.

[15] T. Li, Y. Yang, X. Fu, Z. Sun, Y. F. Li, and S. Liu, "Visual servo control for workspace navigation of nanorobot end-effector inside SEM," *IEEE Trans. Automat. Sci. Eng.*, early access, Jun. 05, 2023, doi: 10.1109/TASE.2023.3279895.

[16] Y. Yang, T. Li, X. Fu, Z. Sun, Y. F. Li, and S. Liu, "Zooming-free hand-eye self-calibration for nanorobotic manipulation inside SEM," *IEEE Trans. Nanotechnol.*, vol. 22, pp. 291–300, 2023.

[17] Z. Gong, B. Tao, H. Yang, Z. Yin, and H. Ding, "An uncalibrated visual servo method based on projective homography," *IEEE Trans. Automat. Sci. Eng.*, vol. 15, no. 2, pp. 806–817, Apr. 2018.

[18] K. Pachtrachai, F. Vasconcelos, G. Dwyer, V. Pawar, S. Hailes, and D. Stoyanov, "CHESS–calibrating the hand-eye matrix with screw constraints and synchronization," *IEEE Robot. Automat. Lett.*, vol. 3, no. 3, pp. 2000–2007, Jul. 2018.

[19] H. Chen et al., "ASpanFormer: Detector-free image matching with adaptive span transformer," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 20–36.

[20] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "LoFTR: Detector-free local feature matching with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8922–8931.

[21] M. G. Forero et al., "Comparative analysis of detectors and feature descriptors for multispectral image matching in rice crops," *Plants*, vol. 10, no. 9, 2021, Art. no. 1791.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.

[23] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *Proc. Brit. Mach. Vis. Conf.*, 2013, pp. 13.1–13.11.

[24] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571.

[25] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. 9th Eur. Conf. Comput. Vis.*, 2006, pp. 430–443.

[26] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2548–2555.

[27] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4938–4947.

[28] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 224–236.

[29] J. Edstedt, I. Athanasiadis, M. Wadenbäck, and M. Felsberg, "DKM: Dense kernelized feature matching for geometry estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 17765–17775.

[30] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[31] Y. Sun, S. Duthaler, and B. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," *Microsc. Res. Technique*, vol. 65, pp. 139–49, 2004.

[32] J. L. Pech Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia, "Diatom autofocusing in brightfield microscopy: A comparative study," in *Proc. IEEE 15th Int. Conf. Pattern Recognit.*, 2000, pp. 314–317.

[33] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Trans. Robot. Automat.*, vol. 5, no. 3, pp. 345–358, Jun. 1989.