Visual Servoing-Based Depth Estimation Technique for Manipulation inside SEM.

Naresh Marturi, Brahim Tamadazte, Sounkalo Dembélé, Nadine Piat

Visual Servoing-Based Depth Estimation Technique for Manipulation inside SEM

Naresh Marturi , *Member, IEEE*, Brahim Tamadazte, *Member, IEEE*, Sounkalo Dembélé, *Member, IEEE* and Nadine Piat

***Abstract*—Depth estimation for micro-nanomanipulation inside a scanning electron microscope (SEM) is always a major concern. So far in the literature, various methods have been proposed based on stereoscopic imaging. Most of them require external hardware unit or manual interaction during the process. In this paper, solely relying on image sharpness information, we present a new technique to estimate the depth in real-time. To improve the accuracy as well as the rapidity of the method, we consider that both autofocus and depth estimation as visual servoing paradigms. The major flexibility of the method lies in its ability to compute the focus position and the depth using only the acquired image information i.e., sharpness. The feasibility of the method is shown by performing various ground truth experiments: autofocus achievements, depth estimation, focus-based nanomanipulator depth control and sample topographic estimation at different scenarios inside the vacuum chamber of a tungsten gun SEM. The obtained results demonstrate the accuracy, rapidity and efficiency of the developed method.**

***Index Terms*—Depth estimation, scanning electron microscope, autofocus, micro-nanomanipulation.**

I. Introduction

Over the last decade, SEM has been emerged as a useful vision sensor in performing robotic manipulation and characterization of micro/nano-objects [1–3]. Its main advantage lies in its ability to produce images with high resolution (better than 1 nanometer), broad range of magnification (up to $\times$ 2,000,000) and high depth-of-focus (DOF). Since the early stages of research using SEM, it has been a great interest for many researchers to extract the depth information from the images to use it in real-time vision-based control applications. Even though it is possible to obtain in-plane information (along X and Y) from the images using *visual servoing* strategies [4], estimating the depth information along Z is more challenging due to the existence of parallel projection [5]. This arise the difficulty in using regular pose estimation methods, well-known in the computer vision community.

By far, stereoscopic imaging-based techniques are widely used for depth estimation and 3D reconstruction of a sample in SEM [6–8]. Since only a single imaging sensor is available, most of them tried to acquire stereoscopic images by tilting the sample concentrically with a specific angle. However, the major challenge is to determine accurate tilt angle and to find correspondences between the acquired noisy and low-textured gray scale SEM images. Alternatively a focused ion beam (FIB) system can be used along with SEM to acquire stereo-images [7]. However, the addition of FIB increases the overall system cost (a FIB system costs about US $ 0.2M). Besides, tilting a sample is not a feasible option when performing automatic tasks like probing, gluing etc. In [9], a touchdown sensor has been used to detect the contact between two carbon nanotubes (CNTs) and simultaneously computing the fine depth. In [1], a sliding-based technique was used to perform Z positioning and in [10], a shadow-based depth estimation technique was used to align the microgrippers along optical axis for handling CNTs.

Apart from stereoscopic methods, depth can be estimated using image sharpness information. The underlying idea is to find the *in-focus* image (with maximum sharpness) among the sequence of images acquired along the optical axis. Depth is then computed by finding the relative camera position that delivers the *in-focus* image. These types of methods are commonly seen in optical microscopy where the DOF is small when compared to the SEM at specific magnification [11]. Using SEM, the manipulation tasks are often performed at higher magnifications ($> 1000\times$) where the DOF is small enough to make depth measurements [9,12,13]. Besides, the computational speed and simplicity makes the focus-based methods more suitable for fast depth measurements during manipulation tasks. However, the only limitation is that the resolution of depth estimation is limited to use them for fine measurements. The relative accuracy using these methods has been reported to be in the range of $10\ \mu m$ [9], where a focus-based method was used for manipulating CNTs. Wich *et al.* implemented a focus-based method for Z-axis coarse adjustments during a nanoassembly task [14]. Ru and To used a similar approach for contact detection and manipulating the nanospheres [12]. In most of these works, variance was used to estimate the image sharpness and depth was computed by scanning the pre-fixed focus range step by step. However, for SEM with a high range of focus (≈ 5000 focus steps), recording the sharpness at each focus step is a time-consuming process. Hence, for fast and dynamic depth estimation, a reliable and accurate autofocus is indispensable.

Usually with SEM, passive autofocus methods are considered where the optimum focus is determined by analyzing the recorded images. So far in the literature, a variety of autofocus techniques are evaluated and compared for different types of microscopies [15, 16]. In case of SEM, to improve the rapidity of autofocusing process, few works have developed various iterative techniques to search for the *in-focus* objective lens position [17]. Even though these methods are quite effective, they are highly dependent on the search history that makes them to suffer from the non-linear hysteresis shown by the electromagnetic lenses. Apart from them, predictive and learning-based autofocus algorithms (mostly used with digital and mobile phone cameras) are also investigated to use with SEM [18]. These methods mainly require a large image datasets obtained

This work is conducted with a financial support from the project NANOROBUST (ANR-11-NANO-006) funded by the Agence Nationale de la Recherche (ANR), France. It is also performed in the framework of the Labex ACTION (ANR-11-LABEX-01-01) and the Equipex ROBOTEX (ANR-10-EQPX-44-01) projects.

The authors are with the Department of Automatic Control and MicroMechatronic Systems (AS2M), FEMTO-ST Institute, 25000 Besançon, France. Corresponding author e-mail: mvnnaresh@gmail.com.

along the optical axis at different conditions to train the system. Moreover, the computational burden will be high to use them for dynamic problems. In addition to these techniques, most of the modern day SEM systems come with an integrated autofocusing module. However, the commercial methods are more intended for manual measurements and are difficult to integrate with automatic applications.

In this work, we propose a new *visual servoing*-based technique to perform both autofocus and depth-estimation for complementing automatic manipulation tasks inside a SEM. The iterative search-based methods that are well-known for SEM autofocus mainly lack in providing enough accuracy, speed and flexibility to use them for dynamic depth estimation tasks. The major problem arises from their dependency on the search history. The developed approach successfully tackles this issue by directly reaching the *in-focus* position without any prior information and by using only the acquired image sharpness. The major flexibility of the method lies in its ability to compute the focus and the depth using only the acquired image information and does not require any external operations like sample or beam tilting. The major scientific contributions include the development of new methodologies for autofocus and dynamic depth estimation via mathematical modelling as visual servoing strategies. More precisely, for autofocus, the Z displacements are linked to the image sharpness through an interaction matrix and the velocity progress is related to the sharpness cost-function gradient. Later this has been extended to estimate the depth, where the entire process is decomposed into two different phases to increase the overall speed. The technical contributions include the extension of the proposed depth estimation for various SEM related problems such as depth control of a nanomanipulator, sample topography and orientation estimation. Indeed, our technique in comparison to the conventional methods allows: (i) an improved accuracy by 44%, (ii) an increased rapidity by a rate of 290%. As this paper deals only with depth estimation and control, we will not speak about full manipulation task that also requires planar positioning control as shown before in [4].

II. Visual Servoing-based Autofocus and Depth Estimation

SEM images are formed by raster scanning the sample surface by means of a focused electron beam. As the device focus is directly connected with the electronic working distance, we make use of this property in measuring the sample depth. Our process starts by estimating image sharpness information. Later, we use this information in designing the visual control law to tune the device focus automatically, which is then used to estimate the depth.

A. Focusing geometry in SEM

In SEM, the two sets of lenses that contribute towards the focusing process are the condenser lenses lying above the aperture and the objective lenses lying below the aperture (see Fig. 1(a)). The condenser lenses control the final spot size and beam diameter whereas the objective lenses focus the spot sized electron beam on to sample surface. The aperture that is present in between these two lens filters out the low energy and non-directional electrons. Total focusing process is illustrated in Fig. 1(b). Coarse focusing is performed by adjusting the electronic working distance W that is attained by varying the current passing through the objective lenses. This is the distance measured in between the objective lens pole piece and the focusing plane. At a distance $\frac{D}{2}$ on both sides of the focus plane for a specific magnification, the beam diameter is two times the pixel diameter and results in blur image. Within the distance D (depth of focus), the image appears to be acceptably in focus.

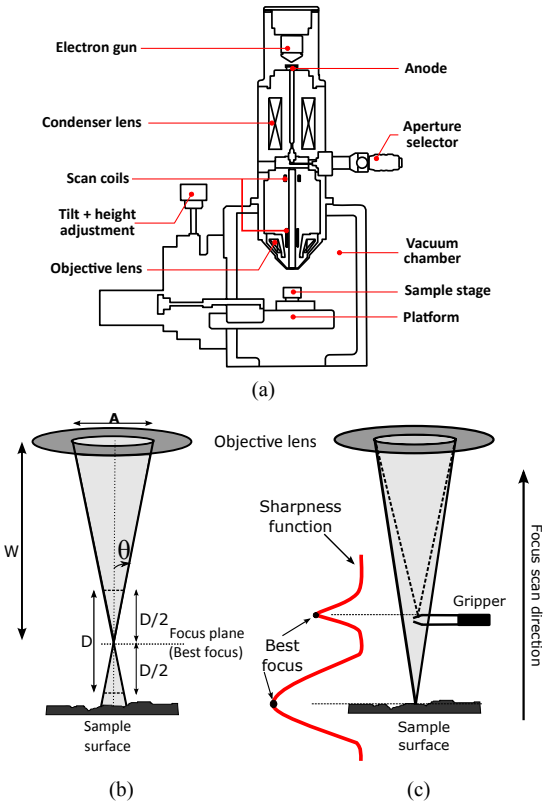


Fig. 1. (a) Cross section diagram of a conventional SEM showing various vital components. (b) Geometry of focusing (c) evolution of a sharpness measure.

B. Image Sharpness Estimation

Numerous efficient sharpness estimators are available in the literature [16]. When applied on the images acquired along the optical axis, maximum score of the function will be obtained for the image on the focal plane (see Fig.1(c)). In our previous work [19], we have evaluated the most commonly used sharpness functions and found that the normalized variance provides good compromise between speed and accuracy. It estimates the sharpness score by computing the gray level intensity variations among the pixels. For a given image $\mathbf{I}_{M\times N}$ with M and N are its width and height, $i(u,v)$ is the pixel intensity at (u,v) and and μ is the pixel mean intensity, the normalized variance $\mathbf{s}(\mathbf{I})$ is given by

$$\mathbf{s}(\mathbf{I}) = \frac{1}{MN}\frac{1}{\mu}\sum_{u=1}^{M}\sum_{v=1}^{N}(i(u,v)-\mu)^2. \qquad (1)$$

C. Visual Servoing-Based Autofocus

In general, *visual servoing* is a closed-loop control mechanism to control any robotic device using image features. It consists of designing a control law to regulate the error obtained from the measured (current and desired) visual features to zero. Using this concept, in this work an autofocusing algorithm has been developed. The principle goal of the method is to find the maximum of sharpness function by controlling the system focus parameters. It follows:

- instead of minimizing the error function, the proposed method maximizes the sharpness function given by (1);
- rather than using local image measurements, the global image information is used;
- most importantly, no reference image or features are used; instead, the method converges to a best focus position i.e., when the sharpness function reaches its maximum value.

The focusing process with the SEM may be represented as a plant whose output is the image sharpness score $\mathbf{s}$, that is always steered by the input focus step F (i.e., position-based control using a step displacement). Physically, this process corresponds to the modification of current passing through the objective lens that changes the electronic working distance. In general cases, if the scene contains a single planar sample, the output sharpness curve contains a single peak and can be easily approximated with a Gaussian function. However, if we consider multiple objects and the system dynamics, this approximation is not valid and in such a case, the plant can be modeled by a non-linear first order equation given by (2)

$$\gamma(F) = a\dot{\mathbf{s}} + \mathbf{s} \tag{2}$$

where, a is the operational time constant and $\gamma(F)$ is the non-linear function (dotted sharpness curve shown in Fig. 2).

By analyzing the sharpness curve for a series of focus steps, it can be observed that for a specific range of focus steps the images contain some details and outside this range, the images are plain with varying noise. From this analysis, it is clear that the primary task is to drive the process near the focus range. Generally, in this case, the usual derivative-based visual control strategies often fail due to the high amount of noise, adequate image details and presence of local minima. In order to tackle this problem, an adaptive control law has been designed, especially to remove the predefined focus step displacement and to avoid the local minima. The corresponding vision-based objective function $\Psi(\mathbf{s})$ is given by

$$\Psi(\mathbf{s}) = C(\mathbf{s}) \left(\frac{\dot{\mathbf{s}}}{\parallel \dot{\mathbf{s}} \parallel} \right) = C(\mathbf{s}) \cdot Sign(\dot{\mathbf{s}}) \tag{3}$$

where, $C(\mathbf{s})$ is a cost-function linked to the image sharpness and the unit vector $\left(\frac{\dot{\mathbf{s}}}{\|\dot{\mathbf{s}}\|}\right)$ provides the driving direction with $\dot{\mathbf{s}} = \frac{\partial \mathbf{s}}{\partial F}$. Two different candidates are evaluated to frame the cost-function: quadratic cost (4) and exponential cost (5) (see Fig. 2(a) and Fig. 2(b), respectively).

$$C_q(\mathbf{s}) = \alpha \left(\frac{\mathbf{s}_0}{\mathbf{s}} \right)^2 \tag{4}$$

$$C_e(\mathbf{s}) = \alpha e^{-\left(\frac{\mathbf{s}-\mathbf{s}_0}{\mathbf{s}_0}\right)} \tag{5}$$

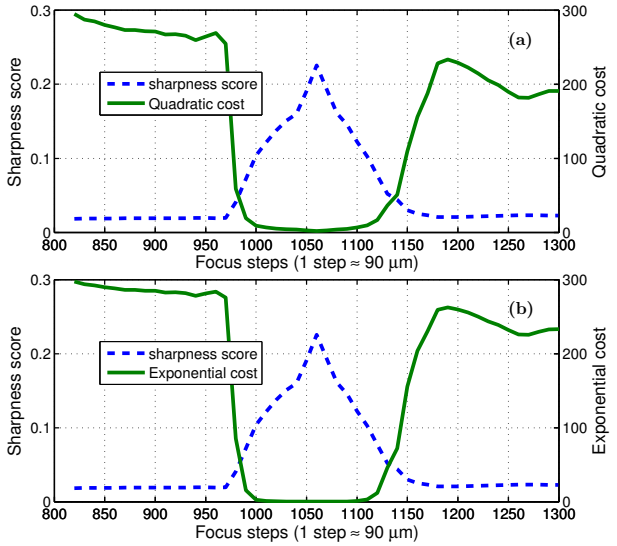


Fig. 2. Cost variation with sharpness (a) quadratic cost (b) exponential cost.

where, α is a positive gain and $\mathbf{s}_0$ is the initial image sharpness score. From analysis, it can be seen that the quadratic cost suits better than the exponential one. Even though the latter shown better convergence, it reaches near zero around 990^{th} step i.e., 87 focus steps before the best focus. Moreover, to have an effect on the focus control the cost should be ≥ 1 at any position. Hence, exponential cost will not have any effect on the process from 990^{th} focus step. The quadratic cost function is designed such that the cost is maximum at a region far from the best focus and is minimum (almost one) near best focus. $\alpha = 300$ was used for cost evaluation. Varying α modifies the process speed.

Using (3) and (4), the primary control law to drive and control the focus step displacement is given by

$$F_{k+1} = F_k + \Psi(\mathbf{s}) = F_k + \alpha \left(\frac{\mathbf{s}_0}{\mathbf{s}} \right)^2 \left(\frac{\dot{\mathbf{s}}}{\parallel \dot{\mathbf{s}} \parallel} \right) \tag{6}$$

where, F_k and F_{k+1} are the current and next focus steps, respectively.

The stability of the designed control law can be discussed by considering the following state variable $Q = [\mathbf{s}\ F]^\top$, that leads in the two non-linear state equations given by (2) and (3). The equilibrium point to be considered is $Q^* = [\mathbf{s}^*\ F^*]^\top$. Here, $\mathbf{s}^*$ corresponds to the maximum of sharpness i.e., $\mathbf{s}_{max}$ and F^* corresponds to the focus step at which sharpness is maximum. The objective is to study the system stability using *Lyapunov theory*. Thus, the state equations (2) and (3) are linearized at the equilibrium point. For this purpose we will consider the new equilibrium point as

$$q^* = Q - Q^* \Leftrightarrow [\varsigma\ f]^\top = [\mathbf{s} - \mathbf{s}^*\ F - F^*]^\top. \tag{7}$$

Let us consider the first state equation given by (2). By definition, the first derivative of $\gamma(f)$ with respect to f at q^* is zero. As a consequence, the linearized first state equation becomes $a\dot{\varsigma} + \varsigma = 0$.

Now, let us consider the second state equation given by (3). At the equilibrium point q^*, the derivative of $C(\varsigma)$ with respect to ς is $b = -2\alpha \left(\frac{\mathbf{s}_0}{\mathbf{s}_{max}} \right)$. This term is negative i.e., $b \in]-\infty\ 0[$.

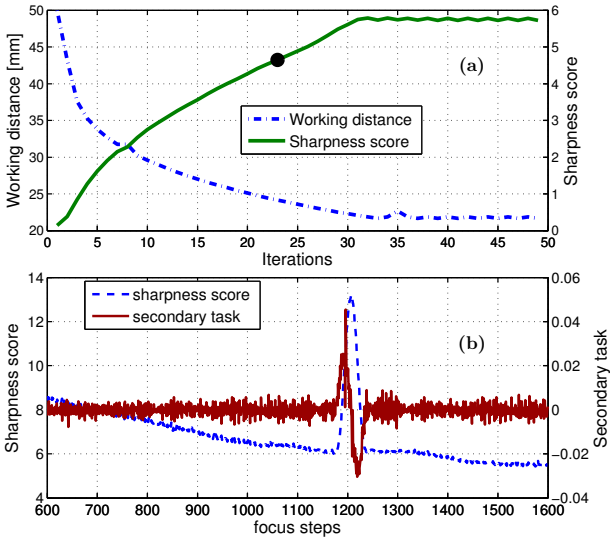


Fig. 3. (a) Evolution of the primary control and (b) relation between secondary function and sharpness at $1000\times$ magnification.

Then the second state equation becomes $\dot{f} = bSign(\dot{\varsigma})\varsigma$. Finally, both the new state equations lead to the following state matrix $\mathbf{\Phi}$ given by

$$\mathbf{\Phi} = \begin{bmatrix} \frac{-1}{a} & 0 \\ bSign(\dot{\varsigma}) & 0 \end{bmatrix} \tag{8}$$

The eigenvalues of this matrix are negative $\left(\frac{-1}{a}\right)$ or zero. Hence, according to the *Lyapunov stability theory*, at the considered equilibrium point the system may have oscillations as shown in Fig. 3(a) (since no desired visual features are used). Thereby, the challenge is to able to stop the process at best focus (since no desired visual features are used) without acquiring any other unnecessary image. To do this, an optional secondary task has been realized. It links the electronic working distance W with the gradient of sharpness function. Assuming the change in the W is small (especially near best focus point), the secondary task function is given by

$$\Gamma(\mathbf{s}) = \left(\frac{\partial s}{\partial W}\right) \mathbf{L}_W{}^{-1} = \left(\frac{\mathbf{s}_k - \mathbf{s}_{k-1}}{W_k - W_{k-1}}\right) \mathbf{L}_W{}^{-1} \tag{9}$$

where, $\mathbf{L}_W = -W$ is the interaction matrix to link gradient with working distance. As the secondary function provides derivative information of the sharpness, it crosses zero when sharpness reaches maximum. The overall process can be stopped at this point. Fig. 3(b) shows the relationship between secondary function and sharpness at $\times 1000$ magnification. Due to the addition of noise, lot of peaks can be seen near zero. To avoid this problem a value of -0.01 has been used instead of 0. Since, secondary task does not have any effect on the main focus control (always < 0.1), it can be integrated it with the primary one. Using (3) and (9), the designed final control law is given by

$$F_{k+1} = F_k + \Psi(\mathbf{s}) + \Gamma(\mathbf{s}). \tag{10}$$

D. Automated Depth Estimation using Autofocus

The resolution of the estimated depth mainly depends on the DOF. Less the latter the better is the resolution (i.e., higher magnification implies better accuracy). Hence, we begin the process by reducing the DOF.

D.1 Reducing DOF

The DOF mainly depends on the semi-angle of the beam. Considering aperture diameter A and working distance W (see Fig. 1(b)) the semi-angle θ can be given by

$$\theta = \tan^{-1}\left(\frac{A}{2W}\right) \simeq \left(\frac{A}{2W}\right) \text{ at } \theta < 100\ mrad. \tag{11}$$

Depending on θ and the sample resolution ϱ_{sample}, the DOF is given by

$$DOF = \frac{\varrho_{sample}}{\theta} = \frac{2\varrho_{screen} W}{AM}\ [\mu m] \tag{12}$$

where, M is the magnification and $\varrho_{sample} = \left(\frac{\varrho_{screen}}{M}\right)$. Hence, the DOF mainly depends on the aperture diameter and working distance at any particular magnification.

From (12), DOF can be reduced by: decreasing W, increasing M or A. In order to reduce W i.e., to move the focal plane towards the lens pole piece, it is required to physically lift up the sample. Nonetheless, this is not a feasible option, as increase in sample height may damage the lens. The minimum acceptable working distance is $\in [3-5]\ mm$. The other option is to increase M which decreases the field-of-view making it difficult to track the objects during a manipulation task. Fig. 4(a) shows the relation between magnification and DOF. The final option is by increasing A, the semi-angle of the beam increases (from (11)) and simultaneously DOF decreases. This option can be used only with the SEMs having multiple exchangeable apertures and has been selected in this work. Four apertures of different diameters ($30\ \mu m$, $50\ \mu m$, $70\ \mu m$ and $100\ \mu m$) are available with the used experimental SEM. By observing the amount of blur, $100\ \mu m$ aperture is selected[1] and is fixed for all the experiments. However, image resolution may be limited using bigger apertures. An important point to note is that this step can be treated as a pre-check and is not required to perform each time when a new depth estimation task is started.

D.2 Depth Estimation

The depth estimation process is performed in two phases: initialization and execution. In the first hand, initialization phase is one time operation and is performed at the beginning of each new depth estimation task. During this phase, the system will perform autofocus on the entire scene to obtain a global best focus. This phase has been performed to speed-up the next execution phase. On the other hand, the execution phase is a continuous process that starts immediately after the initialisation phase to estimate the real-time depth. During this phase, the system will perform autofocus in an automatically selected region of interest (ROI). Instead of starting the autofocus process from the beginning of focus range, the system will now use the new initial step obtained from the initialization phase.

This new step has been selected based on the step difference (between two focus steps) during the initialization phase autofocus (e.g., 2, 3, etc.). A threshold value $\kappa = 3\,(F_b - F_{b-1})$

[1] A calibration (beam shift and tilt) may be required for older SEMs when the aperture is modified.

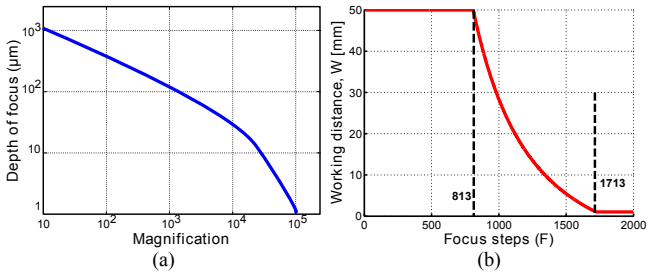


Fig. 4. Relationship between (a) magnification and DOF (b) focus step and working distance in JEOL SEM.

(F_b is the obtained best focus step from initialization phase) has been used for this purpose. The particular focus step (during initialization phase autofocus) from which the step difference $\leq \kappa$ will be retained as the execution phase initial focus step. An example of new initial step is pointed in the Fig. 3(a) with a circle (on solid curve). Furthermore, to increase the estimation speed, the gain α will be automatically changed to five times more than the previous value at the particular focus step. This phase restarts automatically with a change in the ROI.

The used image acquisition system (DISS5) provides a simple control for the focus by linking the working distance with a series of focus steps (i.e, each step modifies the working distance). This relation is shown in Fig. 4(b). Using this relation, the depth value can be computed for any focus step obtained from any of the two phases. It is performed using polynomial approximation given by

$$W = \begin{cases} 50, & \text{if } F \leq 813 \\ 1, & \text{if } F \geq 1713 \\ \sum_{i=1}^{j=5} p_i F^{j-i} & \text{otherwise} \end{cases} \quad (13)$$

where, $p_{i=1\ldots5}$ are the polynomial coefficients.

III. Real-World Validations

A. Experimental System

To validate the proposed method, different experiments are realized using a tungsten source SEM (JEOL JSM 820) along with a DISS5 imaging system (point electronic GmbH) and a work computer (Intel Core 2 Duo, CPU 3.16 GHz, and 3.25 GB of RAM). With SEM, the maximum allowable electronic working distance is $50\ mm$ and the total focus range is divided into 4000 coarse steps. The DISS5 is responsible for transferring beam control commands to the microscope and recording the image data. These images are analyzed in the work computer and a focus control command is issued. For all the tests, the SEM secondary electron images of size 512×512 pixels and an aperture diameter of $100\ \mu m$ are used.

B. Automatic Depth Estimation

Initial experiments are conducted to estimate the depth during a manipulation task. A nanomanipulator (Kleindiek MM3A) whose end effector is fixed with a gold coated tip-less cantilever (Fig. 5) has been used in this work. Silicon microparts of dimensions $500 \times 10 \times 20\ \mu m^3$ are positioned beneath the cantilever. The objective is to bring down the cantilever to the focal

Fig. 5. Kleindiek nanomanipulator MM3A mounted inside the SEM vacuum chamber. (top-inset) End-effector fixed with a tip-less cantilever (bottom inset) SEM image of the cantilever.

plane of micropart by moving the MM3A (at speed-6 provided by Kleindiek) along the optical axis. Manipulator movement commands are transferred from the work computer via serial port. Since this work deals with depth estimation, automatic control of the MM3A degrees of freedom is not discussed in this paper. For this test an acceleration voltage of $5\ kV$ and a scan time of 180 nanoseconds per pixel (maximum allowed by the system) have been used to generate the electron beam and images, respectively.

B.1 Initialization Phase

In this phase, a global autofocus is performed on the entire scene. An α value of 200 is used. Simultaneously, the process speed can be increased using higher values of α (refer to Table I). However, extreme values for α may cause oscillations at the initial step when the *in-focus* position is closer to the starting position and may lead to the task failure. The plots shown in Fig. 6(a) and Fig. 6(b) illustrate the variations of image sharpness along with the cost and secondary task functions, respectively and Fig. 6(c) to Fig. 6(e) depict some of the images acquired at different times during the autofocus at $\times 8300$ magnification. From the obtained results it can be seen that the autofocus has been successfully accomplished and the process stopped at best focus position. Since this is a one-time operation to speed up the depth estimation, the performance of the proposed autofocus has been compared to the iterative search-based method at different magnifications. For the comparison experiment, the step sizes used with search-based method are 100, 20 and 1, respectively in each iteration. A best focus position returned by an expert human operator was used as a reference to estimate the focusing errors with both proposed and iterative methods. Four different runs are performed at each magnification and the obtained results are summarized in Table II (S = standard deviation, R = RMSE). The results obtained by performing the experiment at two different conditions: increased brightness (C1) and decreased scan speed of 720 nanoseconds per pixel (C2) are summarized in Table III. Results prove the method's repeatability and it can also be noticed that the accuracy of proposed method is better than search-based method. Apart from that, it can be also noticed that the accuracy increases with increase in magnification, which shows the method's dependability on DOF. Besides, the average number of images acquired and the total time taken by the proposed method are **19** and **6.5 seconds**, respectively and by

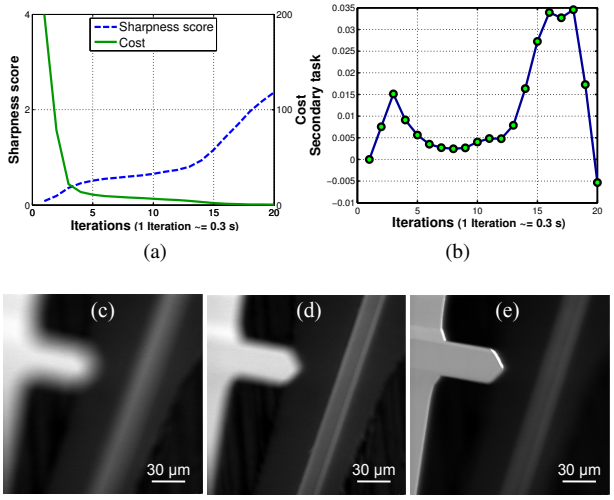


Fig. 6. (a) Sharpness score and cost (b) secondary task variations during the initialization phase autofocus at ×8300 magnification. (c) to (e) Images obtained at 5, 14 and 19 iterations, respectively.

TABLE I

GLOBAL AUTOFOCUS TIME (IN *seconds*) WITH VARIOUS α VALUES.

Mag (×)	$\alpha = 100$	$\alpha = 200$	$\alpha = 300$	$\alpha = 400$
4000	11.7	7.7	5.1	4.1
8000	12.9	6.5	4.8	3.7
10000	10.6	7.9	4.3	4.1

TABLE II

INITIALIZATION PHASE AUTOFOCUS COMPARISON RESULTS.

Mag	Obtained W (mm)			Performance (mm)	
	manual	ours	search	ours	search
4000	9.694 9.679 9.685 9.685	9.637 9.626 9.630 9.623	9.621 9.615 9.631 9.608	S = 0.003 R = 0.056	S = 0.010 R = 0.067
9000	9.597 9.613 9.607 9.589	9.555 9.564 9.570 9.538	9.536 9.571 9.550 9.538	S = 0.005 R = 0.045	S = 0.008 R = 0.053
13000	9.569 9.575 9.556 9.563	9.536 9.552 9.529 9.540	9.456 9.528 9.483 9.465	S = 0.004 R = 0.026	S = 0.029 R = 0.082

TABLE III

PERFORMANCE EVALUATION OF PROPOSED METHOD AT DIFFERENT EXPERIMENTAL CONDITIONS.

Case	Mag	Ours		Search	
		S	RMSE	S	RMSE
C1	4000	0.0045	0.0517	0.0081	0.0704
	9000	0.0049	0.0431	0.0073	0.0513
	13000	0.0037	0.029	0.0089	0.054
C2	4000	0.0027	0.0401	0.0064	0.066
	9000	0.0024	0.0339	0.0041	0.0716
	13000	0.0033	0.0114	0.0052	0.042

the search-based method are 55 and 18.9 seconds, respectively. In others words, the proposed method allows to improve the accuracy with a rate of $\mathbf{44}\%$ and at the same time the speed is increased by a rate of $\mathbf{290}\%$.

B.2 Execution Phase

Once the initialization phase is completed, the execution phase starts automatically. As explained in section D.2, the starting focus step and cost values are adjusted automatically to speed up the process. The magnification is fixed at × 8300 to have both cantilever and micropart in the SEM's field-of-view. Here, the DOF is measured to be 27.3 μm. As an initial step, the micropart and cantilever ROIs are obtained using the template matching technique [20]. The micropart depth has been measured initially by autofocusing in its ROI. When the manipulator moves down, the cantilever depth has been estimated by autofocusing in its ROI. The manipulator movement is automatically stopped when it reaches the micropart's DOF i.e., the required depth to reach. Fig. 7(a) shows the 2D trajectory followed by the manipulator and Fig. 7(b) shows the 3D trajectories (to avoid any biased interpretation) for five different runs starting from different initial positions. The resulting standard deviation of the obtained final depths for these 5 runs is $\mathbf{1.18}$ $\mu\mathbf{m}$. Besides, the average time taken to estimate the depth at a particular position using the above mentioned scan time is **1.3 seconds**. With modern SEMs having high scanning speeds, this time can be even reduced to 10^{th} of a second.

C. Topographic Depth

The sample used for this test is an aluminium sample of dimensions $4{\times}1{\times}2\ mm^3$ containing step patterns (Fig. 8(a)). The magnification is fixed at $\times 3800$ where the DOF is measured to be $55\ \mu m$. The images are generated using an acceleration voltage of $5\ kV$ and scan speed of 180 nanoseconds per pixel. A window (ROI) of size $64.7{\times}97.5\ \mu m^2$ has been used to estimate the depth. The entire region of the sample has been scanned by moving the window. The experimentally obtained depth values are approximated with their average values to find the step depth. Fig. 8(b) shows the obtained and averaged topographic depth map of entire sample. These averaged depth values are compared with the original depth values that are measured us-

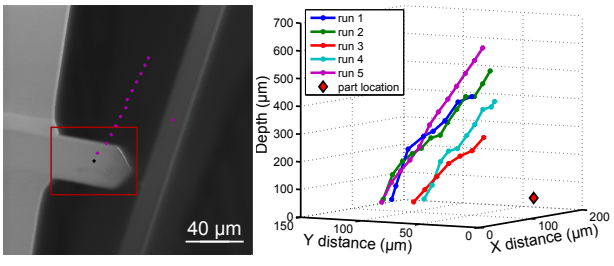
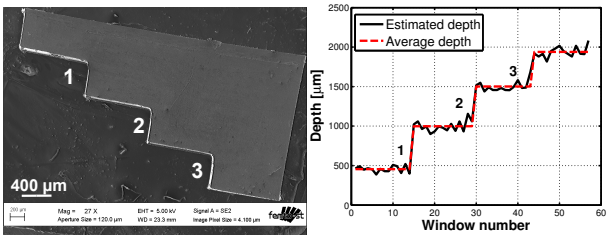


Fig. 7. (a) Image showing the trajectory followed (dotted line) by the manipulator (tracked region is pointed with a rectangle) (b) estimated depths for 5 different runs (part location (desired depth) is pointed with diamond).



(a) (b)

Fig. 8. (a) Aluminium sample with step pattern (b) topographic depth map for entire sample where, mean depth is represented by dotted line. These mean values are then compared with profilometer values to assess the accuracy.

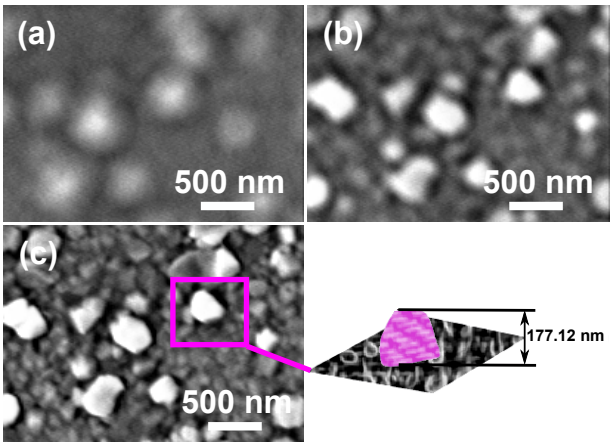


Fig. 9. Focusing series of the gold on silicon sample at $\times$ 25,000 magnification. Focused on the brighter gold particles.

ing a *contact-profilometer* in order to estimate the measuring accuracy of our method. Table IV summarize these results for different steps marked in Fig. 8(a). It can be seen that the depths are estimated with an accuracy of $\mathbf{4.3}$ $\mu\mathbf{m}$ to $\mathbf{11.6}$ $\mu\mathbf{m}$ for real position ranging from 454.7 μm to 515.3 μm. If we consider the relative errors, they are only from $\mathbf{0.85}\%$ (for 4.3 μm of error) to $\mathbf{2.20}\%$ (for 11.6 μm of error). The overall accuracy is computed to be $\mathbf{8.9}$ $\mu\mathbf{m}$ at the used magnification and can be improved by increasing the magnification.

Another validation was performed to illustrate the flexibility of the proposed approach. It concerns the ability of our technique to work at high magnification, for instance $\times$ 25,000. To this, a test is conducted in order to compute the depth between the surfaces of gold and silicon particles (Fig. 9). Similar scan speed and acceleration voltage as above are used. Intensity thresholding has been used to differentiate the brighter (gold) and darker (silicon) regions. The obtained regions are autofocused separately using the proposed method to estimate the depth. Fig. 9 shows a series of images obtained during this experiment. From the obtained results, the depth between gold and silicon particles has been computed to be $\approx \mathbf{177.12\ nm}$.

D. Object orientation estimation

The final experiments are performed to estimate the object's orientation and to evaluate the performance. A microgripper (Femto tools FT G32) mounted on an inclined support inside the SEM vacuum chamber is used for these tests (Fig. 10(a)).

TABLE IV
ORIGINAL, ESTIMATED AND ERROR FOR THE AVERAGE STEP DEPTHS.

Step	Original (μm)	Obtained (μm)	Error (μm)
1	515.3	526.9	11.6
2	503.5	507.8	4.3
3	454.7	445.5	9.2

TABLE V
ESTIMATED ANGLES AT DIFFERENT CONDITIONS.

Case	Original (°)	Computed (°)	error (°)
EC1	27.13	26.97	0.16
EC2	27.13	27.24	-0.11
EC3	27.13	26.59	0.54

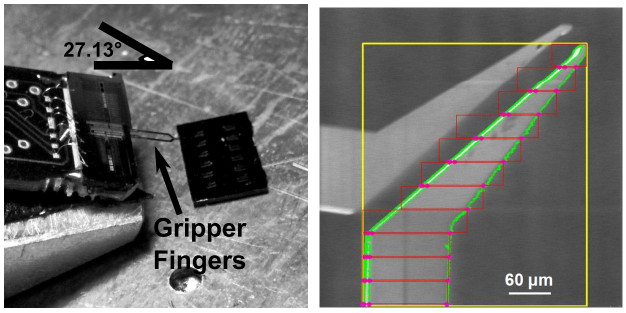
Even though it is possible to use the cantilever mounted on the nanomanipulator for this test, the original inclination is not known for validations. Hence an inclined surface with known angle is used for this experiment. The used acceleration voltage is 10 kV. The magnification is fixed to $\times 5300$ to have both fingers in field-of-view, where the DOF is measured to be 32 μm. In this case the windows are selected automatically by detecting a static gripper finger (Fig. 10(b)). The first experiment concerns using a fixed angle at three different experimental conditions: high scan speed of 180 nanoseconds per pixel i.e., high noise (EC1), low scan speed of 720 nanoseconds per pixel i.e., moderate noise (EC2) and high brightness (EC3). The obtained depth values along with the fitted lines for EC1 ad EC2 are shown in Fig. 10(c). From the obtained result, the angle of inclination i.e., the angle made by the line segment with respect to the horizontal axis has been computed using a 3D regression method by taking into account all the intermediate depth positions. The obtained results are summarized in Table V. It can be seen that the error is less with slower scan speeds. Second experiment is conducted to evaluate the performance at different angles (at EC1). Table VI shows the obtained results.

IV. CONCLUSION

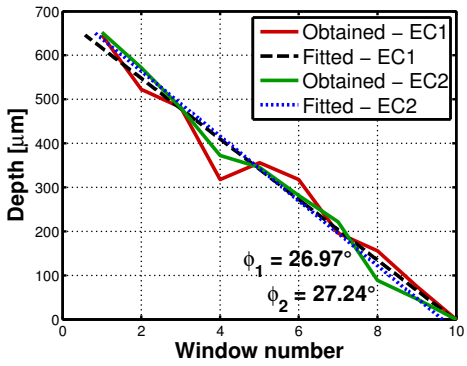
In this paper, we presented a fast and efficient solution to tackle the problem of dynamic depth estimation under SEM. The developed method estimates the depth by finding the point of maximum sharpness in the local sub-regions of a scene. The automatic nature mainly comes from the flexibility of the developed autofocus i.e., by controlling the device focus steps by means of visual servoing. Since, the best focus is obtained in a single run, the method overcomes the problem of lens hystere-

TABLE VI
PERFORMANCE EVALUATION OF THE ANGLE ESTIMATION.

Original (°)	Computed (°)	error (°)
58.70	58.53	0.17
45.19	44.98	0.21
69.31	69.44	-0.13
	RMSE	**0.083**



(a) (b)



(c)

Fig. 10. (a) Mounted microgripper on an inclined surface at an angle of 27.13° inside SEM chamber (b) detected microgripper fingers along with the obtained regions (c) obtained depth maps at different operating conditions. Raw depth values are fitted to perform angle measurements.

sis and problem of entire focus range scan. The overall depth estimation method was divided into two phases (initialization and execution), such that the method can exhibit fast depth estimation. It takes approximately 1 to 2 seconds (with the used set-up at scan speed of 180 nanoseconds per pixel) to estimate the depth for a region during execution phase and can be used with real-time operations. The obtained results showed that the method can be used in micro-nanomanipulation for Z positioning as well as for orientation estimation and for 3D structural characterization. Future experiments are also planned to test the method's performance aside with astigmatism correction and also with the other SEM parameters such as acceleration voltage, emission current etc.

REFERENCES

[1] Changhai Ru, Yong Zhang, Yu Sun, Yu Zhong, Xueliang Sun, David Hoyle, and Ian Cotton, "Automated four-point probe measurement of nanowires inside a scanning electron microscope," *IEEE Trans. Nanotechnol.*, vol. 10, no. 4, pp. 674–681, 2011.

[2] Ezequiel Ponz, Juan Luis Ladaga, and Rita Dominga Bonetto, "Measuring surface topography with scanning electron microscopy. i. ezeimage: a program to obtain 3d surface data," *Microsc. Microanal.*, vol. 12, no. 2, pp. 170–177, 2006.

[3] Oscar Yánez-Suárez and Mahmood R Azimi-Sadjadi, "Identification and measurement of fibers in scanning electron microscopy images using a high-order correlation process," *IEEE Trans. Instrum. Meas.*, vol. 48, no. 1, pp. 55–61, 1999.

[4] Naresh Marturi, Brahim Tamadazte, Sounkalo Dembélé, and Nadine Piat, "Visual servoing schemes for automatic nanopositioning under scanning electron microscope," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2014, pp. 981–986.

[5] Bradley E Kratochvil, Lixin Dong, and Bradley J Nelson, "Real-time rigid-body visual tracking in a scanning electron microscope," *Int. J. Robot. Res.*, vol. 28, no. 4, pp. 498–511, 2009.

[6] Yan Liang Zhang, Yong Zhang, Changhai Ru, Brandon K Chen, and Yu Sun, "A load-lock-compatible nanomanipulation system for scanning electron microscope," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 1, pp. 230–237, 2013.

[7] Robert Tunnell and Sergej Fatikow, "3d position detection with an fib-sem dual beam system," in *10th WSEAS Int. Conf. on communications, electrical & computer engineering*, 2011, pp. 128–133.

[8] Ruggero Pintus, Simona Podda, and Massimo Vanzi, "An automatic alignment procedure for a four-source photometric stereo technique applied to scanning electron microscopy," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 5, pp. 989–996, 2008.

[9] Volkmar Eichhorn, Sergej Fatikow, Thomas Wich, Christian Dahmen, Torsten Sievers, Karin Nordström Andersen, Kenneth Carlson, and Peter Bøggild, "Depth-detection methods for microgripper based cnt manipulation in a scanning electron microscope," *J. Micro-Nano Mechatron.*, vol. 4, no. 1-2, pp. 27–36, 2008.

[10] Volkmar Eichhorn, Sergej Fatikow, Tim Wortmann, Christian Stolle, Christoph Edeler, Daniel Jasper, Ozlem Sardan, P Boggild, Guillaume Boetsch, and Christophe Canales, "Nanolab: A nanorobotic system for automated pick-and-place handling and characterization of cnts," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2009, pp. 1826–1831.

[11] Sounkalo Dembélé, Nadine Piat, Naresh Marturi, and Brahim Tamadazte, "Gluing free assembly of an advanced 3d structure using visual servoing.," in *Micromechanics and Microsystems Europe Workshop*, 2012, pp. 6–11.

[12] Changhai Ru and Steve To, "Contact detection for nanomanipulation in a scanning electron microscope," *Ultramicroscopy*, vol. 118, pp. 61–66, 2012.

[13] Naresh Marturi, Sounkalo Dembélé, and Nadine Piat, "Depth and shape estimation from focus in scanning electron microscope for micromanipulation.," in *Proc. IEEE Int. Conf. Control Autom. Rob. Embed. Syst.*, 2013, pp. 1–6.

[14] Thomas Wich, Christian Stolle, Tim Luttermann, and Sergej Fatikow, "Assembly automation on the nanoscale," *CIRP Journal of Manufacturing Science and Technology*, vol. 4, no. 4, pp. 391–400, 2011.

[15] ME Rudnaya, RMM Mattheij, and JML Maubach, "Evaluating sharpness functions for automated scanning electron microscopy," *Journal of microscopy*, vol. 240, no. 1, pp. 38–49, 2010.

[16] Yu Sun, Stefan Duthaler, and Bradley J Nelson, "Autofocusing algorithm selection in computer microscopy," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* IEEE, 2005, pp. 70–76.

[17] ME Rudnaya, HG Ter Morsche, JML Maubach, and RMM Mattheij, "A derivative-based fast autofocus method in electron microscopy," *J. Math. Imaging. Vision.*, vol. 44, no. 1, pp. 38–51, 2012.

[18] FC Nicolls, G de Jager, and BT Sewell, "Towards a predictive autofocus algorithm for sem," *Proc. Electron Microsc. Soc. South Afr*, vol. 24, no. 11, 1994.

[19] Naresh Marturi, Brahim Tamadazte, Sounkalo Dembélé, and Nadine Piat, "Visual servoing-based approach for efficient autofocusing in scanning electron microscope," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 2677–2682.

[20] Torsten Sievers and Sergej Fatikow, "Real-time object tracking for the robot-based nanohandling in a scanning electron microscope," *J. Micromechatron.*, vol. 3, no. 3/4, pp. 267, 2006.