

Depth and Shape Estimation from Focus in Scanning Electron Microscope for Micromanipulation

Naresh Marturi, Sounkalo Dembélé, and Nadine Piat

Automatic control and Micro Mechatronic Systems (AS2M) department, Institute FEMTO-ST,
25000 Besançon, France.

Email: first name.last name@femto-st.fr

Abstract—Inter-object depth estimation is always a major concern for micromanipulation using scanning electron microscope (SEM). So far, various methods have been proposed for estimating this depth based on stereoscopic imaging. Most of them require external hardware unit or manual interaction during the process. In this paper, using the image focus information, different methods are presented for estimating the inter-object depth for micromanipulation and the local pixel point depth for 3D shape reconstruction. In both cases, the normalized variance has been used as the sharpness criteria. For inter-object depth estimation, a visual servoing-based autofocusing method has been used to maximize the sharpness in object region windows. For Shape reconstruction, a stack of images are acquired by varying the working distance. These images are processed to find the maximum sharpness of each pixel and consequently reconstructing the surface. Developments are validated in a robotic handling scenario where the scene contains a microgripper and silicon microstructures.

Keywords—Scanning electron microscope; depth from focus; shape from focus; micromanipulation

I. INTRODUCTION

Over the past few years, SEM has been emerged as an important tool in performing robotic micromanipulation, especially where the sample dimensions are of few micrometers. It is widely used in various scientific, industrial and biomedical applications. Normally, SEM images are produced by raster scanning the sample surface by means of a focused beam of electrons. As the wavelength of primary beam (electrons) is very small (0.12 Å), a SEM can be characterized by high resolution (better than 1 nanometre), broad range of magnification (up to 500,000 ×) and high depth of field.

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 microhandling and assembly applications. Even though SEM images are capable of providing surface topography information, they are purely two dimensional. The pixel intensity values do not contain any information regarding the height. In order to tackle this problem, many research works have tried to apply stereoscopic imaging-based technique that is commonly seen in the optical systems [1]–[3]. As the system possesses only a single imaging sensor, most of them tried to acquire a stereo pair of images

by concentrically tilting the sample with a specific angle. The main difficulty lies in determining the accurate tilt angle and finding the correspondence between the acquired noisy SEM images [4]. Moreover, it requires a precise calibration. However, tilting a sample is not a feasible option when performing an autonomous task. To overcome this problem, Jahnisch and Fatikow have developed a special hardware system for beam deflection in order to observe the sample with different angles [5]. However, the system is limited by its small angle of deflection. Another solution can be by using a focused ion beam (FIB) system along with SEM [6], but it increases the overall cost of the system and damages the sample. In [7], a touchdown sensor has been used to detect the contact between two micro-objects and simultaneously computing the depth in nanometres. In [8], a laser triangulation method has been used for depth detection utilizing an additional optical microscope. A photometric stereo-based approach using multiple back scattered electron detectors has been used in [9].

Apart from the stereo imaging-based methods that are relatively complex, the depth information can be extracted simply by using the focus information in the images. These types of methods are commonly used for depth estimation and autofocusing in optical microscopes, where the depth of field is narrow [10]. The underlying idea behind these techniques is that they use a sequence of images obtained by moving the microscope along its optical axis with a predefined finite number of steps. Depth is then computed by applying a focus measure on the images to find the camera position at which the object in frame appears to be in focus. In the case of SEM, in spite of having a high depth of field, it is still possible to use the focus-based approach by simple modifications in the system imaging parameters. One such technique has been used in this work.

In this paper, using the image focus information two approaches have been developed; one for estimating the inter-object depth and the other for estimating the individual pixel point depth for reconstructing the micro-objects shape. The developed inter-object depth estimation method uses a visual servoing-based autofocusing algorithm explained in [11] for depth computation. The main advantage associated with the developed method is that it does not require any additional hardware modifications to the existing system. Apart from that, as the autofocusing algorithm is proved to be fast and robust, the implemented depth detection method can be used with real time systems. On the other hand, the developed shape reconstruction method uses multiple images acquired by varying the working distance. It searches for the maximum

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-LABX-01-01) and the Equipex ROBOTEX (ANR-10-EQPX-44-01) projects.

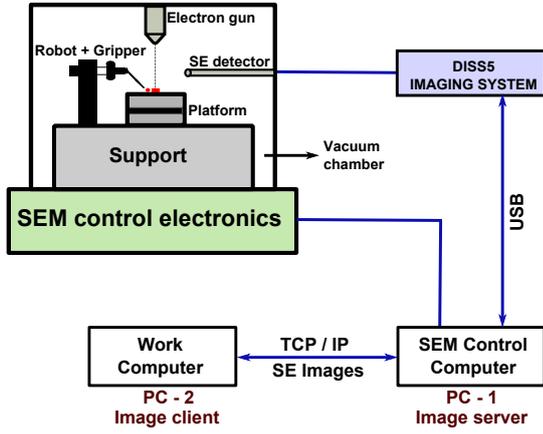


Fig. 1. Experimental setup architecture.

sharpness of each pixel in the images and assign the depth from SEM focusing parameters (i.e. from the working distance).

The rest of the paper is organized as follows: section II presents the details regarding experimental setup and focusing process in a conventional SEM. The developed method for estimating the inter-object depth is presented along with the experimental results in section III. Section IV explains the implemented shape reconstruction method using the computed pixel point depth information.

II. EXPERIMENTAL SETUP AND FOCUSING IN SEM

A. Experimental setup

The experimental setup architecture used for this work is shown in the Fig. 1. It consists of a JEOL JSM 820 SEM, a DISS5 imaging system along with two computers. The electron column of the JEOL SEM is equipped with an electron gun with tungsten filament, an anode, different sets of electromagnetic lenses, an aperture strip, two sets of scan coils, a secondary electron detector and a movable platform. The primary computer (PC 1: Intel Pentium 4, CPU 2.24 GHz and 512 MB of RAM) is connected to the SEM control electronics and imaging system. It is solely responsible for controlling the microscope. The work computer (PC 2: Intel Core 2 Duo, CPU 3.16 GHz, and 3.25 GB of RAM) is connected to the primary one using an Ethernet cross-over cable. The communication between the two computers is accomplished by implementing a client-server model using TCP/IP. The server program runs continuously from the primary computer and is responsible for sending the focus commands to the system and receiving the raw image data from DISS5. This data is later stored as an image. The saved images are transferred to the image client up on receiving a request. Later, the client uses these images for depth estimation and shape reconstruction.

B. Focusing in SEM

The electromagnetic lenses that contribute towards the focusing process in SEM are the condenser lenses lying above the aperture and the objective lenses lying below the aperture. The condenser lenses are mainly responsible for controlling the spot size and beam diameter. At first, these lenses converge the beam to a spot and this spot sized beam flares out again and passes through the aperture, where non-directional electrons are filtered out. The objective lenses that are present below

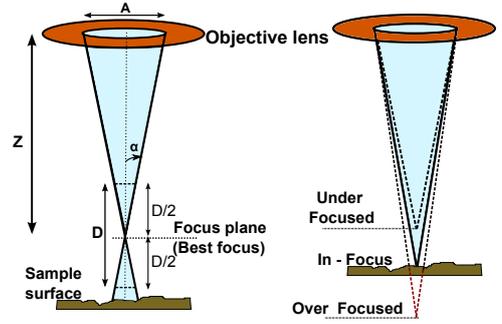


Fig. 2. (a) Focusing geometry in SEM (b) various focusing scenarios.

the aperture, converge the beam once again and focus it on to the sample surface. The focusing process is shown in the Fig. 2(a). Coarse focusing is performed by adjusting the electronic working distance (Z) that is attained by varying the current passing through the lenses. This is the distance measured in between the final pole piece of the lens 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. Using the aperture diameter A and the working distance Z , the semi angle of the electron beam can be given by (1).

$$\alpha = \tan^{-1} \left(\frac{A}{2Z} \right) = \left(\frac{A}{2Z} \right) \text{ at } \alpha < 100 \text{ mrad} \quad (1)$$

If we consider the resolution on the sample is δ_{sample} , the corresponding resolution on the screen is $\left(\frac{\delta_{screen}}{M} \right)$, where M is the magnification. Depending on the angle α and resolution δ , the depth of focus D is given by (2).

$$D = \frac{\delta_{sample}}{\alpha} = \frac{2\delta_{screen}Z}{AM} [\mu m] \quad (2)$$

So the depth of focus is mainly dependent on the size of aperture selected and electronic working distance. Fig. 2(b) shows the various focusing scenarios in SEM.

The JEOL SEM used for the experiments is installed with a dynamic focusing module that modifies the current passing through the objective lens and thus the electronic working distance can be preselected. The DISS5 system 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). The relation between focus steps and electronic working distance is shown in Fig. 3. Subsequently, the value of working distance (Z) for any given focus step (F) at a fixed

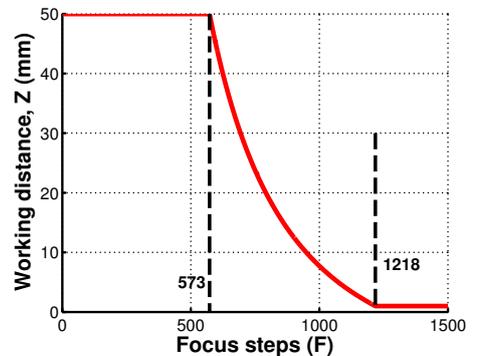


Fig. 3. Relationship between the focus step and working distance in JEOL SEM used for experiments.

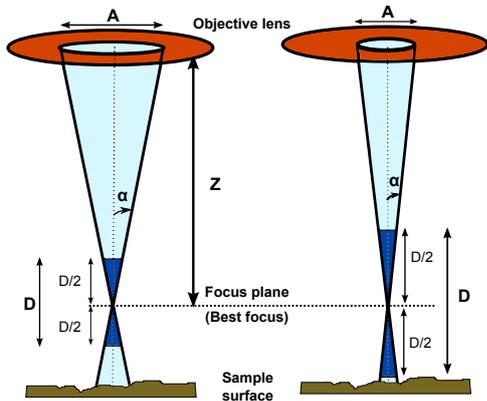


Fig. 4. Relationship between the aperture diameter and depth of focus in SEM.

sample height can be computed by using the curve equation obtained by approximating it with a polynomial given in (3).

$$Z = \begin{cases} 50, & \text{if } F \leq 573 \\ 1, & \text{if } F \geq 1218 \\ p_1 F^4 + p_2 F^3 + p_3 F^2 + p_4 F + p_5, & \text{otherwise} \end{cases} \quad (3)$$

where, $p_1 \dots p_5$ are the polynomial coefficients.

III. DEPTH ESTIMATION FROM FOCUS

Most of the available focus-based depth estimation methods in optical microscopy relies on applying the sharpness function on a stack of images to measure the relative blur. They utilize the advantage of low depth of field and compute the focus of the region that is in focus. However, with SEM, it is highly challenging as the depth of focus is much larger than in optical microscopes. So it is necessary to minimize the depth of focus by varying the SEM imaging parameters before estimating the depth. Once it is reduced, the following tasks are performed to compute the depth: selection of region of interests (ROI) for all the objects present, perform autofocus on all regions and estimate the depth between objects.

A. Reducing the depth of focus

As mentioned in the previous section, from (2), the depth of focus (D) mainly depends on the aperture diameter and the working distance. So in order to reduce D , we have three possible options: increasing the aperture diameter (A), increasing the magnification (M), or reducing the electronic working distance (Z). By increasing A the semi angle of the beam increases and simultaneously D decreases. This option can be used only with the SEMs having multiple exchangeable apertures. Fig. 4 shows the relation between aperture diameter and depth of focus. The other option is to increase M . But increase in magnification limits the field of view making it difficult to track the parts. The relation between the magnification and depth of focus is shown in Fig. 5. Apart from these two options, depth of field can be reduced by decreasing the electronic working distance i.e. to move the sample close towards the objective lens pole piece. However, this option is not a good choice, as increase in sample height may damage the lens. The minimum acceptable working distance is about $5 - 8 \text{ mm}$.

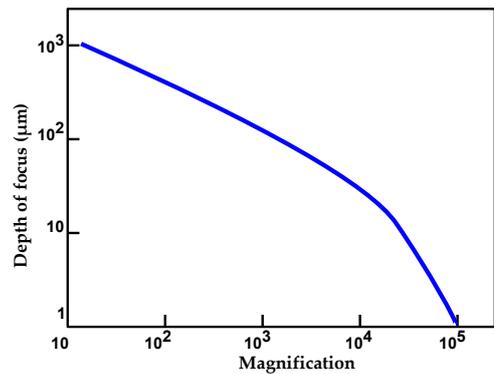


Fig. 5. Relationship between the magnification and depth of focus in SEM.

In this work, we choose the option of varying the aperture diameter. The aperture strip present in the JEOL SEM column contains four apertures of diameters $100 \mu\text{m}$, $70 \mu\text{m}$, $50 \mu\text{m}$ and $30 \mu\text{m}$ respectively. By observing the amount of blur, an aperture with $70 \mu\text{m}$ has been selected for this work. However, high aperture sizes may result in low resolution.

B. Sharpness function

Various sharpness functions are available in literature [12]. Most commonly used are: derivative-based functions like Laplacian, Brenner, statistics-based functions like variance, normalized variance, correlation-based functions and wavelets-based functions. When applied on the images or regions, the maximum value of the sharpness function is found at the image with best focus. For this work, after evaluating all the available sharpness functions, normalized variance has been selected for estimating the sharpness score as it provides a good compromise for speed and accuracy. For a given image $\mathbf{I}_{M \times N}$ where M and N are the width and height, the normalized variance \mathbf{S}_{NV} is given by (4).

$$\mathbf{S}_{NV} = \frac{1}{MN} \frac{1}{\mu} \sum_{u=1}^M \sum_{v=1}^N (I(u, v) - \mu)^2 \quad (4)$$

where, I is the intensity of the pixel at (u, v) and μ is the pixel mean intensity given by (5).

$$\mu = \frac{1}{MN} \sum_{u=1}^M \sum_{v=1}^N I(u, v) \quad (5)$$

It estimates the focus score by computing the variations in grey level intensities among the pixels. Fig. 6 shows the principle of focusing and variation of sharpness function (red curve) in SEM. Fig. 7 shows the sharpness score computed using (4) for a series of focus steps at $1000 \times$ magnification.

C. Inter-object depth estimation

In the context of this work, the inter-object depth estimation has been performed using a FT G32 force sensing microgripper from FEMTO TOOLS (Fig. 8(a)) and silicon microparts of dimensions $10 \times 500 \times 20 \mu\text{m}$ (Fig. 8(b)). The gripper is fixed inside the SEM vacuum chamber and the microparts are placed on the positioning table underneath the gripper. As an initial approach, the sharpness function given by (4) has been applied over the total range of focus steps (i.e. by changing the working distance) to find the individual object

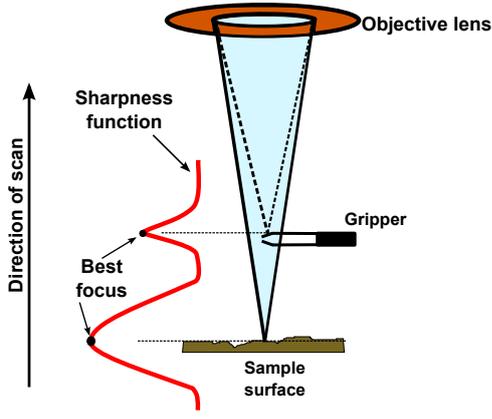


Fig. 6. Principle of focusing and sharpness score variation.

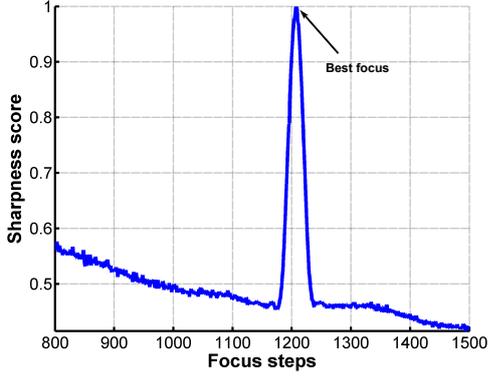


Fig. 7. Sharpness score variation given by (4) with respect to the focus steps at $1000\times$ magnification.

depth. The resulted curve is shown in the Fig. 9. As expected, two local maxima have been found indicating the height of the gripper and the microparts with respect to the lens pole piece. Although this approach determined the absolute height, it is a time consuming process to scan the entire region of focus. In order to overcome this problem, in this work, we first segment the image into different regions of interest (ROI). For each ROI, the best focus position is computed using the visual servoing-based autofocusing technique explained in [11]. Unlike the search-based methods, visual servoing-based method maximises the sharpness using an adaptive gain and reaches the best focus in few iterations. Also, it has an ability to skip the local maxima that affect the focusing process. The autofocusing method is summarized in algorithm 1. By computing the working distance that is associated with the obtained best focus step using (3) provides the height of the object. Finally, the inter-object depth Z_{io} is obtained by (6).

$$Z_{io} = |Z_{part} - Z_{gripper}| \quad (6)$$

where, Z_{part} and $Z_{gripper}$ are the obtained heights of micropart and gripper regions respectively.

For this task, the ROI is determined on the basis of thresholding the pixel intensity values. The value of this threshold has been selected manually. Optionally, the ROI can also be determined by using object recognition or tracking algorithms. Fig. 10(a) and 10(b) show the best focused images of each ROI. The plots shown in the Fig. 11 show the results of autofocusing process applied on different regions. Fig. 11(a) and 11(b) show the heights obtained for gripper and micropart

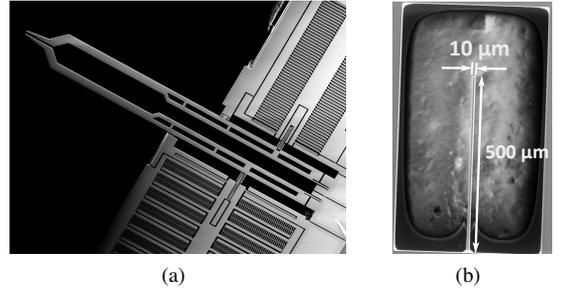


Fig. 8. The objects used for the experiments (a) FEMTO TOOLS FT G32 microgripper (b) silicon micropart.

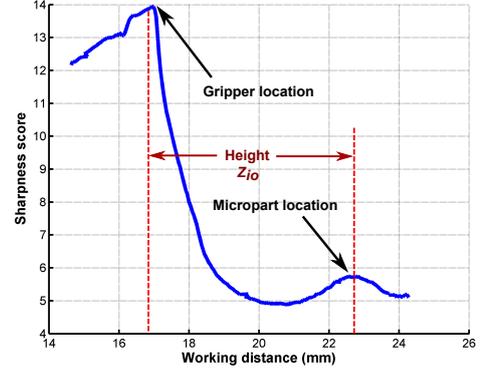


Fig. 9. Sharpness function indicating the gripper and micropart locations.

regions respectively when they are far from each other and the computed Z_{io} is 5.632 mm. Fig. 11(c) and 11(d) show the heights obtained for gripper and micropart regions respectively when they are near and the computed Z_{io} is 1.23 mm.

IV. SHAPE ESTIMATION FROM FOCUS

In this task, we try to reconstruct the 3D shape of the microobject using focus information. Here, instead of computing and maximizing the sharpness of entire region, we maximize the sharpness of each pixel of the ROI in its local neighborhood. The neighborhood size is preselected. In this process, the first step is to perform autofocusing on the global scene in order to acquire a sharp image to compute the ROI. Unlike for depth estimation where a global ROI has been used for each part, multiple ROIs are determined to increase the accuracy of reconstruction. For this task, the ROIs are computed by applying the watershed transformation. Fig. 12 shows the computed ROI demonstrating the segmented gripper fingers from platform surface. Next, multiple images are acquired by varying the electronic working distance. Fig. 13 shows two images out of a series of 10 images acquired with a varying working distance of $50\ \mu\text{m}$. The acquired images

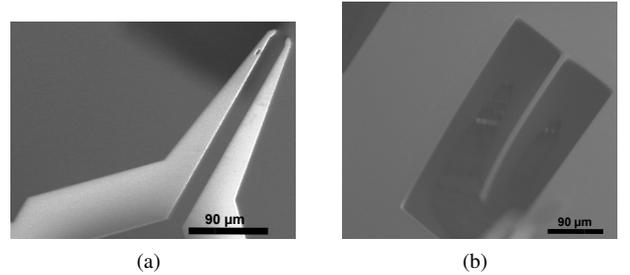


Fig. 10. Focused ROI of (a) FT G32 microgripper (b) silicon microparts.

Algorithm 1 Visual servoing-based autofocusing in SEM.

- 1: Choose initial value for α $\triangleright \alpha$ is a positive value used with gain
 - 2: Acquire initial image at focus step $F = 10$
 - 3: Compute S_0 using (4) $\triangleright S_0$ is initial sharpness score
 - 4: $F_{prev} \leftarrow F$
 - 5: **while true do** \triangleright Loop until best focus is found
 - 6: Set new focus F_{new} \triangleright For first iteration $F_{new} = 11$
 - 7: Acquire current image
 - 8: Compute S_{curr} using (4) \triangleright Current sharpness score
 - 9: Compute gain $\lambda \leftarrow \alpha \left(\frac{S_0}{S_{curr}} \right)^2$
 - 10: Compute cost $C \leftarrow \lambda \left(\frac{S_{curr}}{\|S_{curr}\|} \right)$ \triangleright Here, cost is to be maximised
 - 11: Compute secondary task $T_2 \leftarrow \left(\frac{\partial S_{curr}}{\partial Z} \right) L_Z$ \triangleright Stop criteria; L_Z is interaction matrix
 - 12: Obtain new focus step $F_{new} \leftarrow F_{prev} + C + T_2$
 - 13: **if** $T_2 < -0.001$ **then**
 - 14: $bestfocus \leftarrow F_{prev}$
 - 15: **return** best focus
 - 16: **end if**
 - 17: $F_{prev} \leftarrow F_{new}$
 - 18: **end while**
-

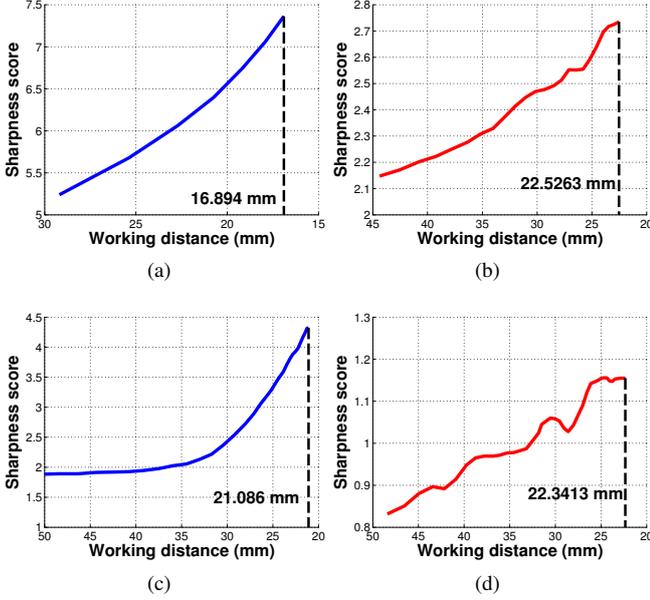


Fig. 11. Results of autofocusing at different regions. (a) Gripper (b) micropart heights when they are far from each other. (c) Gripper (d) micropart heights when they are near.

are indexed with their corresponding focus step. Next, for each pixel point (u, v) in the ROI for each image, normalized variance has been computed in a local neighborhood of size $n \times n$ using (7). In this work, a neighborhood of size 5×5 has been used.

$$S_{local}(K) = \frac{1}{n^2} \frac{1}{\bar{\mu}} \sum_{u=u-\frac{n-1}{2}}^{u+\frac{n+1}{2}} \sum_{v=v-\frac{n-1}{2}}^{v+\frac{n+1}{2}} (I(u, v) - \bar{\mu})^2 \quad (7)$$

where, $K = 1 \dots k$ number of images, $\bar{\mu}$ is the local neighborhood pixels mean, I is the grey level intensity of the pixel.

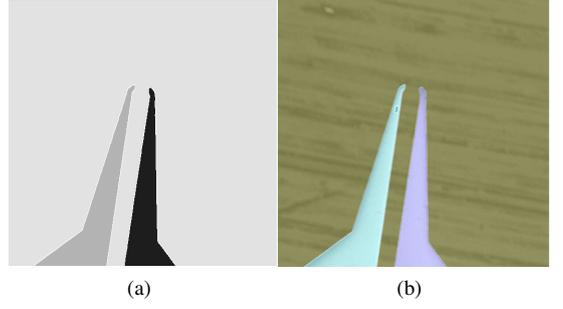


Fig. 12. Segmented regions of the (a) gripper fingers and (b) platform surface formed after applying watershed transformation.

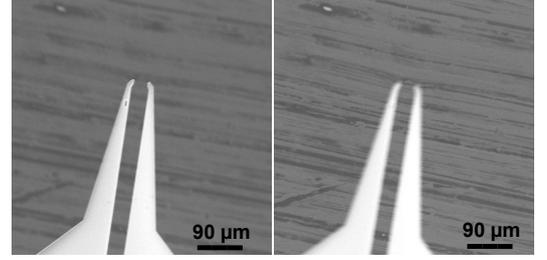


Fig. 13. Images acquired for reconstruction with varying working distance.

Now, the image in which the S_{local} is maximum for the current point is determined. The height of the point is then computed using the index (focus step) of the particular image from (3). As the entire process of reconstruction is depended on the sharpness information, the scene to be reconstructed should contain good textures. Apart from that, some pixel points may lead to erroneous heights due to the quality of image and step between the working distances. These heights are processed using a median filter. Fig. 14(a) shows the initial depth map and Fig. 14(b) shows the pixels remained on the gripper surface after filtering. Fig. 14(c) and 14(d) show the images formed after surface interpolation and overlaying the original texture. The developed method is also validated using a gold on carbon specimen. Fig. 15(a) shows the sample stub for JEOL SEM and the region pointed by red rectangle is used for reconstruction. Fig. 15(b) and Fig. 15(c) show the reconstructed image and the images formed after surface interpolation respectively. Fig. 15(d) shows the texture overlayed image.

V. CONCLUSION AND FUTURE WORK

Microhandling of parts with gripper includes a gripping phase and a release phase. In both phases, the detailed knowledge of the inter-object depth (between gripper and the part) and structure of the scene is crucial. The paper develops a solution to these problems in the context of a SEM depending solely on the image focus information. The inter-object depth has been estimated by finding the point of maximum sharpness separately for each object present in the scene. It uses the visual servoing-based autofocusing technique for this purpose. The results obtained with a tungsten gun SEM for microgripper and a silicon microstructure are convincing: fast and accurate. The second problem of shape reconstruction has been performed by acquiring multiple images and by estimating the depth of each pixel point. It has been validated on the scenes containing microgripper and gold on carbon stub.

Algorithm 2 Shape reconstruction from focus in SEM.

```

1: Perform autofocus on global scene and get working distance  $Z$ 
2: Compute multiple ROIs
3: for  $Z = Z - 0.25 : 0.05 : Z + 0.25$  do
4:   Acquire image and index it with the corresponding focus step
5:   Save the image in stack
6: end for
7: Select the neighborhood size,  $n$ 
8: for ROI = 1: total number of ROI do
9:   for  $K=1$ : total number of images do
10:    for  $u=1$ : image width do
11:     for  $v=1$ : image height do
12:      Compute  $S_{local}$  of each pixel  $(u, v)$  in ROI
using (7)
13:      Get the image index with maximum  $S_{local}$ 
14:      Assign  $Z$  using (3)
15:    end for
16:  end for
17: end for
18: end for
19: Get a reconstructed image adding all ROI pixel depths
20: Perform median filtering and surface interpolation
21: Get final reconstructed image

```

The future work will focus on validating the developed methods under a more advanced field emission gun SEM. It also includes the improvement of the approaches in terms of accuracy.

REFERENCES

- [1] S. Roy, J. Meunier, A. Marian, F. Vidal, I. Brunette, and S. Costantino, "Automatic 3d reconstruction of quasi-planar stereo scanning electron microscopy (sem) images*," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2012, pp. 4361–4364.
- [2] E. Ponz, J. L. Ladaga, and R. D. Bonetto, "Measuring surface topography with scanning electron microscopy. i. ezeimage: a program to

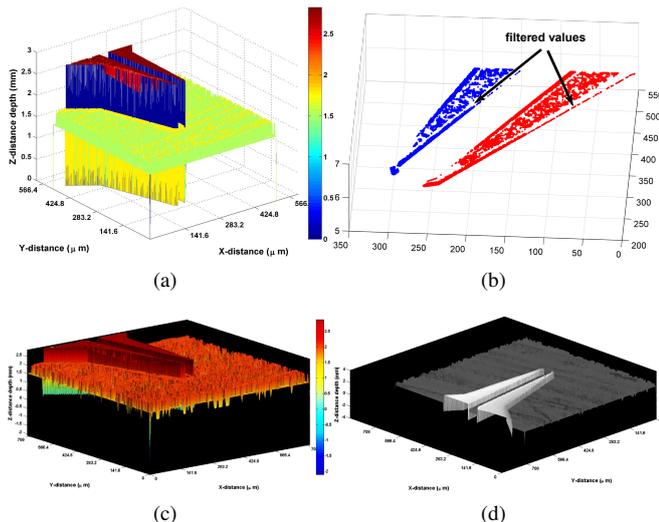


Fig. 14. Reconstructed images of gripper scene. (a) Initially estimated depth map. (b) Filtered surface of gripper fingers (white holes are filtered values). (c) Reconstructed scene after filtering and surface interpolation. (d) Image formed by overlaying the original texture.

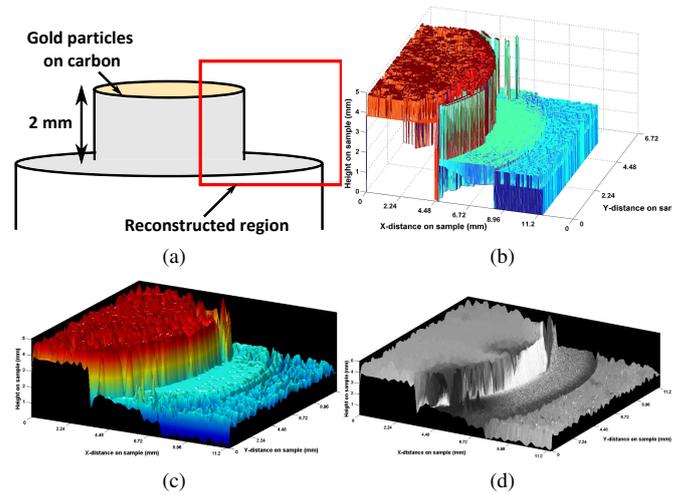


Fig. 15. (a) Gold on carbon sample stub for JEOL SEM (b) reconstructed image (c) Image formed after surface interpolation (d) image formed by overlaying the original texture.

- obtain 3d surface data," *Microscopy and Microanalysis*, vol. 12, no. 2, pp. 170–177, 2006.
- [3] D. Janová and J. Jan, "Robust surface reconstruction from stereo sem images," in *Computer Analysis of Images and Patterns*. Springer, 1995, pp. 900–905.
- [4] N. Marturi, S. Dembélé, and N. Piat, "Performance evaluation of scanning electron microscopes using signal-to-noise ratio." in *International Workshop on MicroFactories*, 2012, pp. 1–6.
- [5] M. Jähnisch and S. Fatikow, "3-d vision feedback for nanohandling monitoring in a scanning electron microscope," *International Journal of Optomechatronics*, vol. 1, no. 1, pp. 4–26, 2007.
- [6] R. Tunnell and S. Fatikow, "3d position detection with an fib-sem dual beam system," in *Proceedings of the 10th WSEAS international conference on communications, electrical & computer engineering, and 9th WSEAS international conference on Applied electromagnetics, wireless and optical communications*. World Scientific and Engineering Academy and Society (WSEAS), 2011, pp. 128–133.
- [7] V. Eichhorn, S. Fatikow, T. Wich, C. Dahmen, T. Sievers, K. N. Andersen, K. Carlson, and P. Bøggild, "Depth-detection methods for microgripper based cnt manipulation in a scanning electron microscope," *Journal of Micro-Nano Mechatronics*, vol. 4, no. 1-2, pp. 27–36, 2008.
- [8] F. Schmoekkel, H. Worn, and M. Kiefer, "The scanning electron microscope as sensor system for mobile microrobots," in *8th IEEE International Conference on Emerging Technologies and Factory Automation*, vol. 2. IEEE, 2001, pp. 599–602.
- [9] D. Chen, A. Miyamoto, and S. Kaneko, "Robust surface reconstruction in sem with two bse detectors," in *Mechatronics (MECATRONICS), 9th France-Japan & 7th Europe-Asia Congress on and Research and Education in Mechatronics (REM)*. IEEE, 2012, pp. 64–70.
- [10] G. Fortier, B. Tamadazte, S. Dembélé, and N. Le Fort-Piat, "Estimating the 3d orientation of a microgripper by processing the focus data from the images delivered by a videomicroscope," in *International Workshop on Robotic and Sensors Environments. ROSE 2008*. IEEE, 2008, pp. 58–63.
- [11] N. Marturi, B. Tamadazte, S. Dembélé, and N. Piat, "Visual servoing-based approach for efficient autofocus in scanning electron microscope," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, in press.
- [12] Y. Sun, S. Duthaler, and B. Nelson, "Autofocusing algorithm selection in computer microscopy," in *IEEE/RSJ International Conference on Intelligent Robots and Systems.(IROS 2005)*. IEEE, 2005, pp. 70–76.