Image-Guided Nanopositioning Scheme for SEM.

Naresh Marturi, Brahim Tamadazte, Sounkalo Dembele, Nadine Piat

Image-Guided Nanopositioning Scheme for SEM

Naresh Marturi, Brahim Tamadazte, Sounkalo Dembélé and Nadine Piat

***Abstract*—Positioning of micro-nanoobjects inside a scanning electron microscope (SEM) for manipulation is a key and challenging task to perform. Often it is performed by skilled operators via teleoperation, which is tedious and lacks repeatability. In this paper, rendering this task as an image-guided problem, we present a frequency domain scheme for automatic control of positioning platform movements. The designed controller uses the relative global image motion computed using the frequency spectral information of the images as visual signal and can provide control upto 5 degrees of freedom. The proposed approach is validated in simulations as well as experimentally using a high resolution piezo-positioning platform mounted inside SEM vacuum chamber. The obtained results quantify the performance of proposed nanopositioning scheme.**

***Note to Practitioners*—The main motivation behind this paper comes from the very need for automatic positioning of objects inside a SEM to perform dynamic analysis and structural characterization. Mostly, the positioning tasks are exhibited by skilled operators via teleoperation. Nevertheless, it is still a difficult task to repeat and hence automatic strategies are indispensable. This can be tackled upto an extent using microscopic vision information. However, the regular vision-guided strategies with integrated feature tracking are hard to use with SEM due to multiple instabilities associated with the imaging process. To address this issue, this paper presents an image frequency-based positioning stage controller that does not require any visual tracking and is capable of dealing with electronic images provided by SEM for automatic nanopositioning. The presented results illustrate the capability of the method in handling various perturbations and demonstrate its performance in terms of accuracy, robustness and repeatability. Due to the existence of orthographic projection, the proposed method is limited to control depth displacements. This can be resolved by combining it with visual servoing-based autofocus methods.**

***Index Terms*—Scanning electron microscope (SEM), Nanopositioning, visual control, Fourier transform, motion control.**

I. Introduction

With a rapid development in micro-nanoscale technologies in the last couple of decades, nanomanipulation has gained a significant research interest. It has a variety of applications in various industrial and biomedical domains and also used in different scientific works to perform dynamic analysis and characterization of object's structural, mechanical, electrical or optical properties [1]–[3]. Besides, manipulation of nanometric objects also benefit in building complex nanoelectromechanical systems (NEMS) [4]. The consequence of this strong demand is the emergence of an active research area concerning the development of automatic technologies within the micro-nanoscale.

So far, a great progress has already been realized in the development of nanomanipulation systems [5], microgrippers [6], nanopositioning systems [7] as well as robust control strategies [8]–[10]. However, the major concern is about the repeatability and accuracy of the tasks being performed. Since human handling is not a feasible option at this small scale, more robust automatic handling tools and techniques are indispensable. In order to perform an automatic nanomanipulation task, the basic operations include identification, positioning and handling of the objects with nanometric resolution. Out of all, positioning is more challenging at this particular scale mainly due to the lack of accurate and effective feedback information. This problem can be resolved by considering an imaging system like SEM, which is capable of producing near real-time images with high resolution and at high magnifications [11]. Most of the nanopositioning tasks using a SEM are performed by skilled operators who use a joystick device to control the positioning platform's movement. However, due to the presence of random image noise, drift and instabilities in the SEM electron column, it is always a difficult and tiresome task to perform [12]–[14]. This can be tackled up to an extent by automating the positioning process using *visual servoing* strategies, which are based on minimizing the error between tracked current and reference image features (e.g., edges, corners) [15]. This tracking process plays a vital role in the development of visual control laws.

Previously, computer vision techniques such as active contours and cross-correlation-based template matching have been successfully demonstrated for tracking stationary micro-objects in semi-automated positioning tasks inside a SEM [11], [16]. However, they are affected by heavy deformations and large displacements observed with SEM imaging. A new position tracking system that controls the SEM scanning system has been demonstrated in [17]. It determines an object's position using few line scans instead of acquiring a global image. A comparison of different tracking frameworks for SEM was presented in [18]. Besides 2D tracking, a 3D model-based approach for tracking the micro-objects within a region of interest was proposed in [19]. Nevertheless, acquiring images to use them with visual servoing inside a SEM is always a challenging task due to the addition of random noise during various stages of image acquisition (especially with fast raster rates) and instability in contrast and brightness levels. Moreover, the pixels are acquired one at a time slowing down the overall acquisition rate. However, the latter can be dealt by considering latency estimation schemes with regular visual controllers [9]. Apart from this, when the objects are in motion, the images either contain multiple occurrences of the object or discontinued edges due to the sequential

This work is conducted with a financial support from the project NANOROBUST (ANR-11-NANO-006) funded by 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. Part of this work is published in the proceedings of the IEEE ICRA 2014.

The authors are with the department of Automatic Control and MicroMechatronic Systems (AS2M), FEMTO-ST Institute, 25000 Besançon, France (e-mail: brahim.tamadazte@femto-st.fr).

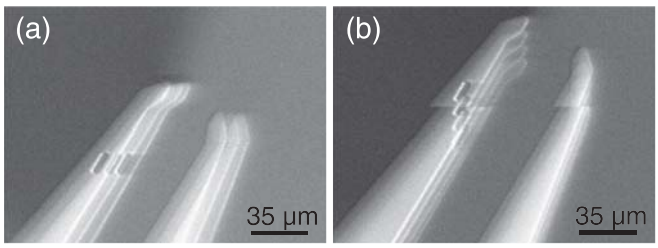


Fig. 1. Secondary electron images of a microgripper fingers obtained when the gripper is in motion. Multiple occurrences of the object and discontinued edges can be seen. Used scan speed is 180 $ns/pixel$.

raster scanning of the surface (see Fig. 1). This phenomenon emanates the difficulty in applying visual tracking algorithms.

To this extent, we develop a tracker-less nanopositioning scheme in this work. Recent developments have shown that the usage of global image information like pixel gray level intensities [20], image histograms [21], spatial sampling kernels [22], etc., for visual servoing can increase the robustness in minimizing the task error, due to the redundant information. Relying on this concept, we develop a new vision-guided nanopositioning scheme that uses image frequency spectral information. In contrast to the available global information-based techniques, which are of type image-based visual servoing, our approach is a position-based visual servoing scheme that relies on the relative image motion. Based on this, we realize an automatic nanopositioning task inside a SEM, which circumvents the feature tracking step. Mostly, Fourier-based methods are widely used for planar image registrations due their robustness to local image variations and various disturbances [23]. The main basis of these methods is the phase correlation function, which enables to estimate the transformation between two images. Previously, with SEM imaging, such type of methods were used to compensate the time-varying distortion, *drift* [14], [24]. Recently in [25], we have demonstrated how a simple Fourier-based method can be used to control 2 degrees of freedom of a positioning platform, where we have also compared its performance with the available *direct visual servoing* technique [20]. In this paper, we extend our previous work in many ways. First, the method has been enhanced to perform a 5 degrees of freedom positioning task. The translation motion estimation has been improved by moving the relative shift computation from pixel level to sub-pixels. Moreover, we integrate a two-way rotation estimation to augment the capabilities of the approach. Also, by using an optimised control law, the convergence of the final task error has been improved. Entirely, new experiments are reported to demonstrate the performance of proposed method at challenging conditions like, variable image intensities, artificial and high quality images as desired images and using samples with repetitive patterns. Furthermore, we also demonstrate the method's performance in terms of computational efficiency in comparison with the regular spatial processing methods.

In this paper, we render the nanopositioning task as a pure image dependent problem and develop a closed loop scheme for controlling the positioning platform's degrees of freedom inside a SEM. We first analytically develop the Fourier-based control law in Section II. The underlying idea is to use the image motion obtained using the Fourier spectral information. A two stage method has been presented for computing the rotations. In the first stage, the rotation around Z has been computed and the second stage includes the computation of out-of-plane rotations (around X and Y) using spherical Fourier transformations. 2D translations with sub-pixel accuracy are computed using the normalized phase correlation technique. Due to the existence of orthographic projection, the proposed method cannot control Z displacements. Nevertheless, this can be accomplished using visual servoing-based autofocus shown in [28]. Used experimental set-up is described in Section III and the experiments performed at various experimental conditions are reported in Section IV. We also provide a supplementary video illustrating positioning results inside SEM. An additional advantage associated with the method is that the positioning errors are corrected in real-time, which also corrects the induced drift distortions.

II. Fourier-based Visual Control for Nanopositioning

In this section we present a new Fourier-based visual servoing scheme for nanopositioning that does not require any locally obtained image measurements. One reason for choosing frequency domain is that with SEM imaging, the brightness and contrast are not constant and it has been known that Fourier domain techniques are robust to these variations and noise [23]. Here, we first present the motion estimation in images, which is then used to design positioning controller.

A. Motion Estimation Basics

Let $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ with $\boldsymbol{x} \doteq [x\ y\ z]^\top$ and $z = 1$ be two different images of same scene with size $M \times N$. These are related by

$$g(\boldsymbol{x}) = f\left(\boldsymbol{R}(\boldsymbol{x}) + \boldsymbol{\Delta x}\right) \tag{1}$$

where, $\boldsymbol{R}_{(\alpha,\beta,\gamma)}(\boldsymbol{x}) = R_x(\alpha)R_y(\beta)R_z(\gamma) \in \boldsymbol{SO}(3)$ is the rotation matrix and $\boldsymbol{\Delta x} = (\boldsymbol{\Delta}x, \boldsymbol{\Delta}y, \boldsymbol{\Delta}z) \in \mathbb{R}^3$ is the translation vector. Since z is constant in this case *i.e.*, $\boldsymbol{\Delta}z = 0$ and we assume image dimensions are equal *i.e.*, $M = N$ (square image), the 2D Fourier transforms (FT) of images can be written as

$$\mathcal{F}(\boldsymbol{k}) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(\boldsymbol{x}) e^{-j2\pi\left\{\frac{\boldsymbol{kx}}{M}\right\}} \tag{2}$$

where, $\boldsymbol{k} = [u\ v]^\top$ are the frequency domain coordinates. To estimate the motion, it is required to compute the relation between Fourier transforms. If we apply the transformation shown in (1) to (2) *i.e.,* if the image $f(\boldsymbol{x})$ is translated and rotated, we get

$$\begin{aligned}\mathcal{F}(\boldsymbol{k}) &= \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f\left(\boldsymbol{R}(\boldsymbol{x}) + \boldsymbol{\Delta x}\right) e^{-j2\pi\left\{\frac{\boldsymbol{kR}(\boldsymbol{x})+\boldsymbol{k\Delta x}}{M}\right\}} \\ &= \left(\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} g(\boldsymbol{x}) e^{-j2\pi\left\{\frac{\boldsymbol{kR}(\boldsymbol{x})}{M}\right\}}\right) e^{-j2\pi\left\{\frac{\boldsymbol{k\Delta x}}{M}\right\}}\end{aligned} \tag{3}$$

As $\boldsymbol{R}$ is rotational matrix, *i.e.,* orthogonal *i.e.*, $\boldsymbol{R}^{\top} = \boldsymbol{R}^{-1}$, one can write (3) as

$$\mathcal{F}(\boldsymbol{k}) = \left(\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(\boldsymbol{x}) e^{-j2\pi \left\{ \frac{\boldsymbol{k}(\boldsymbol{R}^{-1})^{\top}(\boldsymbol{x})}{M} \right\}} \right) e^{-j2\pi \left\{ \frac{\boldsymbol{k}\boldsymbol{\Delta}\boldsymbol{x}}{M} \right\}} \quad (4)$$

Now, if we apply rotation to the Fourier transform, we get

$$\mathcal{F}(\boldsymbol{R}\boldsymbol{k}) = \left(\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(\boldsymbol{x}) e^{-j2\pi \left\{ \frac{\boldsymbol{k}\boldsymbol{x}}{M} \right\}} \right) e^{-j2\pi \left\{ \frac{\boldsymbol{k}\boldsymbol{\Delta}\boldsymbol{x}}{M} \right\}} \quad (5)$$

Finally, using (5), the relation between the Fourier transforms can be deduced as

$$\mathcal{G}(\boldsymbol{k}) = \mathcal{F}(\boldsymbol{R}\boldsymbol{k})\, e^{j2\pi \left\{ \frac{\boldsymbol{k}\boldsymbol{\Delta}\boldsymbol{x}}{M} \right\}} \quad (6)$$

It can be seen from (5), the exponential term outside the brackets contains translation component $\boldsymbol{\Delta}\boldsymbol{x}$ independent of the original spatial image information $\boldsymbol{x}$ and the rotation in spatial domain is commuted to the frequency domain as seen in (6). Such a kind of relation helps in decoupling translation and rotation motions, which can be accomplished by computing magnitude and phase spectra of images *i.e.,*

$$\begin{aligned} \text{magnitude:} \quad & \mid \mathcal{G}(\boldsymbol{k}) \mid = \mid \mathcal{F}(\boldsymbol{R}\boldsymbol{k}) \mid \\ \text{phase:} \quad & \angle\mathcal{G}(\boldsymbol{k}) = \angle\mathcal{F}(\boldsymbol{k}) - 2\pi \left\{ \frac{\boldsymbol{k}\boldsymbol{\Delta}\boldsymbol{x}}{M} \right\} \end{aligned} \quad (7)$$

An important point to be noted is that the DFT assumes the original image is periodic. This affects in producing discontinuities especially at the image boundaries, which in turn leads to unwanted artifacts in the Fourier spectrum. In order to minimize this effect, in this work, the input images are pre-processed using a *Hann window*. Following the general rule, we first compute the rotation and then estimate the translation.

B. Rotation Estimation

Generally, rotations in a space can be represented using the axis of rotation $\mathbf{u} = (u_x, u_y, u_z)$ and an angle θ, from which $\boldsymbol{R}$ can be written as

$$\boldsymbol{R} = \mathbb{I}\cos\theta + \sin\theta[\mathbf{u}]_{\times} + (1 - \cos\theta)\mathbf{u} \otimes \mathbf{u} \quad (8)$$

whose eigenvalues are $\lambda_1 = 1$, $\lambda_2 = e^{i\theta}$ and $\lambda_3 = e^{-i\theta}$ and $\mathbb{I}$ is the identity matrix. Up on computing the eigen-vector $\boldsymbol{\nu}$ corresponding to the real eigenvalue λ_1, axis of rotation can be computed as $\frac{\boldsymbol{\nu}}{|\boldsymbol{\nu}|}$ and simultaneously the angle can be deduced using λ_2 and λ_3. In our case, there are two possible stages of rotations: planar (yaw) and out-of-plane (roll and pitch). We first start by computing the planar rotation as its deduction is straight forward and can be directly estimated using the magnitude spectra.

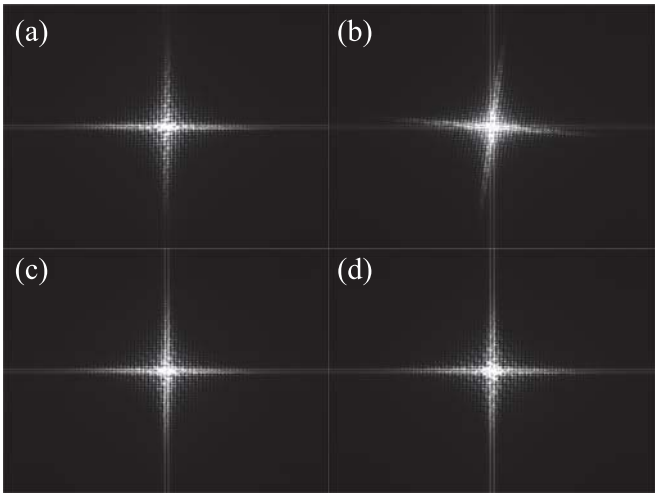


Fig. 2. Log-polar magnitudes for (a) reference image (b) rotated around Z (c) and (d) rotated around X and Y, respectively. In any case the angle of rotation is 3°. The magnitude difference between the reference and tilted images is very small and is not effective for angle estimation.

1) Estimating Planar Rotations: As in [26], planar rotations can be obtained by re-sampling the image magnitude spectra onto rectangular log-polar coordinates. If $g(x, y)$ is rotated by γ_0 (around Z) of $f(x, y)$, then their Fourier transforms are related by

$$\mathcal{G}(u, v) = \mathcal{F}(u\cos\gamma_0 + v\sin\gamma_0, -u\sin\gamma_0 + v\cos\gamma_0) \quad (9)$$

Let $\xi_1 = \mid \mathcal{F} \mid$ and $\xi_2 = \mid \mathcal{G} \mid$ be the magnitude spectra, then from (7)

$$\xi_2(u, v) = \xi_1(u\cos\gamma_0 + v\sin\gamma_0, -u\sin\gamma_0 + v\cos\gamma_0) \quad (10)$$

Now, the angle of rotation can be obtained by representing the magnitude spectra in log-polar form as shown in (11).

$$\xi_2(\rho, \gamma_0) = \xi_1(\rho, \gamma - \gamma_0) \quad (11)$$

where, ρ and γ are the radius and angle in polar coordinates. From (11) it can be seen that, rotation in spatial domain is a pure translational displacement in polar representation. Now, by computing the global correlation between $\xi_2(\rho, \gamma_0)$ and $\xi_1(\rho, \gamma - \gamma_0)$ provides a peak in the correlation surface, which corresponds to an estimate $\hat{\gamma} \in [0, \pi]$ to the angle of rotation.

$$\hat{\gamma} = \operatorname{argmax}\{\operatorname{corr}(\xi_1, \xi_2)\} \quad (12)$$

2) Estimating Out-of-Plane Rotations: By analysis (Fig. 2), the platform tilt angles have a minimal effect on the image magnitudes and are difficult to determine from log-polar magnitude correlation. Moreover, the existence of the orthographic projection in SEM [19] makes this process more challenging. In this case, the tilt angles can be coarsely estimated using image defocus information and spherical Fourier transform. For this purpose we assume that: 1) the primary beam tilt angle is 0°; 2) the positioning platform tilts around a common origin; 3) device working magnification is high ($> 2000\times$) *i.e.*, the depth-of-focus (DOF) is minimum enough[1] ($< 300nm$).

[1]This value has been computed for the used experimental SEM and may differ with other devices.

Since the DOF is small, any tilt in the platform results in the images with focus differences, which is the main basis of our method. The underlying idea is to obtain point cloud images using focus information, which are then represented as extended Gaussian images[2] (EGI) [27]. First, to build the point cloud, it is required to measure the depth of each image point, which is a time consuming process. To solve this, we divide the image into equal sized kernels. Making use of the projection model, depth for each individual kernel is estimated by linking the position corresponding to maximum focus that is computed using the visual servoing technique described in [28] with the SEM working distance. The defocus information $\mathbf{S}(K)$ has been computed using the normalized variance of kernel (with size $k \times k$) and is given by

$$\mathbf{S}(K) = \frac{1}{k^2}\frac{1}{\mu}\sum_{x=x-\left(\frac{k-1}{2}\right)}^{x+\left(\frac{k+1}{2}\right)}\sum_{y=y-\left(\frac{k-1}{2}\right)}^{y+\left(\frac{k+1}{2}\right)}\left(I(x,y)-\mu\right)^2 \quad (13)$$

where, $K = 1 \ldots \kappa$ are defined kernels, μ is the local kernel pixels mean, I is the gray level intensity of the pixel at (x, y).

Once the point cloud is computed for the kernel image, the local point normal are computed to construct the EGI. In this work, we use tools from PCL library [29] to perform this step. Once the EGI is constructed, spherical harmonics are used to estimate the rotations, which are performed in a single step. The spherical harmonics function $Y_{ml} : \mathbb{S}^2 \rightarrow \mathbb{C}$ with $\mathbb{S}^2$ is unit 2D sphere and $\mathbb{C}$ is complex number set is given by [30]

$$Y_{ml}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\left[\frac{(l-m)!}{(l+m)!}\right]}P_{lm}(\cos\theta)e^{(im\phi)} \quad (14)$$

where, (θ, ϕ) are spherical coordinates, P_{lm} is Legendre polynomial, $l > 0$ and $\mid m \mid < l$ are integers with l being the degree of spherical harmonics. Any square integrate function $f(\omega) \in \mathcal{L}^2$ lying on the unit sphere can be represented as the linear sum of the spherical Fourier transforms and spherical harmonics, and is given by

$$f(\omega) = \sum_{l\in\mathbb{N}}\sum_{m=-l}^{l}\mathcal{F}_{ml}Y_{ml}(\omega) \quad (15)$$

where, $\mathcal{F}_{ml}$ are spherical Fourier transforms. Now the goal is to obtain $\boldsymbol{R}$ between two spherical harmonics of images, which can be determined by spherical correlation. As explained, we choose Euler zyz convention for angles *i.e.*, $\boldsymbol{R}(\alpha,\beta,\gamma) \in \mathbf{SO}(3)$. Let $f(\omega)$ and $g(\omega)$ are two functions on sphere with $g(\omega) = \boldsymbol{R}f(\omega)$. The maximum value of spherical correlation of these functions provide an estimate to $\boldsymbol{R}$. This correlation $\mathcal{C}(\boldsymbol{R})$ can be given by

$$\mathcal{C}(\boldsymbol{R}) = \sum_{l\in\mathbb{N}}\sum_{m=-l}^{l}\sum_{p=-l}^{l}\mathcal{F}_{ml}\overline{\mathcal{G}_{pl}}\,\overline{D^{(l)}_{mpl}}(R) \quad (16)$$

where, $\overline{D^{(l)}_{mpl}}(R)$ is the complex conjugate of the Wigner D-function $D^{(l)}_{mpl}(\boldsymbol{R})$ given by

$$D^{(l)}_{mpl}(R) = e^{(-im\gamma)}P^{(l)}_{mp}(\cos\beta)e^{-ip\alpha}. \quad (17)$$

[2]EGIs can be represented as the surface orientation (spherical) histograms that are created by computing the normal of each point lying on the point cloud.

In (16), $\mathcal{F}$ and $\mathcal{G}$ are the spherical orthogonal Fourier transforms (SOFT) [30]. A more detailed derivations for (16) and (17) can be found in [30] and a library to compute these SOFT can be found from [31].

C. Translation Estimation

Once the rotation is computed, translation can be estimated using phase spectra as shown in (7). From (6), a shift in the spatial domain can be seen as multiplication in Fourier domain by a complex exponential function that contains the shift. The correlation between the spectra can be determined by performing the cross power spectrum $\Psi(\boldsymbol{k})$ given by

$$\Psi(\boldsymbol{k}) = \frac{\mathcal{F}(\boldsymbol{k})\overline{\mathcal{G}(\boldsymbol{k})}}{\mid \mathcal{F}(\boldsymbol{k})\overline{\mathcal{G}(\boldsymbol{k})} \mid} = e^{-j2\pi\left\{\frac{\boldsymbol{k}\boldsymbol{\Delta x}}{M}\right\}} \quad (18)$$

In (18), $\overline{\mathcal{G}(\boldsymbol{k})}$ is the complex conjugate of $\mathcal{G}(\boldsymbol{k})$. Here, the cross power spectrum is normalized in order to compensate the intensity variations. Now, (18) can be solved for overall translation $\boldsymbol{\Delta x} = (\Delta_x, \Delta_y)$ in two ways. The first one is to solve directly in the Fourier domain using a coordinate system containing two frequency axes and one phase difference axis. The slopes produced by $\frac{\boldsymbol{k}\boldsymbol{\Delta x}}{M} = 0$ provides (Δ_x, Δ_y). However, it is computationally complex. The other method that is used in this work is to find the inverse FT of (18) that results in a Dirac delta function given by

$$\mathcal{D}(\boldsymbol{\Delta x}) = \mathcal{F}^{-1}(\Psi(\boldsymbol{k})) \quad (19)$$

The obtained function contains a peak at the point of motion, whose coordinates provide an estimate to the displacement

$$\widehat{\boldsymbol{\Delta x}} = argmax\left\{\mathcal{D}(\boldsymbol{\Delta x})\right\} \quad (20)$$

where, $\widehat{\boldsymbol{\Delta x}}$ is the estimated translation vector.

Nevertheless, when dealing with discrete images, the maximum value of the peak is the closest estimate of the integer displacement *i.e.*, the estimated translations are at pixel resolution. In order to improve the accuracy of the positioning task, it is highly required to estimate these translations with sub-pixel accuracy. To do this, we first down-sample the images by factors d_1 and d_2 along X and Y axes, respectively as explained in [23]. Let $f_d(\boldsymbol{x})$ and $g_d(\boldsymbol{x})$ are down-sampled images where, $g_d(x, y) = f_d(x - \Delta_{sx}, y - \Delta_{sy})$ with Δ_{sx} and Δ_{sy} being sub-pixel displacements. In this case, the Dirichlet function *i.e.*, the inverse FT of the cross power spectrum is

$$\mathcal{D}(\Delta_{sx},\Delta_{sy}) = \frac{1}{M^2}\left(\frac{\sin\left\{\pi(d_1+\Delta_{sx})\right\}}{\sin\left\{\frac{\pi}{M}(d_1+\Delta_{sx})\right\}}\right)\left(\frac{\sin\left\{\pi(d_2+\Delta_{sy})\right\}}{\sin\left\{\frac{\pi}{M}(d_2+\Delta_{sy})\right\}}\right) \quad (21)$$

The peak position of the above function (21) provides translations $(\widehat{\Delta_{sx}}, \widehat{\Delta_{sx}})$ in sub-pixels.

$$(\widehat{\Delta_{sx}}, \widehat{\Delta_{sy}}) = \text{argmax}\left\{\mathcal{D}(\Delta_{sx},\Delta_{sy})\right\} \quad (22)$$

Finally, from the obtained translations, the centered translations providing the direction of the motion are calculated using (23) and (24).

$$\tau_x = \begin{cases} \frac{M}{2} - \widehat{\Delta_{sx}} & \text{if } \widehat{\Delta_{sx}} > \frac{M}{2} \\ \widehat{\Delta_{sx}} & \text{elsewhere} \end{cases} \tag{23}$$

$$\tau_y = \begin{cases} \frac{M}{2} - \widehat{\Delta_{sy}} & \text{if } \widehat{\Delta_{sy}} > \frac{M}{2} \\ \widehat{\Delta_{sy}} & \text{elsewhere} \end{cases} \tag{24}$$

Similarly, the translation along Z can also be estimated using Fourier transforms. However, we constrain this motion in this work, as it cannot be controlled due to the orthographic projection. A visual servoing method for Z control is presented in [32].

D. Analysis of Blur and Noise

In practical scenarios with SEM, image acquisition process is affected by different degradations such as the addition of random noise, blur, intensity variations, etc. Previously in [12], we show that the dominant amount of secondary electron image noise follows Gaussian statistics and can be filtered during image recording process. However, it is worth analysing the effect of noise on motion estimation. Let $f(\boldsymbol{x})$ is degraded by blur and random noise $n(\boldsymbol{x})$. It can be modelled as

$$f'(\boldsymbol{x}) = f(\boldsymbol{x}) * h(\boldsymbol{x}) + n(\boldsymbol{x}) \tag{25}$$

where, $f'(\boldsymbol{x})$ is degraded image, $h(\boldsymbol{x})$ is a point spread function (PSF) indicating blur and $*$ is spatial convolution operator. We start by analysing the noise, which is generally associated with the high frequency components. Since the main basis of our motion estimation is component correlation, the amount of correlated noise can be negligible for images with considerable level of signal to noise ratio (SNR). However, if the SNR is low, there might be a significant affect on the observed correlation peak. This has been analysed visually using artificially generated images with varying noise and is illustrated in Fig. 3. From analysis it can be seen that the size of the peak reduced for correlation of noisy images (Fig. 3(d)) when compared with the noise free one (Fig. 3(b)). This phenomenon can be reduced by filtering the images in spatial domain. In this work, we monitor the image SNR at the time of acquisition using our quality estimation method shown in [12]. The images with low SNR (< 15 dB) are subsequently filtered using a Gaussian filter of size 5×5. Moreover, to reduce the noise added due to charge accumulation, the sample surface is connected to the mounting plate by means of a conductive tape.

Now, by neglecting the noise term in (25), the corresponding Fourier transform is given by

$$\mathcal{F}'(\boldsymbol{k}) = \mathcal{F}(\boldsymbol{k}).\mathcal{H}(\boldsymbol{k}) \tag{26}$$

It is worth noting that the convolution now becomes simple multiplication. The phasor form of (26) is

$$\mathcal{F}'(\boldsymbol{k}) = \mid \mathcal{F}(\boldsymbol{k}) \mid e^{\left\{-j2\pi\left\{\frac{\boldsymbol{k}\boldsymbol{x}}{M}\right\}\right\}} \tag{27}$$

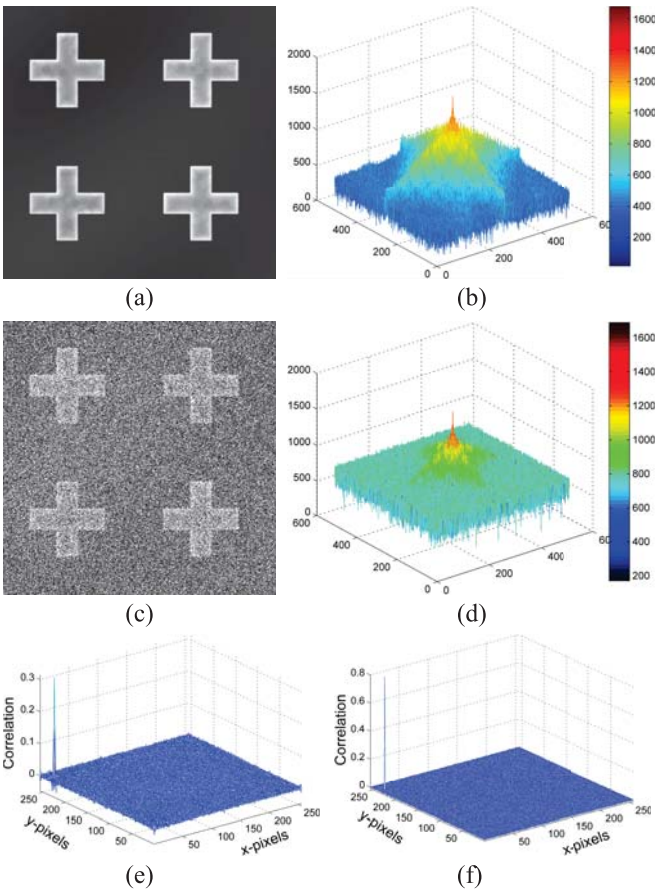


Fig. 3. Images illustrating the affect of noise. (a) and (c) are artificially generated images with varying noise. (a) is noise-free image and (c) is degraded by a Gaussian noise of $\sigma = 0.5$. (c) and (d) are Fourier spectra of (a) and (c), respectively. (e) is the phase correlation peak obtained in case of noise-free images and (f) is for noisy images with $\sigma = 0.5$. One of the images is translated by 10 pixels along X and 20 pixels along Y to perform correlation.

Here after, for simplicity we represent the exponential part with $\Phi_{f'}(\boldsymbol{k})$. By normalizing (27) we will be left out with the phase part *i.e.,*

$$\frac{\mathcal{F}'(\boldsymbol{k})}{\mid \mathcal{F}(\boldsymbol{k}) \mid} = e^{-j\Phi_{f'}(\boldsymbol{k})} = e^{-j\{\Phi_f(\boldsymbol{k}) + \Phi_h(\boldsymbol{k})\}} \tag{28}$$

If we assume that the PSF is centrally symmetric, $\mathcal{H}(\boldsymbol{k})$ will be real with two probable values of its phase being $\{0 \vee \pi\}$. Now using the periodicity property[3], for any integer a we can write

$$\left[e^{-j\Phi_{f'}(\boldsymbol{k})}\right]^{2a} = e^{-j2a\Phi_f(\boldsymbol{k})}.e^{-j2a\Phi_h(\boldsymbol{k})} = \left[e^{-j\Phi_f(\boldsymbol{k})}\right]^{2a} \tag{29}$$

From (29), it can be noted that the image Fourier transform with even power *i.e.,* $e^{-j\Phi(\boldsymbol{k})2a}$ is invariant to the centrally symmetric blur (convolved with the original signal).

Apart from that, as the correlation peak is due to the phase difference at each component, its location is not affected by the low bandwidth noise. This enhances that the method is not affected by the contrast variations, since the image contrast

[3]Periodic property: $e^{-j\frac{2\pi}{N}(\boldsymbol{k}+\frac{N}{2})} = e^{-j\frac{2\pi}{N}\boldsymbol{k}}e^{-j\pi} = e^{-j\frac{2\pi}{N}\boldsymbol{k}}$

information is associated with the low frequency components (dc-component) and thus represent narrow bandwidth variations.

E. Control Scheme

For positioning the platform, a control scheme has been designed considering the translation and rotation motions computed from the previous steps as the observed visual features *i.e.*,

$$\mathbf{s}(t) = [\mathbf{s}_v, \mathbf{s}_\omega]^\top \tag{30}$$

where, $\mathbf{s}_v = [\tau_x \ \tau_y \ \tau_z]^\top$ are the estimated translations, and $\mathbf{s}_\omega = \mathbf{u}\hat{\theta}$ is the rotation feature with $\mathbf{u}$ being the axis and $\hat{\theta}$ being the estimated angle of rotation. This can simply be $[\alpha \ \beta \ \gamma]^\top$. Since Z is not used, we approximate it to be zero. By using such features, the resulting control can allow us to decouple the control of rotation and translation. Once visual features are obtained, the final objective is to drive the positioning stage to the desired location *i.e.*, $\mathbf{s}^* = [\mathbf{0}_{1\times5}]^\top$. Thus, the error to be regulated is given by

$$\mathbf{e} = \begin{bmatrix}\mathbf{e_v}\\ \mathbf{e}_\omega\end{bmatrix}_{5\times1} = \mathbf{s} - \mathbf{s}^* = \left[\tau_x \ \tau_y \ \mathbf{u}\hat{\theta}\right]^\top \tag{31}$$

If we consider the problem of error minimization as an optimization problem, the primary goal will be to minimize the cost function C given by

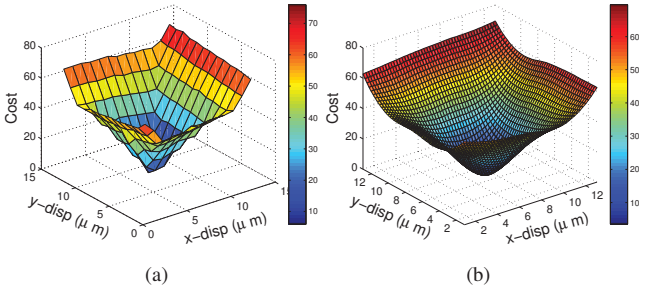
$$C = \mathbf{e}^\top\mathbf{e} = \tau_x^2 + \tau_y^2 + \hat{\theta}^2 \tag{32}$$

Similar to the conventional visual servoing approaches, the goal image $\mathbf{I}^*$ is defined by teaching, where the positioning stage is moved to a desired position and the image is acquired. This image remains as the reference image until the process is converged and its Fourier transformations will be used during the correlation process for estimating the motion. Once, the target is defined, the main goal of the controller is to regulate the cost function given by (32) from an unknown initial position. When C is minimum, the current position corresponds to the desired position. To visually reflect this cost function, a set of images are acquired by moving the platform around the target position. The cost is then computed offline using these images and is shown in the Fig. 4(a). For smooth convergence, the obtained cost function has been regularized using a B-spline function of order two as shown in Fig. 4(b). Upon observing the shape of cost function it is clear that the proposed method provides satisfactory convergence, which can be seen as a result of its robustness to intensity, brightness and contrast variations and image noise.

The positioning task can be accomplished by iteratively updating the velocity. This requires the relationship between time variation of visual features $\dot{\mathbf{s}}$ and the camera instantaneous velocity $\mathbf{v} = [\boldsymbol{\upsilon} \ \boldsymbol{\omega}]^\top$, which is given by

$$\dot{\mathbf{s}} = \begin{bmatrix}\dot{\mathbf{s}}_v\\ \dot{\mathbf{s}}_\omega\end{bmatrix} = \mathbf{L}_f \begin{bmatrix}\boldsymbol{\upsilon}\\ \boldsymbol{\omega}\end{bmatrix} \tag{33}$$

where, $\mathbf{L}_f$ is a 5×5 interaction matrix. As the motion estimation can be decoupled, we can have a decomposed interaction matrix *i.e.*, $\mathbf{L}_f = \begin{bmatrix}\mathbf{L}_v & \mathbf{0}\\ \mathbf{0} & \mathbf{L}_\omega\end{bmatrix}$ where $\mathbf{L}_v$ is a 2×2 matrix to link linear velocities and $\mathbf{L}_\omega$ is a 3×3 matrix to link angular velocities. Such type of strategy allows us to control the platform translation and rotation independently. In this work, we consider $\mathbf{L}_v = \mathbb{I}$ and $\mathbf{L}_w$ is given by (34) [33]



(a) (b)

Fig. 4. (a) Obtained cost (b) fitted cost using second order B-splines.

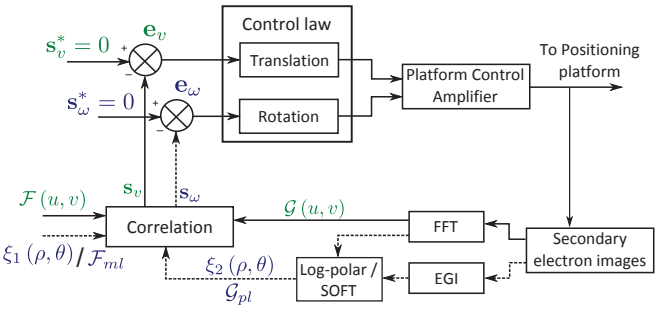


Fig. 5. Control block diagram depicting the decoupled Fourier-based nanopositioning scheme. (dotted lines represent rotation processing)

$$\mathbf{L}_w = \mathbb{I} - \frac{\theta}{2}[\mathbf{u}]_\times + \left(1 - \frac{\operatorname{sinc}\theta}{\operatorname{sinc}^2\frac{\theta}{2}}\right)[\mathbf{u}]_\times^2. \tag{34}$$

As a matter of stability, $\mathbf{L}_w$ can be singular only when $\theta = 2\pi$. To improve the convergence, in this work we implement an adaptive gain λ_a that varies with the variance of error $\sigma_\mathbf{e}$ as given by

$$\lambda_a = \frac{\sigma_\mathbf{e}}{\Lambda} \tag{35}$$

where, Λ is a positive value allows to adjust the gain. To improve the controller performance, a non-linear control inspired from *Levenberg-Maquardt* optimization technique has been used in this work. The final control law to regulate the positioning stage velocity is given by

$$\mathbf{v} = -\lambda_a \left(\mathbf{H} + diag(\mathbf{H})\right)^{-1} \mathbf{L}_f^\intercal \mathbf{e}\, {}^c\mathbf{V}_p \tag{36}$$

where, $\mathbf{H} = \mathbf{L}_f^\intercal\mathbf{L}_f$ is the Hessian matrix and ${}^c\mathbf{V}_p$ is the transformation matrix from camera frame $\mathcal{R}_c$ to positioning stage frame $\mathcal{R}_p$. The overall control block diagram for nanopositioning is shown in Fig. 5.

As the used positioning platform is driven by PZT actuators, it requires voltage commands for generating the displacements. For this purpose, the obtained velocities are converted to the platform displacements $\mathbf{d}_{(x,y)}$ using

$$\mathbf{d}_{(x,y)} = \mathbf{v}_{avg} t \tag{37}$$

where, $\mathbf{v}_{avg} = \frac{\mathbf{v}_0 + \mathbf{v}_{cur}}{2}$ is the average velocity, $\mathbf{v}_0$ is initial velocity, $\mathbf{v}_{cur}$ is the current velocity and t is the measured

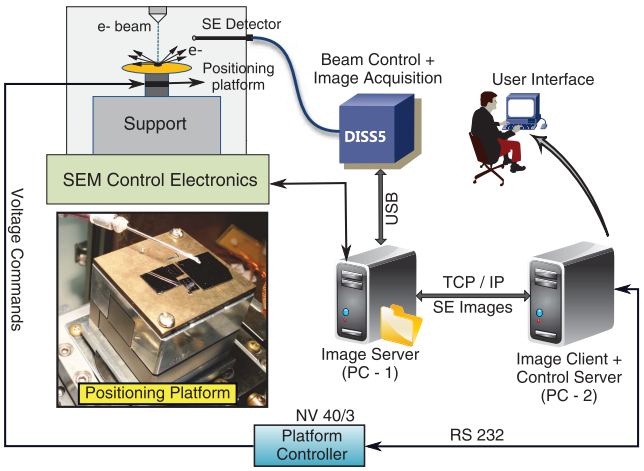


Fig. 6. Experimental set-up architecture (inset) mounted piezo-positioning platform inside the SEM vacuum chamber.

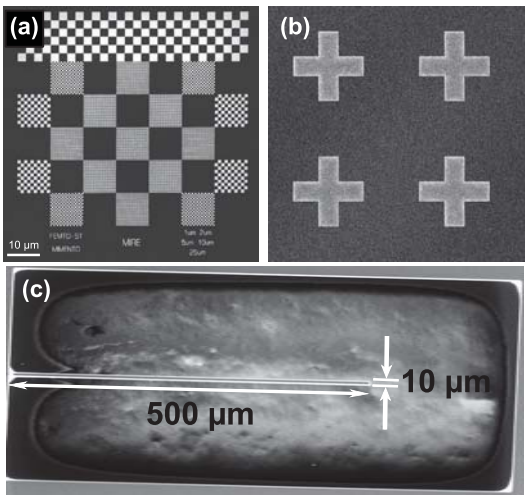


Fig. 7. Experimental samples: (a) microscale calibration sample fabricated at FEMTO-ST institute (b) artificially generated image of periodic crosses using Artimagen library (c) silicon micropart.

sampling time. Later, these displacements are mapped to the respective channel voltages using our displacement-voltage model explained in [25].

F. Stability analysis

The stability of the control law given by (36) can be discussed using Lyapunov's second method for stability. For this purpose, let us consider the following Lyapunov function candidate

$$\mathcal{L}(\varphi) = \frac{1}{2} \parallel \mathbf{e}(t) \parallel^2 \tag{38}$$

where, $\varphi = \mathbf{e}(t)$ and $\mathcal{L}(\varphi)$ is positive definite i.e., $\mathcal{L}(\varphi) \geq 0 \; \forall \; \varphi$. The first derivative of (38) is given by

$$\dot{\mathcal{L}} = \mathbf{e}^\top \dot{\mathbf{e}} \tag{39}$$

Using $\dot{\mathbf{e}} = \mathbf{L}\mathbf{v}$ we get,

$$\dot{\mathcal{L}} = \mathbf{e}^\top \mathbf{L}\mathbf{v} \tag{40}$$

By substituting for $\mathbf{v}$, we get

$$\dot{\mathcal{L}} = -\lambda \mathbf{e}^\top \mathbf{L}\mathbf{L}^\top \mathbf{e} = -\lambda \mathbf{e}^\top \mathbf{e} \tag{41}$$

$$\Rightarrow \quad \dot{\mathcal{L}} < 0 \quad \forall \, \mathbf{e} \neq 0, \; \lambda_a > 0 \tag{42}$$

From (42), for all $\lambda_a > 0$, the system is asymptotically stable in the task space.

III. Experimental Set-up

The experimental set-up used for this work is shown in the Fig. 6. It consists of a JEOL JSM 820 SEM, an image acquisition system (DISS5 from Point electronic GmbH), a 3 degrees of freedom (X, Y, Z) open-loop piezo positioning platform (TRITOR 100 from Piezosystem Jena GmbH) and two computers. The positioning platform is mounted inside SEM vacuum chamber and is controlled using a 3 channel piezo-controller NV 40/3. The maximum possible motion on all axes is up to 100 μm with a resolution of 0.2 nm. The primary computer (PC 1: Intel Pentium 4, CPU 2.24 GHz and 512 MB of RAM) running image server is connected to the SEM control electronics and imaging system. It is solely responsible for controlling the microscope and transfer secondary electron images to client. The work computer running image client (PC 2: Intel Core 2 Duo, CPU 3.16 GHz, and 3.25 GB of RAM) communicates with the primary one over TCP/IP. The control server which is also running on PC-2 uses the received images (via client) to compute control and to generate voltage commands to the platform controller via serial port (RS 232). Apart from this, a graphical user interface (APROS3) program (that runs from the work computer) has been developed to control the platform movements manually as well as to monitor the overall process. All the software programs are developed in C++ and matrix computations are performed using ViSP (Visual Servoing Platform) library [34]. 2D and spherical Fourier transforms are performed using FFTW [35] and SOFT [31] libraries, respectively.

IV. Experimental Validations

A. Task Description

Nanopositioning studied in this work uses two different real-world samples: a gold on silicon calibration sample containing multiple chessboard patterns (Fig. 7(a)) and silicon microparts (Fig. 7(c)) of dimensions $10 \times 500 \times 10$ μm^3. Additionally, we use artificially generated SEM images of periodic crosses by Artimagen library [36] for simulation experiments (Fig. 7(b)). Since, the experimental positioning platform does not features any rotational degrees of freedom, angular positioning is demonstrated by means of simulations. For planar translational positioning, the samples are arbitrarily placed on the piezo-positioning platform and the main objective is to position them automatically at the desired location by controlling the platform's movement. For demonstration, the desired location has been selected by an operator using APROS3. Once the desired location is selected, the platform moves back to its initial position from where the servoing starts. For all experiments, the SEM secondary electron images of size 512×512

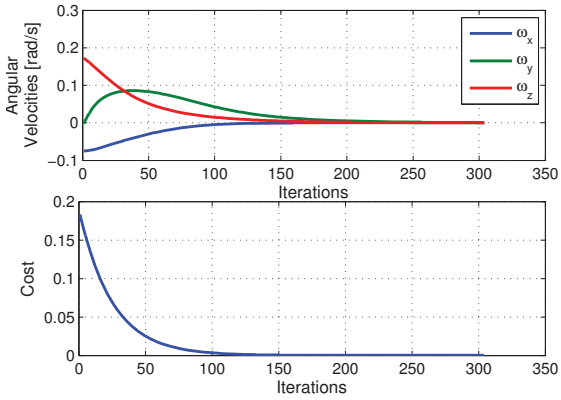


Fig. 8. Variation of (a) angular velocities (w_x, w_y, w_z) (b) cost during the task.

TABLE I
ESTIMATED ANGLES OF ROTATION FOR VARIOUS AXIS.

α (°)		β (°)		γ (°)	
True	Estimated	True	Estimated	True	Estimated
15	14.932	-15	-15.201	70	69.971
8	8.17	-8	-7.181	30	30.007
3	-2.646	3	2.83	-5	-4.98

pixels are used. Real-time images for translation control are generated using an acceleration voltage of 15 kV.

B. Simulation Results: Rotation Control

The first simulation experiment using artificially generated images with Artimagen is performed to validate the behavior of rotational control. The blur in the tilted images is modeled using a Gaussian function, based on which the focusing method explained in [28] is simulated to estimate the kernel depth. Initial rotations considered are $7°, 12°$ and $20°$ for α, β and γ, respectively. Fig. 8 show the evolution of angular velocities and the overall cost. From the obtained results, it can be seen that the algorithm converges successfully to the desired position and a smooth convergence of cost-function is observed.

The second experiments are performed to evaluate the accuracy in estimating the rotations. Table I summarizes the known and computed angles around various axes using the proposed method. From the obtained results, it can be seen that the planar rotation γ can be computed more accurately than α and β. This is due to the effect of orthographic projection and coarse estimation of kernel depth. For real-world scenarios, this phenomenon may raise additional challenges for visual control of rotations around X and Y.

C. Real-time Validations: Translation Control

In this subsection, the experimental results concerning the planar positioning task using the piezo-positioning platform are presented. Tests are performed at different operating conditions; optimal as well as the more challenging and special cases such as unstable imaging conditions, binary mask as desired image, high quality desired image and repeated textured object with high noise are presented here. Each experiment has been performed 10 times with different goal positions and out of them; three trials (minimum, medium and maximum displacements) are shown. Additionally, all the real-world experiments are visualized in the provided supplementary video.

1) At Optimal Conditions: These experiments are conducted to validate the proposed method using optimal scan rates that balance the image acquisition speed and the level of noise. Scan speed used is 720 ns/pixel (oversampling rate[4] of 8) which provides a frame rate of 2.2 frames per second. Magnification is fixed to $\times 1200$ in order to have one complete micropart in the field of view. The images shown in Fig. 9(a)-(d) illustrate the positioning task. Fig. 9(a) and Fig. 9(b) show the desired and initial positions of micropart, respectively. Fig. 9(c) and Fig. 9(d) depict, respectively the error ($\mathbf{I} - \mathbf{I}^*$) at initial and final positions during the task. Platform voltages (for x and y channels) and the positioning error variations are shown in the corresponding plots of condition-1. Obtained results show that the positioning task has been successfully accomplished. However, because of the effect of variable gain and the low frame rate (compared to the optical systems), the system converges faster when the distance is more.

2) At noisy conditions: Second tests are performed to position the microparts using the images acquired with an increased raster scanning speed. Normally in SEM imaging, high scanning rates during image acquisition lead to the increased noise levels in images [12]. The main goal of this test is to check the method's efficiency in reaching the desired position under noisy conditions. For this test a scan speed of 180 ns/pixel (maximum allowed) that provides a frame rate of 3.1 frames per second has been used. Fig. 9(e)-(h) illustrates the positioning process during this test. The voltage and the cost variations are shown in the corresponding plots of condition-2. Despite having high amount of noise, the desired position has been reached successfully (Fig. 9(d)).

3) At high magnification: These experiments are performed to position the microparts at high magnification. Simultaneously, the method is also validated with increased scan speed at the selected magnification. The magnification and the scan speed used for this test are $\times 5000$ and 360 ns/pixel, respectively. The sequence of images shown in the Fig. 9(i)-(l) demonstrates the positioning task. Final error during the process has been shown in Fig. 9(d). The voltage and cost variations are shown in the corresponding plots of condition-3. As can be seen from the obtained results, the control law converges to the desired position.

4) At Unstable Conditions: These sets of experiments are conducted at unstable conditions such as extreme variations in the image intensities. With SEM, these variations are produced due to the induced charging and carbon contamination of the sample surface. Besides, the imaging contrast and brightness are never constant. The main objective of this test is to validate the proposed method's robustness to the occurred disturbances during the positioning process. For this test a scan speed of

[4]Oversampling rate is a parameter modified using DISS5 to change the scan speed. It adjusts the amount of analog to digital conversions per pixel. Thus, the final image signal-to-noise ratio can be modified.

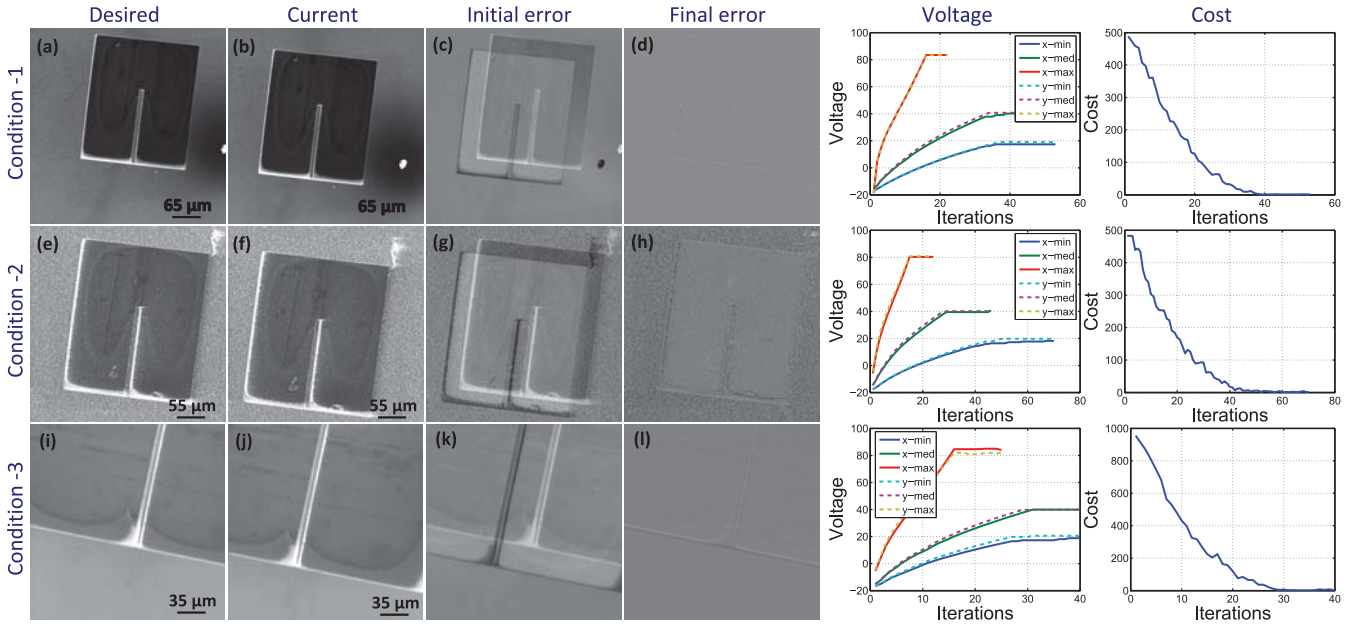


Fig. 9. Nanopositioning at three different conditions: 1) Optimal 2) noisy 3) increased magnification. Column wise: 1) desired image 2) initial image in the positioning task 3) initial error and 4) final error. Voltage and cost variations throughout the process for each condition has been shown.

360 ns/pixel (oversampling rate of 8) that provides a frame rate of 2.2 frames per second has been used. The magnification is fixed to $\times 1200$. In this case, the desired image (Fig. 10(a)) has been obtained with optimum brightness and contrast. After having moved the platform to its initial position, the contrast and the brightness values are changed manually throughout the process. Fig. 10(b) shows initial image during this task, where the contrast and brightness are increased. The channel voltage and cost-function variations are shown in the corresponding plots of condition-4. From the obtained results, it can be seen that the control law converged to the desired position despite having modified operating conditions. It is mainly due to the robustness of Fourier-based method to image noise and intensity variations.

5) Using High Quality Desired Image: These sets of experiments are conducted to check the method's efficiency in positioning when a high quality image is used as a desired image. The desired image shown in Fig. 10(e) has been obtained with a scan speed of 5.8 μs/pixel. Before initiating the control, the scan speed has been increased to 360 ns/pixel. From Fig. 10(f), it can be seen that the noise level increased predominantly in the initial image when compared to the desired image. Fig. 10(g) and Fig. 10(h) depict, respectively the initial and the final error during the positioning task. The channel voltage and cost variations are shown in in the corresponding plots of condition-5. From the obtained results, it is clear that the positioning task has been successfully accomplished. Table II summarizes the final positioning errors with different quality images. Interestingly, it has been found that the positioning error reduced with an increase in the quality of the desired image.

6) Using a Mask as a Desired Image: These sets of experiments are performed using a binary mask of the object as a desired image. The objective is to demonstrate the positioning task using minimal information. Moreover, it also benefits in automatic positioning of an object in between the gripper fingers during a handling task without using a known desired image. In other words, it removes teaching step defining desired image, which is one of the major problems of visual servoing. For this test, the binary mask (Fig. 10(i)) of the micropart has been obtained by intensity thresholding. For demonstration, the four corner points of the mask region are supplied using mouse clicks. Fig. 10(i)-(l) shows the positioning task. In spite of having more than 60% discrepant pixels in the desired image, the goal position has been reached successfully. The only limitation that has been observed is that the poor performance of method in case of having repeated patterns similar to the shape of mask.

TABLE II
POSITIONING ERRORS WITH DIFFERENT QUALITY IMAGES.

Oversamp. rate	scan speed ($\mu s/pixel$)	error-x (pixels)	error-y (pixels)
24	4.3	0.65	0.81
32	5.8	0.48	0.69
48	8.6	0.33	0.51
64	11.5	0.26	0.39

7) Using Sample with Repeated Pattern: The final sets of experiments are conducted to demonstrate the method's performance when operating in a scene containing repeated patterns, heavy textures and high level of noise. With SEM, image noise increases with increase in scan speed. A scan speed of 180 ns/pixel (maximum allowed with an oversampling rate of 2) that provides a frame rate of 3.1 frames per second has been used. Calibration sample is used for this test. Fig. 11(a) and Fig. 11(b) show the initial image and final error, respectively and Fig. 11(c) and (d) illustrates the variable variations during

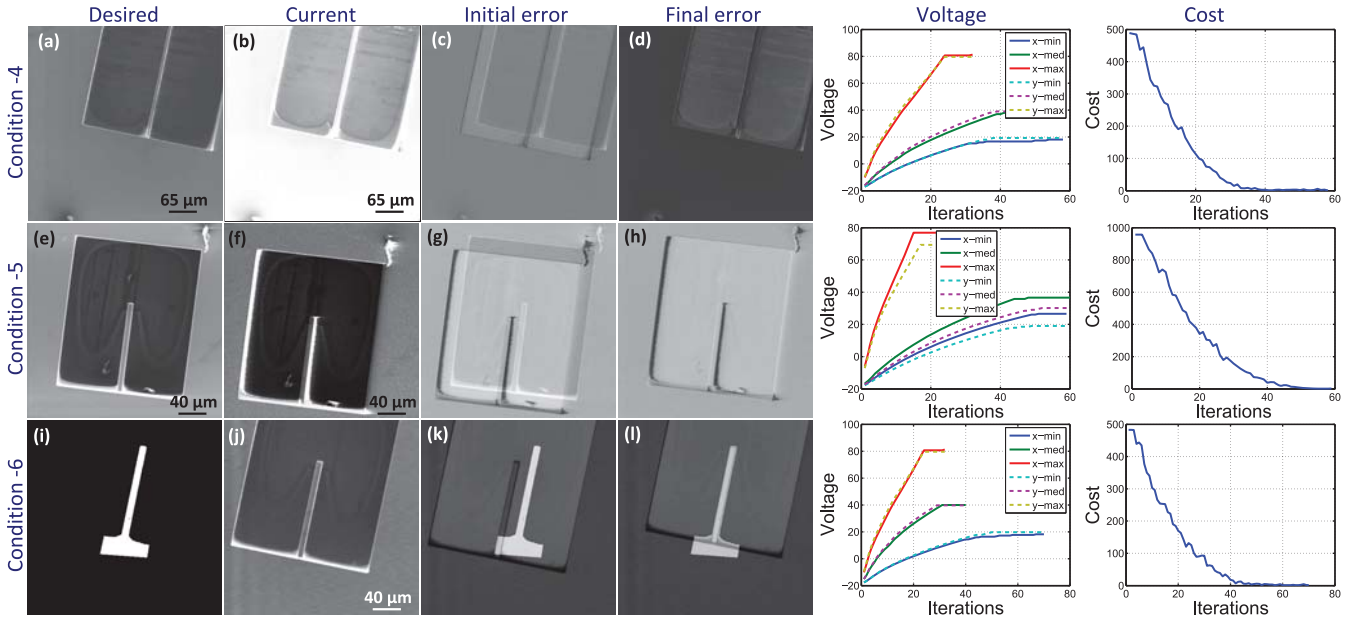


Fig. 10. Nanopositioning at three different conditions: 1) Unstable imaging conditions, 2) using high quality desired image and 3) artificial mask as desired image. Column wise: 1) desired image 2) initial image in the positioning task 3) initial error and 4) final error. Voltage and cost variations throughout the process for each condition has been shown.

the positioning process. From the obtained results, it can be seen that despite having repeated patterns and high amount of noise, the control law successfully converges to the desired position. It clearly points out the proposed method's robustness towards the object and scene being viewed.

D. Positioning Accuracy

Since the used SEM does not have enough room (inside vacuum chamber) for the external displacement measuring devices like laser interferometers, the positioning accuracy has been computed using the final positioning error obtained from the image measurements. Even though, the estimated accuracy is uncertain (due to the presence of image noise), it has been counted to roughly demonstrate the method's efficiency. With real measuring devices, it might be more satisfactory as the level of measuring noise is less in compared to the image measurements. In the current case, it is computed from the product of error pixels and the pixel dimension measured on sample. The pixel dimension P on the sample is computed using (43).

$$P = \frac{D}{G} \quad [\mu m] \tag{43}$$

where, D is the pixel dimension on the screen (constant) and G is the magnification. With our system, the computed D value is 212.3 for a screen size of 512×512 pixels. The accuracies computed with different tests are summarized in Table III along with their standard deviations over all 10 trails. For simplicity, we denote the tests with case numbers: 1) nominal conditions, 2) noisy conditions, 3) high magnifications, 4) unstable conditions, 5) high quality desired image, 6) binary mask as desired image. Besides, to evaluate the method's performance, two other parameters *i.e.*, root mean squared error (RMSE) and peak signal-to-noise ratio (PSNR)

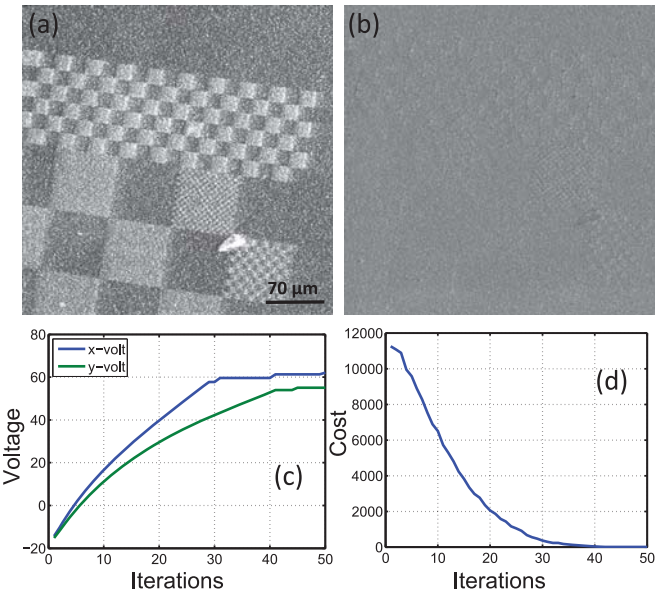


Fig. 11. Nanopositioning of calibration sample. (a) Initial image (b) final error (c) voltage and (d) cost variations.

are estimated. Both these parameters are computed using the initially obtained desired image and the image obtained after convergence *i.e.,* the final image in the process. Since, the latter three tests does not use a normal desired image, these parameters are not estimated for them.

E. Performance Analysis

Due to the amount of theoretical calculations involved, it is worth evaluating the method's performance in real-time applications and also in comparison with the well-known

TABLE III
ANALYSIS OF METHOD'S POSITIONING ACCURACY.

Case	Accuracy (μm)		Std. dev. (μm)		RMSE	PSNR
	x	y	x	y		
1	0.11	0.19	0.0248	0.0187	1.60	38.73
2	0.14	0.23	0.0108	0.0216	1.87	32.70
3	0.04	0.11	0.0061	0.0053	1.79	33.68
4	0.22	0.37	0.0147	0.0286	–	–
5	0.09	0.13	0.0089	0.0127	–	–
6	0.29	0.47	0.0497	0.0589	–	–

TABLE IV
EVALUATION OF PROCESSING TIMES ($milliseconds$) FOR VARIOUS METHODS.

Case	Ours	Flow-based	SURF	FAST
1	38	29	911	73
2	33	34	797(4)	49(2)
3	37	30	1063	65(5)
4	38	×	×	×
5	36	34(5)	826(3)	71(4)
6	39	33	×	×

[1] × indicate that the task failed all 6 times.
[2] Value in (.) indicate the number of times the positioning task succeeded.

spatial processing methods like optical flow-based gradient approach [20], SURF and FAST. In this section, we start by analysing the computational cost involved with our method and then show the speed comparisons. Summarizing our approach, the major steps include: correlating two image frequency spectra and computing the correlation peak. Suppose, the shift between two images is P in each direction and correlation is performed on two images of $\mathrm{N} \times \mathrm{N}$. This results in a total of $\mathrm{P}^2 + \mathrm{P}$ shift probables, out of whom the final correlation output has to be computed. This can be accomplished efficiently using fast Fourier transforms[5] (FFT) instead of the regular discrete Fourier transform (DFT) [35]. Main reason is, to compute a simple 1D DFT, it requires $\mathcal{O}(\mathrm{N}^2)$ computations, and for 2D DFT it goes up to $\mathcal{O}(2\mathrm{N}^3)$. Hence, it is computationally complex when dealing with large images. This can be reduced using FFT to just $\mathcal{O}(\mathrm{N}^2 \log_2 \mathrm{N})$ computations. In case of rotations, we used a lookup table-based approach for estimating the kernel image depth to speed-up the estimation process. Besides, relying on the fact of motion decoupling, rotations and translated are processed in parallel. The average total time taken to process one iteration is measured to be 38 ms.

Various tests are performed to verify the processing performance at previously mentioned 6 experimental conditions. Table IV shows the overall time taken to process initial iteration in the control process (on average of 6 runs). The value in brackets suggest the number of times the task has been succeeded out of the total 6 trials. It can be seen that local feature tracking methods (SURF and FAST) has performed poorly. Moreover, their processing speed is also affected by the number of keypoints detected and tracked. Even though, the optical flow-based method shown comparatively better performance, it failed at times. In contrast to the spatial processing methods, the presented method has exhibited reliable performance. Some facts about the presented method can be considered here: no need to measure or track local features, phase differences are treated on equal basis, and invariant to various disturbances.

[5]FFT algorithms compute N-D DFT in successive lower order DFTS.

V. CONCLUSION AND PERSPECTIVES

The problem of automatic nanopositioning inside a SEM using visual servoing was studied in this paper. Considering the problem of using visual feature tracking for SEM, we presented a Fourier-based automatic nanopositioning scheme, which operates by estimating the relative motion between two images using their frequency spectral information. The control strategy was designed to compensate this motion in image space by controlling the positioning platform movements. Rotational control was validated with simulations and translational control was validated using an experimental set-up containing the tungsten gun SEM equipped with a piezo-positioning platform. From the obtained results, it appears that the method is robust to different disturbances that occur in a SEM-based environment. The future work will attempt to control the positioning of a 3 degrees of freedom microactuator (Kleindiek MM3A) containing a microgripper to manipulate the micron sized silicon walls for surface characterization.

REFERENCES

[1] X. Ye, Y. Zhang, C. Ru, J. Luo, S. Xie, and Y. Sun, "Automated pick-place of silicon nanowires," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 554–561, 2013.

[2] V. Eichhorn, M. Bartenwerfer, and S. Fatikow, "Nanorobotic assembly and focused ion beam processing of nanotube-enhanced afm probes," *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 4, pp. 679–686, 2012.

[3] C. Ru, Y. Zhang, Y. Sun, Y. Zhong, X. Sun, D. Hoyle, and I. Cotton, "Automated four-point probe measurement of nanowires inside a scanning electron microscope," *IEEE Trans. Nanotechnol.*, vol. 10, no. 4, pp. 674–681, 2011.

[4] K. Aoki, H. T. Miyazaki, H. Hirayama, K. Inoshita, T. Baba, K. Sakoda, N. Shinya, and Y. Aoyagi, "Microassembly of semiconductor three-dimensional photonic crystals," *Nat. Mater.*, vol. 2, no. 2, pp. 117–121, 2003.

[5] Y. L. Zhang, Y. Zhang, C. Ru, B. K. Chen, and Y. Sun, "A load-lock-compatible nanomanipulation system for scanning electron microscope," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 1, pp. 230–237, 2013.

[6] B. Tamadazte, M. Paindavoine, J. Agnus, V. Pétrini, and N. Le-Fort Piat, "Four dof piezoelectric microgripper equipped with a smart cmos camera," *J. Microelectromech. Syst.*, vol. 21, no. 2, pp. 256–258, 2012.

[7] G.-Y. Gu, L.-M. Zhu, C.-Y. Su, H. Ding, and S. Fatikow, "Modeling and control of piezo-actuated nanopositioning stages: A survey," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 1, pp. 313–332, Jan 2016.

[8] R. Oubellil, A. Voda, M. Boudaoud, and S. Regnier, "Robust control strategies of stick-slip type actuators for fast and accurate nanopositioning operations in scanning mode," in *Proc. Mediterranean Conf. Control. Autom.*, June 2015, pp. 650–655.

[9] S. Xiao and Y. Li, "Visual servo feedback control of a novel large working range micro manipulation system for microassembly," *J. Microelectromech. Syst.*, vol. 23, no. 1, pp. 181–190, 2014.

[10] H. Tang, Y. Li, and Q. Yang, "Tracking control of a compliant xy nanopositioner under plant uncertainty using a transfigured loop-shaping $H\infty$ controller," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, Aug 2012, pp. 103–108.

[11] S. Zimmermann, T. Tiemerding, and S. Fatikow, "Automated robotic manipulation of individual colloidal particles using vision-based control," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 5, pp. 2031–2038, 2015.

[12] N. Marturi, S. Dembélé, and N. Piat, "Scanning electron microscope image signal-to-noise ratio monitoring for micro-nanomanipulation," *Scanning*, vol. 36, no. 4, pp. 419–429, 2014.

[13] N. Marturi, S. Dembélé, and N. Piat, "Fast image drift compensation in scanning electron microscope using image registration," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, 2013, pp. 807–812.

[14] A. C. Malti, S. Dembélé, N. Piat, C. Arnoult, and N. Marturi, "Toward fast calibration of global drift in scanning electron microscopes with respect to time and magnification," *Int. J. Optomechatronics*, vol. 6, no. 1, pp. 1–16, 2012.

[15] S. Zimmermann, T. Tiemerding, T. Li, W. Wang, Y. Wang, and S. Fatikow, "Automated mechanical characterization of 2-d materials using sem based visual servoing," *Int. J. Optomechatronics*, vol. 7, no. 4, pp. 283–295, 2013.

[16] C. Diederichs, M. Bartenwerfer, M. Mikczinski, S. Zimmermann, T. Tiemerding, C. Geldmann, H. Nguyen, C. Dahmen, and S. Fatikow, "A rapid automation framework for applications on the micro-and nanoscale," in *Proc. Australasian Conf. on Rob. and Auto.*, 2013.

[17] D. Jasper, "High-speed position tracking for nanohandling inside scanning electron microscopes," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2009, pp. 508–513.

[18] C. Dahmen and T. Tiemerding, "Fast and robust position determination in the scanning electron microscope," in *Proc. Australasian Conf. on Rob. and Auto.*, 2013.

[19] B. E. Kratochvil, L. Dong, and B. 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.

[20] B. Tamadazte, N. Le-Fort Piat, and E. Marchand, "A direct visual servoing scheme for automatic nanopositioning," *IEEE/ASME Trans. Mechatronics*, vol. 17, no. 4, pp. 728–736, 2012.

[21] Q. Bateux and E. Marchand, "Direct visual servoing based on multiple intensity histograms," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 6019–6024.

[22] V. Kallem, M. Dewan, J. P. Swensen, G. D. Hager, and N. J. Cowan, "Kernel-based visual servoing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 1975–1980.

[23] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–200, 2002.

[24] P. Cizmar, A. E. Vladár, and M. T. Postek, "Real-time scanning charged-particle microscope image composition with correction of drift," *Microscopy and Microanalysis*, vol. 17, no. 02, pp. 302–308, 2011.

[25] N. Marturi, B. Tamadazte, S. Dembélé, and N. Piat, "Visual servoing schemes for automatic nanopositioning in scanning electron microscope," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2014, pp. 981–986.

[26] B. S. Reddy and B. N. Chatterji, "An fft-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, 1996.

[27] P. Brou, "Using the gaussian image to find the orientation of objects," *Int. J. Robot. Res.*, vol. 3, no. 4, pp. 89–125, 1984.

[28] N. Marturi, B. Tamadazte, S. Dembélé, and N. 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.

[29] A. Aldoma, Z.-C. Marton, F. Tombari, W. Wohlkinger, C. Potthast, B. Zeisl, R. B. Rusu, S. Gedikli, and M. Vincze, "Point cloud library," *IEEE Robot. Autom. Mag.*, vol. 1070, no. 9932/12, 2012.

[30] P. J. Kostelec and D. N. Rockmore, "FFTs on the rotation group," *J Fourier. Anal. Appl.*, vol. 14, no. 2, pp. 145–179, 2008.

[31] P. J. Kostelec and D. N. Rockmore, "SOFT: SO(3) Fourier Transforms," *Department of Mathematics, Dartmouth College, Hanover, NH*, vol. 3755, 2007.

[32] N. Marturi, B. Tamadazte, S. Dembélé, and N. Piat, "Visual servoing-based depth estimation technique for manipulation inside SEM," *IEEE Trans. Instrum. Meas.*, in press.

[33] F. Chaumette, "Visual servoing," in *Robot Manipulators: Modeling, Performance Analysis and Control*, E. Dombre and W. Khalil, Eds. ISTE, 2007, ch. 6, pp. 279–336.

[34] E. Marchand, F. Spindler, and F. Chaumette, "ViSP for visual servoing: a generic software platform with a wide class of robot control skills," *IEEE Robot. Autom. Mag.*, vol. 12, no. 4, pp. 40–52, Dec 2005.

[35] M. Frigo and S. G. Johnson, "The design and implementation of fftw3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005.

[36] P. Cizmar, A. Vladar, B. Ming, and M. Postek, "Simulated sem images for resolution measurement," *Scanning*, vol. 30, no. 5, pp. 381–391, 2008.

Naresh Marturi received the M.S. degree in robotics and intelligent systems from the 'Orebro University, Sweden, and the Ph.D. degree in automation and computer Science from the "Université de Franche-Comté', France, in 2010 and 2013, respectively. Since 2015, he has been a researcher with KUKA Robotics UK Ltd., Wednesbury, UK and University of Birmingham, UK. His main research interests include: vision-based control of industrial robots and development of 2D and 3D computer vision techniques.

Brahim Tamadazte holds a Ph.D. in Automation and Computer Science from the "Université de Franche-Comté" in 2009 and a MS degree in Robotics and Intelligent Systems from the "Université Pierre et Marie Curie (Paris VI)" in 2005, in France. Since 2012, he is with the FEMTO-ST Institute in the AS2M Department as Senior Scientist CNRS researcher (Besançon, France). Currently, he works in surgical microrobotics surgery (conception and control) and visual servoing using OCT imaging systems.

Sounkalo Dembélé received his Ph.D. in Systems Control at University of Franche-Comté Besançon in 1993. He became Electronics Engineer at ENSMM (National Engineering School of Mechanics and Microtechniques) Besançon in 1994 where he developed a mobile robot, Associate Professor at University of Franche-Comté Besançon in 1995. His is a member of FEMTO-ST Institute where he works on robotic assembly of MEMS and MOEMS. After swarm robotics, control of an ultrasonic linear micromotor and fringe based tracking, his researches currently involve Novel View Synthesis by mosaicing and trifocal transfer, shape-from-focus based tracking, feature based tracking, calibration of microscope, vision based control.

Nadine Piat received her Ph.D. in Control Systems and Signal Processing at the University of Technology in Compiégne (UTC) in 1984. She became Associate-Professor at UTC Compiégne in 1984 and Professor at the ENSMM (National Engineering School of Mechanics and Microtechniques - Besançon) in 1998. She is a member of the AS2M department of the Institute FEMTO-ST, UMR CNRS 6174, in Besançon. Her research areas concern perception and advanced control strategies based on visual servoing and reinforcement learning for automation of distributed microrobotics.