Three-Dimensional Visual Tracking and Pose Estimation in Scanning Electron Microscopes

Le Cui, Eric Marchand, Sinan Haliyo, Stéphane Régnier

Three-Dimensional Visual Tracking and Pose Estimation in Scanning Electron Microscopes

Le Cui[1], Eric Marchand[1], Sinan Haliyo[2] and Stéphane Régnier[2]

Abstract— **Visual tracking and estimation of the 3D posture of a micro/nano-object is a key issue in the development of automated manipulation tasks using the visual feedback. The 3D posture of the micro-object is estimated based on a template matching algorithm. Nevertheless, a key challenge for visual tracking in a scanning electron microscope (SEM) is the difficulty to observe the motion along the depth direction. In this paper, we propose a template-based hybrid visual tracking scheme that uses luminance information to estimate the object displacement on x-y plane and uses defocus information to estimate object depth. This approach is experimentally validated on 4-DoF motion of a sample in a SEM.**

I. Introduction

Over the last decade, visual tracking is investigated for automated (or semi-automated) micro/nano-manipulation tasks within Scanning Electron Microscopes (SEM). Nevertheless, only a few tracking algorithms have been successfully implemented inside a SEM currently. An active-contours-based and correlation-based pattern matching method for nanohandling in a SEM has been proposed [1], in which the pose on 3 DoFs (translation along x- and y-axes, rotation around z-axis) are estimated. This method has been improved and applied to a microrobot system inside a SEM [2], [3] for semi-automatic nanohandling. In [4], instead of acquiring the whole image, dedicated few line scans are used to detect the motion of a nano-object or a reference pattern. This approach can be applied to a closed-loop positioning task. Advantages of these template-matching-based methods are their simple implementation and their robustness to additive noise of the SEM image. However, these methods highly depend on the templates and could be sensitive to clutter environment. Alternatively, the model-based tracking method has been proposed and implemented for precise automated manipulation and quantified in a SEM [5]. The 3D model-based approaches perform well on estimating the posture of the object, although they show less robustness to additive noise and highly depend on the 3D model and the feature extraction. Recently, [6] have proposed a visual tracking framework using CAD model for MEMS micro-assembly. Although this approach was initially implemented for optical microscope, experiments have also been conducted within a SEM. Additionally, an improved template matching-based contour model was proposed for the tracking task in a SEM [7] and was applied to vision-guided nanomanipulation of nanowires using four nanoprobes [8]. In this method, a gradient based subpixel method has been introduced for a nanoprobe contour tracking task to improve the accuracy. Moreover, nanoprobe tips tracking for translational motions has been implemented [9].

It remains that two major aspects of image formation within a SEM have to be considered for proper visual tracking. The former is that, from a geometrical aspect, a SEM obey to a parallel projection model [10], [11]. A consequence is that a motion along the depth direction is not observable and, thus, depth can not be estimated from the observation of geometric features. An alternative to geometric observation has then to be considered. Dealing with the later, it should be noticed that most of the current visual tracking methods ignore the particularity that the SEM image sharpness varies when the sample moves along the depth direction, especially at high magnifications. The acquired images are blurred due to the defocus. This may lead to inaccuracy on the feature extraction or the template matching process. Indeed, when the images are significantly blurred, the detection of points or lines for model-based tracking approach is highly affected and the visual tracking task fails. When considering template-based methods only planar translational motions can be considered since motion along depth axes highly impacts the reference template.

In order to estimate a more 3D position of the object with the highest accuracy, the defocus can be considered as a source to recover the depth information. In the literature, Dahmen [12], [13] has proposed to record the normalized variance of the image intensity at various positions off-line and then to recover the position using a lookup table-based method on-line. Nevertheless, this method highly depends on the data set and the estimation is affected by the random image noise hence lacks robustness.

This study addresses a visual tracking and posture estimation framework that consider both defocus and luminance informations for 3D motion of the sample in a SEM. The image sharpness information is integrated into a template-based matching process. The manuscript is organized as follows: Section II introduces the proposed visual tracking method. Section III describes the posture estimation approach from visual tracking. Experimental results obtained on a parallel microrobotic workcell are shown in Section IV.

II. Visual tracking framework

In a SEM, it is observed that the image sharpness varies with the sample motion along the depth direction. In this article, we consider both luminance and defocus blur in the observed image into the visual tracking framework. The

[1]Le Cui and Eric Marchand are with Université de Rennes 1, Lagadic group, IRISA, Inria Rennes, Rennes, France marchand@irisa.fr

[2]Sinan Haliyo and Stéphane Régnier are with Sorbonne Universités, UPMC Univ Paris 06, UMR 7222, ISIR, Paris, France

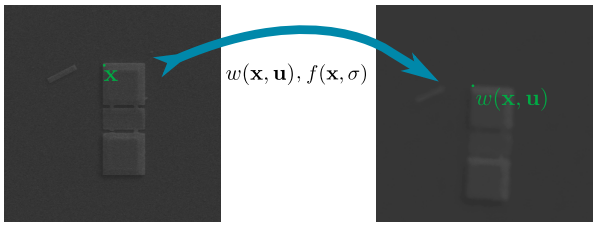

Fig. 1. Visual tracking based on minimizing the dissimilarities of both displacements and blur level

considered method is a template-based tracking approach where the appearance of the image is employed.

A. Template registration for visual tracking

Considering that the appearance of the object is learned from a reference template I^* with pixel position $\mathbf{x} \in W$, the idea of template registration [14] is to look for a new location of these pixels $w(\mathbf{x},\mathbf{u})$ in the current image I (where $\mathbf{u}$ is the displacement parameters) by minimizing the dissimilarity between the reference image and the current image. Considering the sum of squared differences (SSD) as this dissimilarity function:

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmin}} \sum_{\mathbf{x}\in W} (I(w(\mathbf{x},\mathbf{u})) - I^*(\mathbf{x}))^2 \tag{1}$$

Using the Gauss-Newton optimization method to solve this non-linear problem, for each pixel, the first order Taylor expansion of the error $C(\mathbf{u}) = I(w(\mathbf{x},\mathbf{u})) - I^*(\mathbf{x})$ is given by:

$$C(\mathbf{x},\mathbf{u}+\delta\mathbf{u}) \approx I(w(\mathbf{x},\mathbf{u})) + \nabla I \frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}} \delta\mathbf{u} - I^*(\mathbf{x}) \tag{2}$$

where $\delta\mathbf{u}$ is the increment of the displacement parameters, $\Delta I = (\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})^\top$ is the gradient of the image evaluated at $w(\mathbf{x},\mathbf{u})$ and $\frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}}$ is the Jacobian of the warp. Injecting equation (2) into (1):

$$C(\mathbf{x},\mathbf{u}+\delta\mathbf{u}) = \sum_{\mathbf{x}\in W} (I(w(\mathbf{x},\mathbf{u})) + \nabla I \frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}} \delta\mathbf{u} - I^*(\mathbf{x}))^2 \tag{3}$$

The partial derivative of equation (3) with respect to $\delta\mathbf{u}$ is:

$$\begin{aligned} \frac{\partial C(\mathbf{x},\mathbf{u}+\delta\mathbf{u})}{\partial \delta\mathbf{u}} =& 2 \sum_{\mathbf{x}\in W} (\nabla I \frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}})^\top (I(w(\mathbf{x},\mathbf{u})) \\ &+ \nabla I \frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}} \delta\mathbf{u} - I^*(\mathbf{x})). \end{aligned} \tag{4}$$

It is evident that when the cost function C reaches its minimum, equation (4) equals zero. In this case, the increment of the displacement can be then estimated using:

$$\delta\mathbf{u} = H^{-1} \sum_{\mathbf{x}\in W} (\nabla I \frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}})^\top (I^*(\mathbf{x}) - I(w(\mathbf{x},\mathbf{u}))), \tag{5}$$

where H is the Gauss-Newton approximation of the Hessian matrix:

$$H = \sum_{\mathbf{x}\in W} (\nabla I \frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}})^\top (\nabla I \frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}}). \tag{6}$$

The displacement parameters $\mathbf{u}$ can be then updated by $\delta\mathbf{u}$ in each iteration during the non-linear optimization process until the convergence.

To express the displacement of an object in the given image with respect to a reference template, the warp functions $w(\cdot)$ has to be defined. In our case the 4 DoFs motion of the sample is considered. As stated above, a motion along the depth axis will lead to important change in the image due to defocus. Therefore, this template tracker only considers the translations along x, y axes and the rotations around z axis, $\mathbf{u} = (\theta, t_x, t_y)$ between two pixel locations. This can be modeled as:

$$\mathbf{x}_2 = \mathbf{R}\mathbf{x}_1 + \mathbf{t} \tag{7}$$

where $\mathbf{t} = (t_x, t_y)^\top$ is a translation vector and $\mathbf{R}$ is a 2D rotation matrix:

$$\mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The Jacobian of warp $\frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}}$ is given by:

$$\frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}} = \begin{pmatrix} -x\sin\theta - y\cos\theta & 1 & 0 \\ x\cos\theta - y\sin\theta & 0 & 1 \end{pmatrix}. \tag{8}$$

B. Visual tracking using luminance and defocus information

In order to perform visual tracking of three-dimensional motions of a micro-scale object in a SEM, we propose to consider the variation of the sharpness of the image caused by the motion of the sample along the depth direction. Assuming that the reference template is in-focus in the visual tracking task, the general idea is to determine the defocus level σ and the displacement parameters $\mathbf{u} = (\theta, t_x, t_y)$ by minimizing the dissimilarity on both image appearance and image sharpness between the observed image I and the artificially blurred reference image I_b^* using a non-linear optimization process. This problem can be written as:

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmin}} \sum_{\mathbf{x}\in W} (I(w(\mathbf{x},\mathbf{u})) - I_b^*(\mathbf{x},\sigma))^2 \tag{9}$$

and

$$\hat{\sigma} = \underset{\sigma}{\operatorname{argmin}} \sum_{\mathbf{x}\in W} (G(w(\mathbf{x},\mathbf{u})) - G_b^*(\mathbf{x},\sigma))^2 \tag{10}$$

where G is the image gradient of image I defined by:

$$\begin{aligned} G &= \sum_{x=0}^{M}\sum_{y=0}^{N} \|\nabla I(x,y)\|^2 \\ &= \sum_{x=0}^{M}\sum_{y=0}^{N} (\nabla I_x^2(x,y) + \nabla I_y^2(x,y)), \end{aligned} \tag{11}$$

and G_b^* is the image gradient of the blurred reference template.

Recalling the SEM image blur model [15], a blurred image $I_b(x,y)$ can be expressed as the convolution of a sharp image $I_s(x,y)$ and the Gaussian kernel:

$$I_b(x,y) = I_s(x,y) * f(x,y) \tag{12}$$

where the Gaussian kernel $f(x,y)$ can be expressed by:

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \quad (13)$$

where σ is the standard deviation of the Gaussian kernel. Since we assume that the reference template is in-focus, in our visual tracking scheme the reference template is blurred artificially using equation (12). In this case, σ is considered as the blur level to be optimized.

The Jacobian $J_\sigma = \frac{\partial G}{\partial \sigma}$ linking σ and the gradient G is obtained by:

$$\frac{\partial G}{\partial \sigma} = \sum_{x=0}^{M}\sum_{y=0}^{N} 2(\nabla I_x(x,y)\frac{\partial \nabla I_x(x,y)}{\partial \sigma} + \nabla I_y(x,y)\frac{\partial \nabla I_y(x,y)}{\partial \sigma}). \quad (14)$$

With the Gauss-Newton optimization method, the minimization problem is solved by updating $\mathbf{u}$ and σ alternatively in each iteration:

$$\partial \mathbf{u} = -\mathbf{J_u}^{+}(I(w(\mathbf{x},\mathbf{u})) - I_b^*(\mathbf{x},\sigma)) \quad (15)$$

and

$$\partial \sigma = -J_\sigma^{-1}(G(w(\mathbf{x},\mathbf{u})) - G_b^*(\mathbf{x},\sigma)) \quad (16)$$

where the Jacobian $\mathbf{J_u}$ is defined as $\mathbf{J_u} = (\ldots, \nabla I \frac{\partial w(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}}, \ldots)^\top$. In each iteration of the minimization process, the reference image I^* is artificially blurred (using updated σ computed by equation (16)) to be compared with the current image I; the displacement $\mathbf{u}$ between these two images is then updated using equation (15).

III. Posture estimation

In the visual tracking process, the parameters in the warp function and the blur level σ are estimated. With these parameters, the posture of the object in the camera coordinate frame or in the world coordinate frame can be then recovered. It should be noted that, in SEM vision, the motion along the depth direction can not be observed by measuring the sample scale since the parallel projection model is applied [10]. In this case, the estimation of the position along the depth direction and the partial posture on other axes should be performed separately.

A. Partial posture estimation by 3D registration

Considering a 3D point ${}^w\mathbf{X} = ({}^wX, {}^wY, {}^wZ, 1)^\top$ in an object reference frame, its projection on the image plane (expressed in pixels) $\mathbf{x}_p = (u, v, 1)^\top$ can be modeled by

$$\mathbf{x}_p = \mathbf{K}\mathbf{\Pi}{}^c\mathbf{T}_w{}^w\mathbf{X} \quad (17)$$

where $\mathbf{K} = \begin{pmatrix} p_x & 0 & 0 \\ 0 & p_y & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\mathbf{\Pi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and ${}^c\mathbf{T}_w = \begin{pmatrix} {}^c\mathbf{R}_w & {}^c\mathbf{t}_w \\ \mathbf{0}_{3\times 1} & 1 \end{pmatrix}$ is a homogeneous matrix that describes the relation between the object frame and the camera frame. In our model,

$${}^c\mathbf{R}_w = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

describes the rotation around the z-axis. In general, the pixel/meter ratio p_x, p_y can be easily obtained from the SEM software, from calibration procedure [10] or simply computed from a known object measurement in meters vs pixels.

Since the pixel position of a point ${}^i\mathbf{x}$ on the image can be estimated from the tracking task, we are able to obtain its 3D posture $\mathbf{r}$ of the object by minimizing the registration error between the re-projected pixel position ${}^i\mathbf{x}_p(\mathbf{r})$ and the tracked pixel position ${}^i\mathbf{x}_p^*$ using a non-linear optimization. The problem can be written as:

$$\hat{\mathbf{r}} = \underset{\mathbf{r}}{\operatorname{argmin}} \sum_{i=1}^{N} ({}^i\mathbf{x}_p(\mathbf{r}) - {}^i\mathbf{x}_p^*)^2 \quad (18)$$

where N is the number of points used to estimate the posture.

The update in each iteration using the Gauss-Newton optimization method is:

$$\delta\mathbf{r} = -\lambda \mathbf{J}^{+}(\mathbf{x}_p(\mathbf{r}) - \mathbf{x}_p^*) \quad (19)$$

where $\mathbf{J}$ is a Jacobian linking the variation of the posture $\mathbf{r}$ and the pixel location $\mathbf{x}_p$ on image.

It should be noted that the parallel projection model is considered in a SEM. The depth motion is hence unobservable from the variation of the pixel position in the image, or the scale of the sample that is projected on the image plane. In this case, the depth information can no longer be recovered from the 3D registration and only 3 DoFs are considered in the Jacobian. In this case, the Jacobian is given by:

$$\mathbf{J} = \begin{pmatrix} -1 & 0 & y \\ 0 & -1 & -x \end{pmatrix}. \quad (20)$$

B. Estimating depth position using particle filter

In general, the depth position can be computed by the estimated blur level σ using a lookup table-based method (similar to [13]). However, the results would be less reliable than the estimation on the position along other axis using the approach presented in the previous paragraph due to inaccurate image sharpness estimations. Those play an important role in the estimations of the depth position. This inaccuracy is due to system noise, which describes the inaccuracy of the supposed system dynamics model and observation model, and by image noise (caused in the SEM image formation process).

Alternative techniques should be employed to achieve a robust and accurate depth estimating process. This problem involves the estimation of the position along the depth direction, by measuring the observations (in our case image sharpness). The particle filter [16] is considered to solve this estimation problem. Particle filters are Bayesian-based methods for performing inference in state-space models for

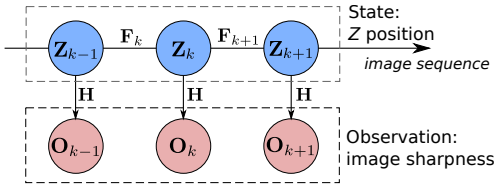

Fig. 2. Estimation of the position along the depth direction from image sharpness

a dynamic system via noisy observations. They comprise a broad family of sequential Monte Carlo algorithms that approximates inference in partially observable Markov chains. The general idea of particle filter techniques is to represent the required posterior density function by a set of random samples (particles) with associated weights and to estimate the internal state of the dynamic system based on these samples and weights [16].

A particle filter is based on a system dynamics model that describes the time-dependent evolution of the state is proposed as:

$$\mathbf{Z}_k = \mathbf{F}(\mathbf{Z}_{k-1}, \boldsymbol{\nu}_{k-1}) \tag{21}$$

where $\mathbf{Z}_k$ is the state vector at kth frame in the tracking, $\mathbf{F}$ is a possibly nonlinear function of the state $\mathbf{Z}_{k-1}$. $\boldsymbol{\nu}$ is an independent and identically distributed (i.i.d.) system noise sequence. equation (21) represents the evolution of a state vector $\mathbf{Z}$ from frame $k-1$ to frame k. In our tracking framework, we denote the state vector by $\mathbf{Z}_k = (Z_k, \dot{Z}_k)^\top$. Using a constant velocity evolution model (depth velocity is supposed to be constant), equation (21) can be rewritten as:

$$\begin{pmatrix} Z_k \\ \dot{Z}_k \end{pmatrix} = \begin{pmatrix} 1 & \Delta t \\ 0 & \alpha \end{pmatrix} \begin{pmatrix} Z_{k-1} \\ \dot{Z}_{k-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \beta \end{pmatrix} \nu_{k-1} \tag{22}$$

where $\dot{Z}_k$ is the velocity along the depth direction, Δt is the time interval between k and $k-1$, α and β are system parameters and $\nu \in \mathcal{N}(0, \sigma_\nu)$ is the stochastic velocity disturbance.

The objective of a tracking task is to recursively estimate the state $\mathbf{Z}_k$ from the observation $\mathbf{O}_k$ defined by:

$$\mathbf{O}_k = \mathbf{H}(\mathbf{Z}_k, \boldsymbol{\varepsilon}_k). \tag{23}$$

where $\mathbf{O}_k$ represents an observation vector at frame k. $\mathbf{H}$ is a possibly nonlinear function and vector $\boldsymbol{\varepsilon}$ is an i.i.d. observation noise sequence. We consider the image gradient as the observation: $\mathbf{O}_k = G_k$.

By testing numerous image sequences, it appears that equation (23) can be approximated using a quadric rational function (see Fig. 3):

$$G_k(Z_k) = \frac{p_0 + p_1 Z_k + p_2 Z_k^2}{q_0 + q_1 Z_k + Z_k^2} + \varepsilon, \quad p_2 \neq 0. \tag{24}$$

The distribution and the variance of the noise ε can be estimated by varying the depth position and observing the corresponding image sharpness.

The posterior predictive distribution of the state $\mathbf{Z}_k$ conditional on the observed image gradient $G_{1:k-1} = \{G_1, G_2, \ldots, G_{k-1}\}$ up to frame $k-1$ can be computed recursively:

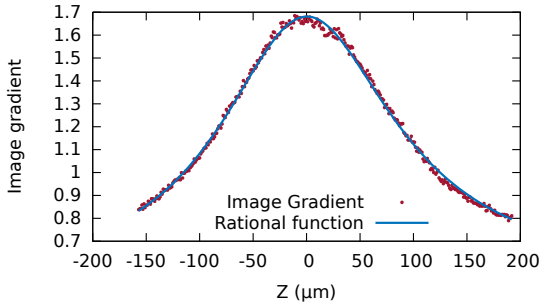

Fig. 3. Image gradient and its approximation using rational function with respect to depth position, respectively

$$p(\mathbf{Z}_k | G_{1:k-1}) = \int p(\mathbf{Z}_k | \mathbf{Z}_{k-1}) p(\mathbf{Z}_{k-1} | G_{1:k-1}) \mathrm{d}\mathbf{Z}_{k-1} \tag{25}$$

According to Bayes' theory, at kth frame the posterior can be updated with the observation G_k:

$$p(\mathbf{Z}_k | G_{1:k}) = \frac{p(G_k | \mathbf{Z}_k) p(\mathbf{Z}_k | G_{1:k-1})}{p(G_k | G_{1:k-1})} \tag{26}$$

where the normalization constant $p(G_k | G_{1:k-1})$ depends on the observation likelihood $p(G_k | \mathbf{Z}_k)$ defined by the observation model (23). Applying sequential importance sampling, the posterior density $p(\mathbf{Z}_k | G_{1:t})$ is then approximated using a set of weighted particles (random samples) $\{\mathbf{Z}_k^i, \omega_k^i\}$ where ω_k^i represents the weight of $\mathbf{Z}_k^i$:

$$p(\mathbf{Z}_k | G_{1:k}) \approx \sum_{i=1}^{N_p} \omega_k^i \delta(\mathbf{Z}_k - \mathbf{Z}_k^i) \tag{27}$$

where $\delta(\cdot)$ is Dirac delta measure ($\delta(\mathbf{Z}_k - \mathbf{Z}_k^i) = 1$ when $\mathbf{Z}_k = \mathbf{Z}_k^i$) and N_p is number of particles. Usually, the weighted particles can be updated using [16]:

$$\omega_k^i \propto \omega_{k-1}^i p(G_k | \mathbf{Z}_k^i) \tag{28}$$

In our tracking framework, we model the observation likelihood $p(G_k | \mathbf{Z}_k)$ using a registration error $\epsilon_k = \| G_k - \mathbf{H}(\hat{\mathbf{Z}}_k) \|^2$:

$$p(G_k | \mathbf{Z}_k) \propto e^{-\tau \epsilon_k} \tag{29}$$

where $\tau \in \mathbb{R}^+$ is a constant.

In our tracking and position estimation framework, a range of particles (with depth position and velocity along the depth direction) are generated randomly (in a given range) and assigned the same weight at first. For each frame in the tracking stage, the image gradient of the current image is computed and the particles are updated using the system dynamics model (equation (22)). The weight of each particle is then recomputed according to equation (28). The estimation of the state is then computed through equation (27).

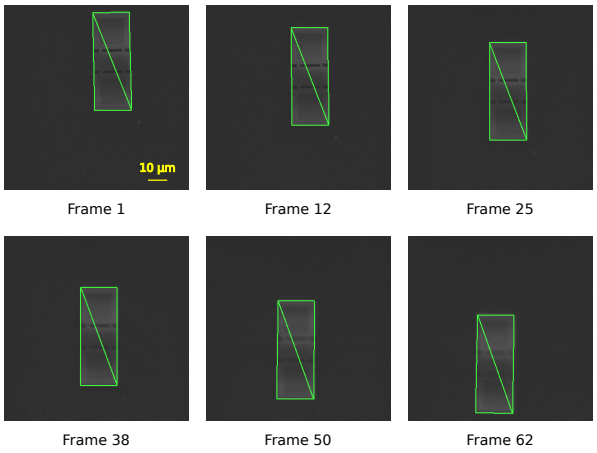

Fig. 4. Snapshots in visual tracking using proposed method, with medium scan speed

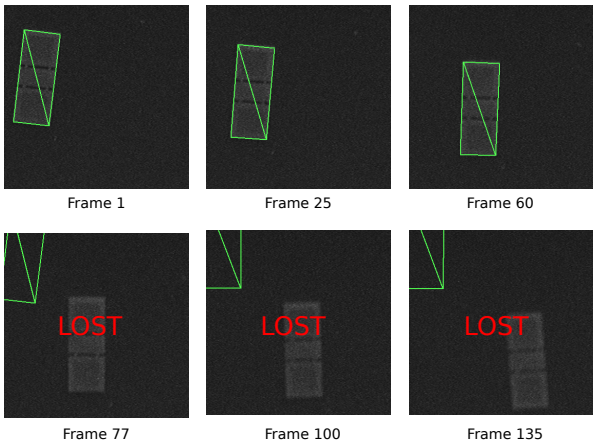

Fig. 5. Snapshots in visual tracking using traditional SSD method, with high scan speed

IV. Experimental validations

Experiments have been performed to evaluate the proposed visual tracking framework in the presence of defocus blur. The sample is a indium phosphide and silicon thin membrane, $20\mu m$ x $10\mu m$ and $200nm$ thick. Images (size 360×360 pixels) are acquired in the SEM Zeiss EVO 25 LS (at ISIR-UPMC, France). The sample was positioned on 4 DoFs (translations along x-, y-, and z-axes, rotations around z-axis) and the magnification is fixed at $1000\times$ during the visual tracking task.

A. Experimental validations of visual tracking

First experiment has been performed with a medium scan speed (about 3.3 µs/pixel) of the SEM. Fig. 4 shows the snapshots of some frames in the experiments. The sample becomes blurred since its position varies along the depth direction. Although the rotation around z-axis varies slightly (about 0.04 degree/frame), the evolution of the angle can still be estimated during the visual tracking task.

In order to evaluate the proposed approach with respect to the traditional SSD-based one in noisy conditions, an experiment has been performed at a high scan speed (about 0.72 µs/pixel) using the same sample at the same magnification. The snapshots of the experiments using the two methods above are shown in Fig. 5 and 6, respectively. It is found in the figures that the tracking task could fail using traditional SSD-based method if the image is highly degraded from blur and noise. A reason is that traditional SSD-based template matching method consider only the geometrical transformation of the object on geometry. When blur is present in the images, the dissimilarity function no longer applies. Alternatively, the proposed method shows robustness since the image blur is modeled in the minimization process of the dissimilarity function.

B. Experimental results on posture estimation

Experiments have been performed to evaluate the estimation of the position and the orientation of the object. The images are acquired with a high scan speed (about

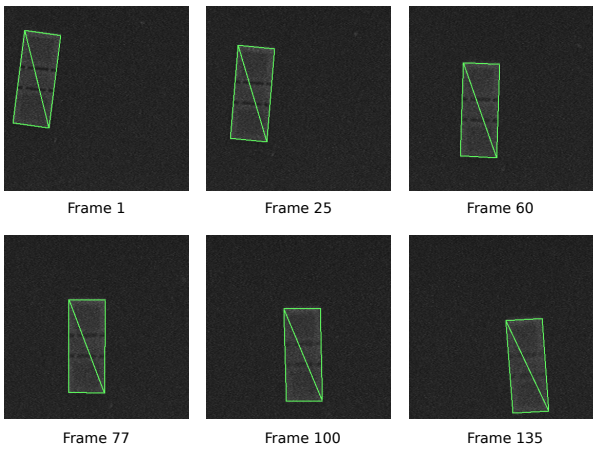

Fig. 6. Snapshots in visual tracking using proposed method, with high scan speed

0.72 µs/pixel) at $1000\times$. An image sequence is acquired in the same condition of the SEM by varying the position on the depth direction to provide the training data. In this experiment, the sample moves on 4 DoFs as previous experiments.

To compute the posture of the object from 3D registration, the calibration process [10] has been performed to provide the SEM intrinsic parameters. Fig. 7 shows the evolution of the position on x- and y-axes and rotation around z-axis estimated by 3D registration. Small oscillations are found in the estimation of the rotation around z-axis (yellow curve in the figure). Actually, since the increment of this rotation is small, corresponding to a small displacement on image, it is difficult to determine this value accurately from the blurred image.

The position of the sample on the depth direction is estimated using the proposed image gradient and particle filter-based approach. Actually, in the visual tracking task, the optimization process of both blur level and displacement are performed simultaneously. Considering the high noise level on the SEM image, the cost function computation for the blur level estimation could be affected by the noise and

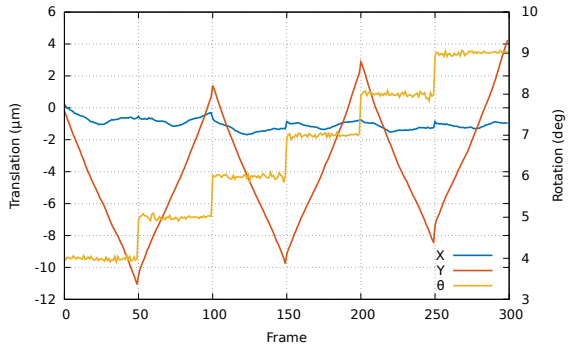

Fig. 7. Estimated position on x, y and orientation around z

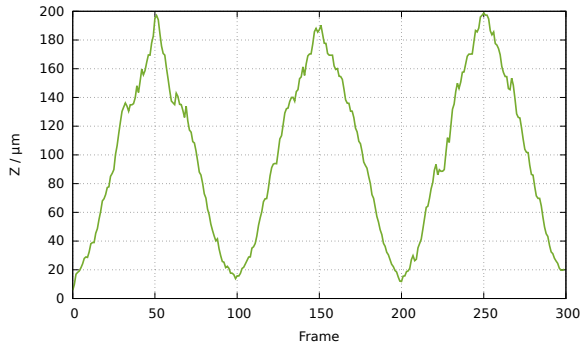

Fig. 8. Evolution of estimated position on the depth direction Z with respect to frames

the variation of the displacement during the warp process. Since the image gradient is computed directly from the tracked zone, it is more reliable than the blur level that is estimated using optimization process. In the experiment using particle filter, the number of particles is set to 200. In the experiments, we find that this number represents a good compromise between the performance and the time consumption in our experiments. A very large number of particles does not obviously improve the performance in our experiments. Fig. 8 shows the results of the estimation of the sample position on the depth direction. The sample motion along the depth direction is clearly shown by the estimated position along the depth direction. It can be find that this estimation is less accurate than the estimation on the other DoFs computed by the 3D registration. This is mainly because that the image gradient could be highly affected by the image noise at a high scan speed in the SEM.

V. Conclusion

In this article, we address a three-dimensional visual tracking and posture estimation approaches for SEM applications. To overcome the problem that the extraction of the visual feature could be no longer reliable in the presence of the defocus blur when the sample is moved along the depth direction, we propose to consider the defocus blur level in the template-based visual tracking scheme. In our method, the posture of the object is estimated in three dimensions. The positions on x- and y-axes and the rotation around the z-axis are estimated by a 3D registration-based method. Since the motion along the depth direction can not be observed by the scale of the sample, we propose to use the particle filter to estimate the motion along the depth direction by observing the image sharpness. The proposed approaches are validated by the experiments in a SEM at 1000$\times$ in 4 DoFs. The further work will be improving the accuracy on the depth position estimation and applying this method in visual guidance of automated micro/nano-positioning.

Acknowledgment

This work has been realized in the context of the French ANR P2N Nanorobust project. The microrobotic stage is supported by French Robotex platform. The authors would like to acknowledge Camille Dianoux and Jean-Ochin Abrahamians for their help in the experiments at ISIR.

References

[1] T. Sievers and S. Fatikow, "Real-time object tracking for the robot-based nanohandling in a scanning electron microscope," *Journal of Micromechatronics*, vol. 3, no. 3, pp. 267–284, 2006.

[2] S. Fatikow, T. Wich, H. Hülsen, T. Sievers, and M. Jähnisch, "Microrobot system for automatic nanohandling inside a scanning electron microscope," *Mechatronics, IEEE/ASME Transactions on*, vol. 12, no. 3, pp. 244–252, 2007.

[3] S. Fatikow, V. Eichhorn, C. Stolle, T. Sievers, and M. Jähnisch, "Development and control of a versatile nanohandling robot cell," *Mechatronics*, vol. 18, no. 7, pp. 370–380, 2008.

[4] D. Jasper and S. Fatikow, "Line scan-based high-speed position tracking inside the sem," *International Journal of Optomechatronics*, vol. 4, no. 2, pp. 115–135, 2010.

[5] B. E. Kratochvil, L. Dong, and B. J. Nelson, "Real-time rigid-body visual tracking in a scanning electron microscope," *The International Journal of Robotics Research*, vol. 28, no. 4, pp. 498–511, 2009.

[6] B. Tamadazte, E. Marchand, S. Dembélé, and N. Le Fort-Piat, "Cad model-based tracking and 3d visual-based control for mems microassembly," *The International Journal of Robotics Research*, 2010.

[7] C. Ru, Y. Zhang, H. Huang, and T. Chen, "An improved visual tracking method in scanning electron microscope," *Microscopy and Microanalysis*, vol. 18, no. 03, pp. 612–620, 2012.

[8] 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," *Nanotechnology, IEEE Transactions on*, vol. 10, no. 4, pp. 674–681, 2011.

[9] Z. Gong, B. K. Chen, J. Liu, and Y. Sun, "Robotic probing of nanostructures inside scanning electron microscopy," *Robotics, IEEE Transactions on*, vol. 30, no. 3, pp. 758–765, 2014.

[10] L. Cui and E. Marchand, "Calibration of scanning electron microscope using a multi-images non-linear minimization process." in *IEEE Int. Conf. on Robotics and Automation, ICRA'14*, 2014.

[11] N. Cornille, D. Garcia, M. A. Sutton, S. McNeill, and J.-J. Orteu, "Automated 3-d reconstruction using a scanning electron microscope," in *SEM annual conf. & exp. on experimental and applied mechanics*, 2003.

[12] C. Dahmen, "Focus-based depth estimation in the sem," in *International Symposium on Optomechatronic Technologies*. International Society for Optics and Photonics, 2008, pp. 72 661O–72 661O.

[13] ——, "Defocus-based threedimensional tracking in sem images," in *Informatics in Control Automation and Robotics*. Springer, 2011, pp. 243–254.

[14] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *Int. Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.

[15] F. Nicolls, G. de Jager, and B. Sewell, "Use of a general imaging model to achieve predictive autofocus in the scanning electron microscope," *Ultramicroscopy*, vol. 69, no. 1, pp. 25 – 37, 1997.

[16] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, 2002.