Full 3D Rotation Estimation in Scanning Electron Microscope

Andrey Kudryavtsev, Sounkalo Dembele, Nadine Piat

Full 3D Rotation Estimation in Scanning Electron Microscope

Andrey V. Kudryavtsev, Sounkalo Dembélé and Nadine Piat

Abstract— **Estimation of 3D object position is a crucial step for a variety of robotics and computer vision applications including 3D reconstruction and object manipulation. When working in microscale, new types of visual sensors are used such as Scanning Electron Microscope (SEM). Nowadays, micro- and nanomanipulation tasks, namely components assembly, are performed in teleoperated mode in most of the cases. Measuring object position and orientation is a crucial step towards automatic object handling. Current methods of pose estimation in SEM allow recovering full object movement using its computer-aided design (CAD) model. If the model is not known, most methods allow to estimate only in-plane translations and rotation around camera optical axis. In the literature, SEM is considered as a camera with parallel projection or an affine camera, which means image invariance to z-movement and bas-relief ambiguity. In this paper, authors address the problem of measuring full 3D rotation of the unknown scene for uncalibrated SEM without additional sensors. Rotations are estimated from image triplets by solving a spherical triangle from fundamental matrices only, without need of intrinsic calibration, allowing to avoid parallel projection ambiguities. The presented results, obtained in simulation and on real data, allow validating the proposed scheme.**

I. INTRODUCTION

Estimation of 3D object position and orientation is a crucial step for a variety of robotics and computer vision applications including 3D reconstruction and object manipulation. Due to the non-stopping development of micro-nanoscale technologies, new types of visual sensors started being used for providing visual feedback for vision-based control approaches in microscale. One of the examples of such sensors is a SEM. Nowadays, it is mostly used as a visualization tool, however, by applying computer vision theory it can be transformed into a measuring instrument.

When working at microscale, special imaging conditions should be taken into account. SEM is considered as a camera with parallel projection (or an affine camera) for magnification values bigger than $\times 1000$, which is confirmed in the literature [1], [2]. This statement is supported by the following fact: when the distance between camera and object becomes much bigger than the object itself, the projection rays become parallel, which is the case when working with SEM, especially at high magnification. This distinctive feature of SEM imaging raises the following ambiguities. First, images are invariant to motion along optical axis (usually z) which makes very challenging the estimation of depth coordinate (Fig. 1,a): most existing solutions are based on focus
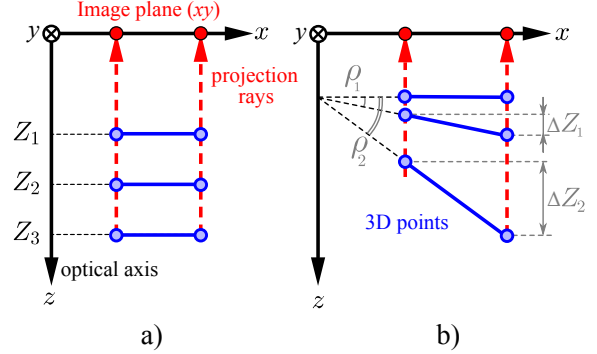

Fig. 1: Motion ambiguities under parallel projection: a) invariance of 2D projection (image) to z translation, b) bas-relief ambiguity.

information and not motion [3], [4]. Secondly, bas-relief ambiguity, which consists in fact that, the 2D projection of two different objects may be the same depending on combination of depth variation ΔZ and fronto-parallel rotation $\rho \in$ SO(3) (Fig. 1,b) [5]. This property creates big limitations to the stereoscopy in SEM, because it also results in impossibility of measuring out-of-plane rotations R_x and R_y from two images only. Recent works demonstrate that it is possible to extract the information about object position from SEM images but only using some additional information. They can be subdivided in two groups.

The first group of motion estimation techniques is based on pose computation. In [6], positions along x- (t_x) and y-axis (t_y), and rotation about an optical axis (R_z) were computed using Gauss-Newton method by minimizing the sum of the projection errors of some points of a pre-defined 2D model. Authors assumed that the object, while moving, stays on the plane parallel to the image plane. However, if the object performs a more complex 3D movement, the results may be inaccurate as the impact of rotations R_x and R_y cannot be neglected. In [7], the developed solution based on augmented reality approach used the search for 3D CAD model in the 2D images of SEM to minimize the distance between lines extracted from images and those of the model. The pose, comprising position and orientation, was computed efficiently, but the method works only for polyhedral structures. In another example, while working on 3D reconstruction in SEM, Tafti *et al.* obtain the full object motion information but the tilt angle of the stage and SEM calibration matrix were known [8].

The second group of motion estimation methods is based on the matching of a pre-defined 2D model. In [9] and

[10], cross-correlation was used, whereas in [11] model was represented by active contours and motion was estimated from the minimization of the active contours and the object detected edges. Both methods have the same drawback of [6]: as the rotations R_x and R_y are not considered, measured quantities may be inaccurate. In [12] an interesting method was proposed, it is based on the work described in [13] which used spherical Fourier transform to compute the rotations R_x, R_y and R_z. The method is promising, however, it was only tested on simulated SEM images and not in real conditions.

A. Contribution

The present work describes a method allowing to fully recover the 3D object rotations not by using robot sensors or defocus information but by adding one more image only, thus, a method of measuring the full 3D rotation from three SEM images. While being based on epipolar geometry and spherical trigonometry, the usage of image triplets allows avoiding the ambiguities of parallel projection presented above. An additional advantage associated with the method is that the SEM calibration (instrinsic parameters matrix) is not needed. All test images were obtained by moving the robotic stage inside the SEM, however, for the ease of demonstration, we consider that not the object but the camera (SEM) is moving and the object is fixed (the situations are geometrically equivalent).

The paper is organized as follows: Section II presents general information about image formation under parallel projection assumption. Then, in Section III, the geometry of motion between two and three image frames is analyzed and presented which leads to the equations of spherical trigonometry allowing to recover all three components of rotation. After theoretical developments, Section IV presents the validation results, first, on simulation with virtual image sequence, and then on real images of a pollen grain and an end-effector of a microgripper.

II. PARALLEL PROJECTION MODEL

The parallel projection of a 3D point $\mathbf{Q} = (Q_x, Q_y, Q_z, 1)^\top$ (in homogeneous coordinates) to the image frame can be written as follows:

$$\mathbf{q} = \mathbf{K}\mathbf{\Pi}^c\mathbf{T}_w\mathbf{Q} \tag{1}$$

with $\mathbf{K} = \begin{pmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\mathbf{\Pi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ where f_x and f_y define the focal distances in x and y directions of image frame. $^c\mathbf{T}_w$ denotes a 4×4 homogeneous matrix that describes the transformation between world and camera frames. It can be further decomposed in rotation and translation components,

$$^c\mathbf{T}_w = \begin{pmatrix} ^c\mathbf{R}_w & ^c\mathbf{t}_w \\ \mathbf{0}_{1\times3} & 1 \end{pmatrix} \tag{2}$$

The fact of parallel projection imposes the following properties on the process of image formation. First, image is invariant to the object displacement along the camera
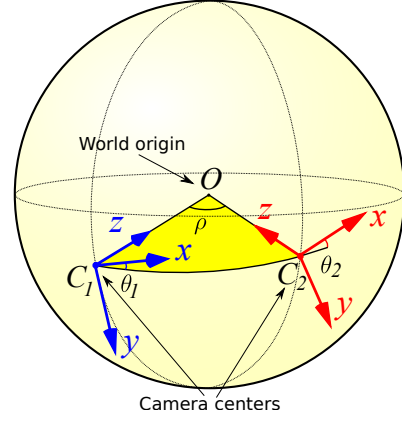

Fig. 2: Geometry of motion between two views taken with an affine cameras C_1 and C_2. Transformation between $\mathcal{R}_1$ and $\mathcal{R}_2$ is a composition of three rotations with angles θ_1, ρ, and θ_2. O is the centroid of 3D points with coordinates $(0, 0, 0, 1)^\top$.

optical axis. In other words, image is invariant to the distance between camera and object, which is the case in SEM. For instance, moving the sample closer to the electron beam or moving it away will not change the resulting 2D projection. As a result, the depth coordinate is lost in the process of image formation in case of parallel projection.

Secondly, the projection of the object is independent of translations in x and y directions of image frame if the relative coordinates are used both in 3D object frame and in camera frame [14]. If such translation is performed, only the position of the object in the image changes, but not the disposition of its feature points. For any given reference point $(q_x^r, q_y^r)^\top$ in image frame and $(Q_x^r, Q_y^r, Q_z^r)^\top$ in world frame, the expressions for relative coordinates ($\check{q}$ in image frame and $\check{Q}$ in a world frame) can be written as follows:

$$\begin{pmatrix} \check{q}_x \\ \check{q}_y \\ 1 \end{pmatrix} = \begin{pmatrix} q_x - q_x^r \\ q_y - q_y^r \\ 1 \end{pmatrix} \tag{3}$$

$$\begin{pmatrix} \check{Q}_x \\ \check{Q}_y \\ \check{Q}_z \\ 1 \end{pmatrix} = \begin{pmatrix} Q_x - Q_x^r \\ Q_y - Q_y^r \\ Q_z - Q_z^r \\ 1 \end{pmatrix} \tag{4}$$

Thus, for every camera, we can choose a reference point defined as the centroid of the set of points detected in this image and furthermore work with relative coordinates. It results in a fact that the centroid of the 3D points in world frame is projected into the centroid of the points in the image plane. The centroid of 3D points is considered having coordinates $(0, 0, 0, 1)^\top$.

These two properties allow drawing the following conclusions:

- Parallel projection camera is invariant to translations in image frame, relative coordinates can be used: all detected interest points are translated into $(0, 0, 1)^\top$. It means that the centroid of 3D points is in $(0, 0, 0, 1)^\top$.
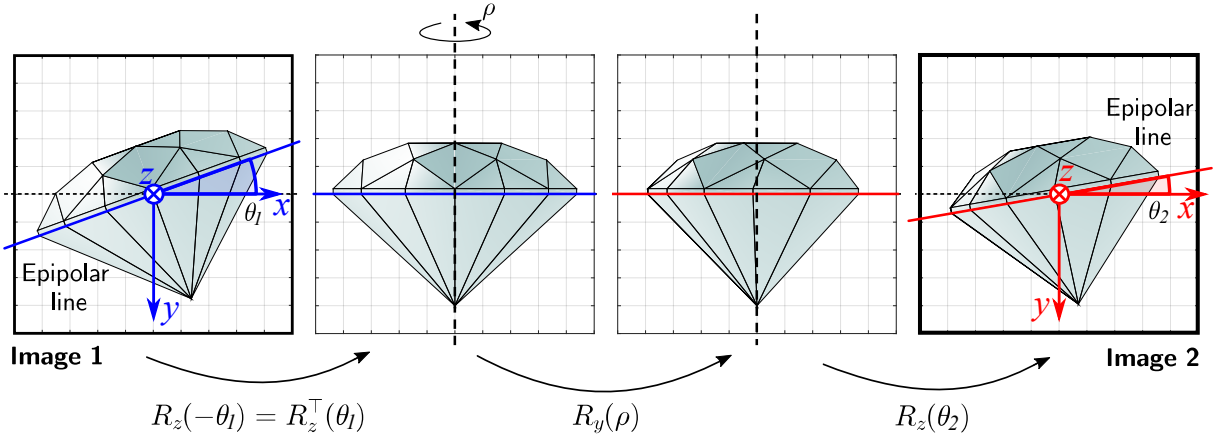
Fig. 3: Rotations decomposition between two images.

$$R_z(-\theta_1) = R_z^\top(\theta_1) \qquad R_y(\rho) \qquad R_z(\theta_2)$$

Therefore, in the relative coordinates, for all taken images z-axis of camera is pointing towards the same point, the world origin;

- From the second property, as the object-camera distance is chosen to be one for all views, all camera centers are at the same distance from the world origin.

Thus, when using relative coordinates, at every iteration of movement, the camera lies on the surface of the sphere (Fig. 2).

III. MOTION GEOMETRY

A. Two Views

For the remainder of this paper, it is assumed that the frame of the first camera $\mathcal{R}_1$ coincides with the world frame, that

$$^1\mathbf{R}_w = \mathbf{I}_{3\times3} \tag{5}$$

Consider the situation where two views were obtained. If feature points are already detected and matched, so that 2D projections $^1\mathbf{q}$ in left image frame match the projections $^2\mathbf{q}$ in the right image frame, the centroids can be translated to $(0,0,1)^\top$ in order to further use relative coordinates. It leads to the configuration represented in Fig. 2. It is important to notice that the plane C_1OC_2, that passes through camera centers and world origin, represents the epipolar plane. It intersects each camera's image plane where it forms the epipolar lines. Thus, the angles θ_1 and θ_2 are the slopes of epipolar lines in image I_1 and I_2 respectively, and will be further referred as slope angles.

As a result, the rotation between two views C_1 and C_2 can be then decomposed as follows:

$$^2\mathbf{R}_1 = \mathbf{R}_z(\theta_2)\mathbf{R}_y(\rho)\mathbf{R}_z^\top(\theta_1) \tag{6}$$

The steps of transformation between images are represented in Fig. 3.

In order to measure slope angles, one can use epipolar geometry. It represents a set of geometric constraints between 3D points and their projections onto two 2D images. Algebraically, epipolar geometry constraints are encapsulated in the 3×3 matrix $\mathbf{F}$, fundamental matrix. In case of parallel projection, it has a special form:

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{pmatrix} \tag{7}$$

where e is generally chosed to be one as the matrix is normalized, and a,b,c,d are real numbers. In the present work, we use Gold Standard algorithm dedicated for camera with parallel projection [5] to estimate the fundamental matrix from a set of point correspondences across the images. In addition, we use this algorithm inside RANSAC scheme that allows filtering outliers in the correspondence set and increasing the robustness of fundamental matrix estimation. Being iterative, the RANSAC scheme implies a random selection of a correspondence subset for further model estimation. It should be noted that in case of parallel projection at least four point correspondences are needed to find $\mathbf{F}$ which represents the minimal configuration. Then, from the chosen subset, fundamental matrix is estimated using Gold Standard algorithm and the number of inliers is compared with the one on previous iterations. These steps are repeated until the maximum number of inliers is found. In other words, the Gold Standard algorithm is used to fit the model (fundamental matrix) to a randomly selected set of four point correspondences. The Gold Standard algorithm comprises the following steps. Assume one correspondence is represented by the vector $\mathbf{c}_i$:

$$\mathbf{c}_i = \left(q'_x, q'_y, q_x, q_y\right)^\top \tag{8}$$

Then, in order to work with relative coordinates all points are centered in $(0,0)^\top$:

$$\check{\mathbf{c}}_i = \mathbf{c}_i - \bar{\mathbf{c}} \tag{9}$$

where $\bar{\mathbf{c}}$ is the centroid of points computed as:

$$\bar{\mathbf{c}} = \frac{1}{N}\sum_i^N \mathbf{c}_i \tag{10}$$

with N the total number of correspondences found. It allows the construction of $N\times4$ matrix $\mathbf{A}$ with rows $\check{\mathbf{c}}_i^\top$. Then, if
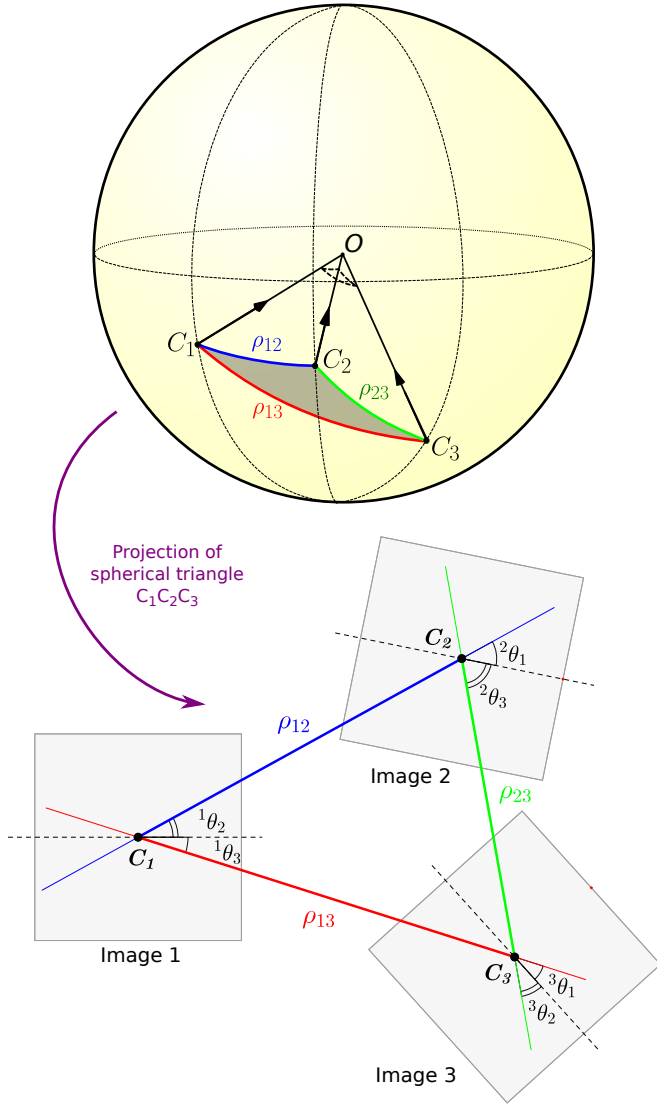
Fig. 4: Three-view geometry: camera centers C_1, C_2 and C_3 are located on the surface of a unit sphere and form a spherical triangle.

the singular vector corresponding to the smallest singular value of $\mathbf{A}$ is denoted as $\mathbf{N}$, all five elements of $\mathbf{F}$ can be found using:

$$\begin{aligned} (a, b, c, d) &= \mathbf{N}^\top \\ e &= -\mathbf{N}^\top \bar{\mathbf{c}} \end{aligned} \qquad (11)$$

The fundamental matrix is then obtained using (7) and normalized to have $e = 1$. The slopes of epipolar lines are then estimated as follows:

$$\theta_l = \arctan\left(-\frac{d}{c}\right), \theta_r = \arctan\left(-\frac{a}{b}\right) \qquad (12)$$

Thus, from two views it is possible to recover two out of three rotation angles. However, ρ angle can not be obtained from only two views due to the bas-relief ambiguity as it was stated previously. The solution proposed in this paper is to use three-views to solve this problem.

B. Three-view Geometry

Three view configuration in case of an affine camera is represented in Fig. 4. Assuming that relative coordinates are used, the origins of all image frames lie on the sphere with the world origin O as its center. The polygon $C_1 C_2 C_3$ (on the surface of the sphere) represents a spherical triangle as it is formed by intersection of three great circles, one for each pair of frame centers. Great circle is a circle that has the same radius as the sphere. It means that it is possible to use the whole branch of geometry describing the relations between sphere elements, spherical geometry and trigonometry. Spherical triangle has the following properties.

1) Angles $\sphericalangle C_1, \sphericalangle C_2, \sphericalangle C_3$ (symbol $\sphericalangle$ denotes spherical angle). The angles of the triangle are equal to the the angles between the tangent vectors of the great circle arcs where they meet at the vertices. In our case, it represents the angle between epipolar lines from two other images, e.g., the angle $\sphericalangle C_1$ is measured in the first view as the angle between epipolar lines from images 2 and 3 (Fig. 4):

$$\sphericalangle C_1 = {}^1\theta_2 - {}^1\theta_3 \qquad (13)$$

where ${}^i\theta_j$ is the slope of epipolar line in image i defined by image j. These angles are found using 12.

2) Sides $C_1 C_2$, $C_1 C_3$ and $C_2 C_3$. According to the theory of spherical geometry, on the unit sphere the lengths of the sides of spherical triangle are numerically equal to the radian measure of the angles that the great circle arcs subtend at the centre. It means that, revising our configuration, one can conclude that the sides of the triangle are equal to the angles ρ that could not be measured in two-view case:

$$\rho_{12} = \angle C_1 O C_2 \qquad (14)$$

Thus, we have all the elements of the triangle expressed using rotational parameters of camera positions:

$$\begin{aligned} \sphericalangle C_1 &= {}^1\theta_2 - {}^1\theta_3 \\ \sphericalangle C_2 &= {}^2\theta_1 - {}^2\theta_3 \\ \sphericalangle C_3 &= {}^3\theta_1 - {}^3\theta_2 \end{aligned} \qquad (15)$$

and

$$\begin{aligned} \rho_{12} &= C_1 C_2 = \angle C_1 O C_2 \\ \rho_{13} &= C_1 C_3 = \angle C_1 O C_3 \\ \rho_{23} &= C_2 C_3 = \angle C_2 O C_3 \end{aligned} \qquad (16)$$

In the presented configuration, the angles of the spherical triangle are recovered using image pairs and Eq. 12. The problem of solving spherical triangle, with its angles known, is quite common and can be solved by applying a supplemental cosine law of spherical trigonometry. Using presented notations, it has the following form:

$$\begin{aligned} \rho_{12} &= \arccos\left(\frac{\cos(C_3) + \cos(C_1)\cos(C_2)}{\sin(C_1)\sin(C_2)}\right) \\ \rho_{13} &= \arccos\left(\frac{\cos(C_2) + \cos(C_1)\cos(C_3)}{\sin(C_1)\sin(C_3)}\right) \\ \rho_{23} &= \arccos\left(\frac{\cos(C_1) + \cos(C_2)\cos(C_3)}{\sin(C_2)\sin(C_3)}\right) \end{aligned} \qquad (17)$$

Thus, all rotational parameters are recovered: slope angles ${}^j\theta_i$ were calculated using (12) and ρ_{ij} using (17). The rotation ${}^j\mathbf{R}_i$ matrix can be then obtained by substituting these values in (6).

IV. METHOD VALIDATION

In order to evaluate the performance of the proposed method, two types of experiment were conducted. First, the method was tested on the manually generated sequence of virtual images using MATLAB. Secondly, 2 image datasets coming from SEM Carl Zeiss Auriga 60 were used. The features were obtained using AKAZE descriptors [15] from OpenCV library [16] and then matched. Authors of this algorithm declare that it outperforms BRISK, ORB, SURF and SIFT in most of the scenarios. The fundamental matrices were estimated using Gold Standard algorithm inside RANSAC scheme as it was stated previously.

A. Validation through simulation

Virtual image sequence represents an image set containing 150 images of a diamond (Fig. 3) with predefined pose. The orientation of the object (diamond) was estimated for all frames using the method presented above. In the absence of noise, the resulting graphs (Fig. 5) allow to compare the estimated values with the predefined ones. As the result the error stays inferior to 1 microdegree for all orientation Euler angles, which allows first validation of the method. It is important to notice that it is possible to measure object orientation for the second image but only at the moment when the third one becomes available. This fact is also reflected in Fig. 5. Next step consisted in a test of method robustness: for this, random noise was added to the coordinates of the extracted points (Fig. 6). For noise amplitudes below 0.5 pixel, the mean error of rotation estimation stays inferior to 0.1 degrees for all three rotation components. With further increase of noise (up to 1 pixel), the error don't exceed the value of 0.5 degrees. This error comes from the errors in fundamental matrices as the quality of correspondences degrades with increasing noise.

B. Real SEM images

Estimation of camera rotations was conducted on two SEM image datasets. First, seven images of *Potamogeton*, a pollen grain of an aquatic plant acquired with Carl Zeiss AU-RIGA 60 FE-SEM (Fig. 7(a)). The rotation was performed by tilting the stage of 3 degrees for every image. Second, *End effector* dataset which contains seven images of a tip of the end effector of a microgripper (Fig. 7(b)). The movement between images was realized using a 6-DoF robot mounted inside the microscope, first, by the steps of 3 degrees about each axis and then by 5 degrees. Features were extracted and matched, the fundamental matrices were found using Gold Standard algorithm inside RANSAC scheme and the rotations were then recovered using the presented method. The estimated values are presented in Table I. Obtained angles are very close to true ones with a deviation inferior to 0.1 degree. The error can be explained by the fact that during image acquisition the axis of tilt was not ideally vertical for the first dataset, i.e., the tilt was not a pure rotation around y-axis.
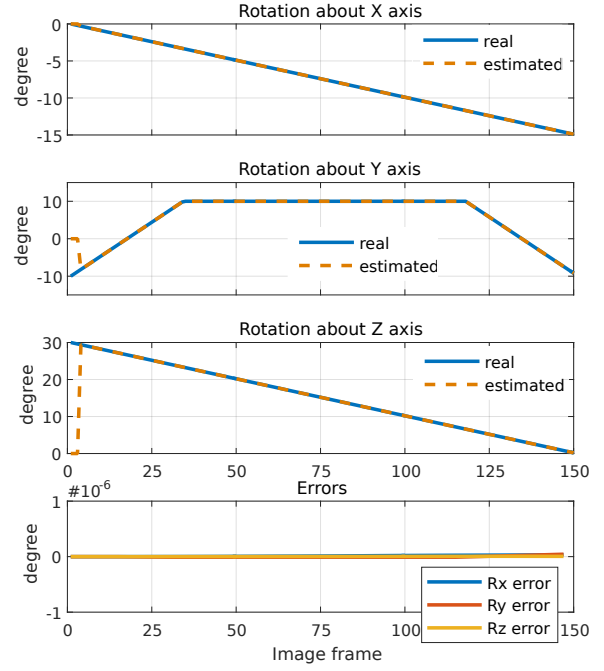

Fig. 5: Errors in orientation measurement for virtual image sequence of a diamond. Sequence contains 150 images.
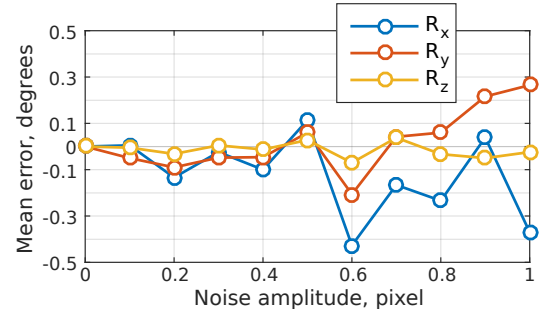

Fig. 6: Mean error of rotations estimation for a virtual image sequence with noise injected.

TABLE I: Comparison between true Euler angles and estimated (in degrees) for two SEM image datasets, *Potamogeton* and *End effector*.

|	Euler angles, in degrees			Estimated angles, in degrees		
	R_z	R_y	R_x	$\hat{R}_z$	$\hat{R}_y$	$\hat{R}_x$
Potamogeton:						
I_1	0	0	0	0	0	0
I_2	0	3.00	0	-0.03	3.02	0.07
I_3	0	6.00	0	0.07	6.03	0.10
I_4	0	9.00	0	0.19	9.02	0.17
I_5	0	12.00	0	0.19	12.03	0.22
I_6	0	15.00	0	0.38	15.06	0.29
I_7	0	18.00	0	0.33	18.04	0.32
End effector:						
I_1	0	0	0	0	0	0
I_2	0	0	3.00	0	0	2.96
I_3	0	3.00	3.00	0	2.98	2.96
I_4	3.00	3.00	3.00	2.98	2.98	2.96
I_5	3.00	3.00	8.00	2.98	2.98	8.00
I_6	3.00	8.00	8.00	2.98	7.93	8.00
I_7	8.00	8.00	8.00	7.97	7.93	8.00
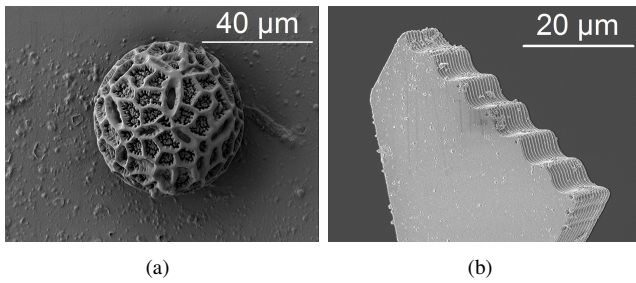
Fig. 7: One of the images from: a) *Potamogeton* dataset, pollen grain (magnification: ×1000, image size: 2048 × 1536); b) *End effector* dataset, end effector of a microgripper (magnification: ×2000, image size: 1024 × 768).

V. CONCLUSIONS

The paper has investigated the problem of out-plane rotations estimation inside a SEM. At high magnification, starting from ×1000, projection rays becomes parallel which leads to the following motion ambiguities: invariance to translation along optical axis of the camera, and bas-relief ambiguity in stereo-image case. These ambiguities make challenging the computation of full 3D motion, particularly the three rotations. The paper stated how the stereovision geometry can be described through a spherical geometry. Then, considering three-view geometry, we showed the way to recover the three Euler angles through affine epipolar constraints and spherical trigonometry without any use of additional sensors. The method was applied to virtual images and the error between real angles and their estimations was about 1 microdegree. It was then applied to two sets of images acquired with Carl Zeiss AURIGA 60 FE-SEM. The resulting angles differ from the true ones for about 0.1 degree. It can be explained by the fact, that during image acquisition it was assumed that the stage was perfectly parallel to the image plane, which was not the case. The experiments were conducted on different magnifications: ×1000 and ×2000. It should be noted that the presented method can be used in situations where parallel projection model is applicable, thus, for magnification ×1000 and bigger. Being based on algebraic equations, the method is easy to implement. It resolves affine camera ambiguities present in stereo case, and can be used for an arbitrary 3D rotational motion.

Future work will be focused on the deepening of the solution, particularly its application to automatic nanoobjects handling, assembly and characterization using vision-based control schemes. It should be noted that in case of 3D rotation, the translations cannot be recovered using registration technique due to the presence of out-of-plane rotations. This situation can be avoided if, in visual servoing scenario, one starts with rotation compensation. Then, when current object orientation is equal to desired one, it is possible to use registration methods (interest points or Fourrier-based methods) to measure translation in image plane.

REFERENCES

[1] N. Cornille, D. Garcia, M. A. Sutton, S. McNeill, and J.-J. Orteu, "Automated 3-d reconstruction using a scanning electron microscope," in *SEM Conference on Experimental and Applied Msechanics*, 2003.

[2] L. Cui and E. Marchand, "Calibration of scanning electron microscope using a multi-image non-linear minimization process," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 5191–5196.

[3] V. Eichhorn, S. Fatikow, T. Wich, C. Dahmen, T. Sievers, K. N. Andersen, K. Carlson, and P. Bøggild, "Depth-detection methods for microgripper based cnt manipulation in a scanning electron microscope," *Journal of Micro-Nano Mechatronics*, vol. 4, no. 1-2, pp. 27–36, 2008.

[4] N. Marturi, B. Tamadazte, S. Dembélé, and N. Piat, "Visual servoing-based depth estimation technique for manipulation inside sem." *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 8, pp. 1847–1855, 2016.

[5] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[6] L. Cui, E. Marchand, S. Haliyo, and S. Régnier, "Three-dimensional visual tracking and pose estimation in scanning electron microscopes," *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, October 2016.

[7] 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.

[8] A. P. Tafti, A. B. Kirkpatrick, Z. Alavi, H. A. Owen, and Z. Yu, "Recent advances in 3d sem surface reconstruction," *Micron*, vol. 78, pp. 54–66, 2015.

[9] 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.

[10] 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.

[11] S. Fatikow, T. Wich, H. Hulsen, T. Sievers, and M. Jahnisch, "Micro-robot system for automatic nanohandling inside a scanning electron microscope," *IEEE/ASME Transactions on Mechatronics*, vol. 12, no. 3, pp. 244–252, 2007.

[12] N. Marturi, B. Tamadazte, S. Dembélé, and N. Piat, "Image-guided nanopositioning scheme for sem," *IEEE Transactions on Automation Science and Engineering*, no. 99, pp. 1–12, 2016.

[13] P. J. Kostelec and D. N. Rockmore, "FFTs on the rotation group," *Journal of Fourier analysis and applications*, vol. 14, no. 2, pp. 145–179, 2008.

[14] L. Quan, "Self-calibration of an affine camera from multiple views," *International Journal of Computer Vision*, vol. 19, no. 1, pp. 93–105, 1996.

[15] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Trans. Patt. Anal. Mach. Intell*, vol. 34, no. 7, pp. 1281–1298, 2011.

[16] G. Bradski, "The OpenCV library," *Dr. Dobb's Journal of Software Tools*, 2000.