Depth Estimation of Optically Transparent Laser-Driven Microrobots

Maria Grammatikopoulou, Lin Zhang, and Guang-Zhong Yang, *Fellow, IEEE*

***Abstract*— Six degree-of-freedom (DoF) pose feedback is essential for the development of closed-loop control techniques for microrobotics. This paper presents two methods for depth estimation of transparent microrobots inside an Optical Tweezers (OT) setup using image sharpness measurements and model-based tracking. The *x-y* position and the 3D orientation of the object are estimated using online model-based template matching. The proposed depth estimation methodologies are validated experimentally by comparing the results with the ground truth.**

I. INTRODUCTION

Examining cells on an individual basis provides insights into their mechanical behaviour since properties such as Young's modulus are useful in determining cell pathology [1]. Precise cell manipulation could also improve procedures, such as in-vitro fertilization. OT are capable of highly precise and multi-point actuation. However, direct cell manipulation with the laser can cause photodamage. To avoid this problem, microrobots can be fabricated to handle the cell indirectly.

Three-dimensional visual feedback of microrobots is necessary in order to develop microrobotic systems with manipulation capabilities comparable to that of large-scale robotic systems. Depth estimation techniques proposed for the macro scale cannot be directly applied to the micro-scale due to the difference in the projection model that describes the microscope-camera system. In addition, in most microscope set-ups only one camera view is available.

Model-based approaches, using the microrobot's geometrical design for reference, estimate the microstructure's pose for assistance in microassembly [2], [3]. Shape estimation for haptic feedback has also been considered in [4]. A CAD-based optimization solution is proposed in [5] for intraocular microrobot localization. In [6], the velocity of helical micro-swimmers along the microscope z-axis is estimated for surface reconstruction using prior calibration of the microrobot's velocity along the z-axis. Depth recovery has been investigated for set-ups inside a Scanning Electron Microscope (SEM) [7]. Cui *et al.* [8] have proposed a method for estimating the 3D position and the rotation on the x-y plane. In [9], estimation of multiple depths for objects in different heights within the scene is performed using the sharpness measurements locally in the image. Other applications include depth recovery of microgrippers and Carbon Nanotubes (CNT) for automation purposes using variance-based sharpness calculations [10].

One difference between the SEM and the optical microscope set-up is that object transparency is visible in the latter. The transparency issue cannot be overcome by 3D-printing opaque objects since modifying the fabrication properties with respect to transparency has an impact on the relative refractive index between the object and the medium. As a result, features from within the micro-object are observable and therefore the sharpness measurement is a combination of overlapping depths. Hence, in optical micromanipulation the relative refractive index is directly connected to the amplitude of the grasping force by the laser beam. A transparent object complicates the pose estimation, nonetheless and crucially, less transparent or opaque objects hinder trapping force due to the refractive index.

In this paper, two methods for depth estimation of 3D-printed transparent microstructures within an OT set-up is proposed, which incorporates position tracking and model-based 3D orientation estimation. Two methods are presented; I) using the overall sharpness of the image frame which corresponds to features in multiples depths from within the object and II) utilizing the local sharpness around distinct features for single depth recovery at specific regions of the object. The first method aims to provide an overall measurement for the z-displacement of the object, while the second takes into account local changes of sharpness which might be attributed to object 3D rotation rather than displacement along the microscope z-axis. The second method takes into account the estimated orientation of the microstructure as proposed in [11]. The motivation for the global and the local approach stems from the fact that the overall sharpness measure can not infer directly the depth of a 3D microstructure in case of simultaneous 3D translation and rotation. The proposed methodology is validated through experimental results which are compared with ground truth data.

II. DEPTH ESTIMATION

A. Framework Overview

This section presents an overview of the proposed methodology for recovering the 3D pose of optically transparent laser-driven microrobots in an OT set-up. The main issues that are addressed are the unobservable motion along the z-axis and object transparency.

The first method presented in Section II-C describes a strategy for recovering the z-translation of a co-ordinate frame attached to the microstructure with respect to the focus plane. This strategy requires calibration for estimating the global sharpness model as a function of a single depth value that describes the object displacement along the microscope z-axis (Section II-C.1). By the term global sharpness model, we refer to the function that maps a sharpness value to

All authors are with the Hamlyn Centre for Robotic Surgery, Imperial College London, London, United Kingdom. (email: m.grammatikopoulou14@imperial.ac.uk)

the corresponding object z position relative to the focus plane approximated by a sum of Gaussian curves. The object depth is reconstructed by the set of functions derived from the calibration routine as explained in Section II-C.2. This method gives a depth estimation that describes the overall object displacement.

The second method attempts to estimate a local sharpness model for single features. By the term local sharpness, we denote the approximation of the mapping function that relates a single sharpness value to the corresponding depth of an individual object feature by one Gaussian curve. A calibration routine is also performed in this case for a number of object features (Section II-D.1). The corresponding local depths are estimated by solving an optimization problem as described in Section II-D.2.

The motivation for developing these two methods is that the first method does not tackle the problem of interfering object features at different depths that are visible due to transparency. In addition, a holistic focus measurement cannot distinguish whether changes in the sharpness are cause by displacement along the z-axis or by 3D object rotation, as this motion also results to parts of the object being focused or defocused.

B. Projection Model

The projection of a 3D point on the image plane by the microscope-camera system is described by the parallel projection model. We assume a known CAD model for the object. Let $\mathbf{P_{M_i}} \in \mathbb{R}^3$, $i = 1, .., N$ be the i_{th} vertex of the registered geometrical model expressed with respect to a fixed global co-ordinate frame and $\mathbf{P_{c_i}} \in \mathbb{R}^2$, $i = 1, .., N$ be the corresponding projection of the i_{th} point on the image plane. Hence, the projection $\mathbf{P_{c_i}}$ is equal to

$$\mathbf{P_{c_i}} = K\,P\,(R\,\mathbf{P_{M_i}} + \mathbf{t}) \tag{1}$$

$$K = \begin{bmatrix} p_x & 0 \\ 0 & p_y \end{bmatrix}, P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{2}$$

with K the intrinsic parameters of the microscope-camera system, P the projection matrix, $R \in \mathbb{R}^{3x3}$ the rotation matrix that describes the object's orientation and $\mathbf{t} \in \mathbb{R}^3$ the object's translation with respect to the global frame. As seen from the third column of matrix P in (2), the translation along the z-axis is not observable.

In the Sections II-C and II-D, the depth information is estimated using image sharpness measurements and the information derived from the model-based orientation estimation. It is worth noting that the transparency of the object must be taken into consideration when using sharpness measurements to recover the three-dimensional position of the microrobot in an OT environment.

C. Depth estimation using global sharpness model approximation

The first method aims to reconstruct the overall z-displacement of the object by using a global representation of the sharpness model. The full image feed is used for the sharpness calculation.

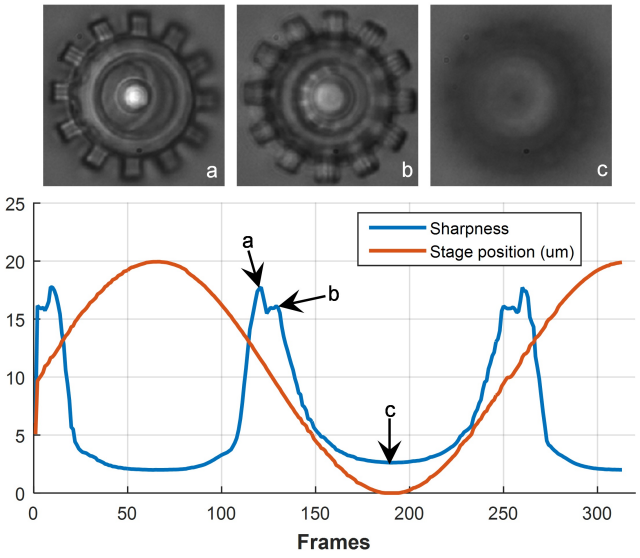

Fig. 1. Sharpness measurement and ground truth trajectory used in calibration for mapping function approximation

In order to estimate the translation along the z-axis of the object, the mapping between the sharpness measurement for each image frame and the corresponding z-position of the object must be established. The focus measure ϕ that is used in this case is the Gaussian Derivative [12]:

$$\phi = \sum (I * G_x)^2 + (I * G_y)^2,\; G(x,y) = \frac{1}{2\pi\sigma^2} e^{(-\frac{x^2+y2}{2\sigma^2})}$$
$$G_x = \frac{\partial G(x,y)}{\partial x},\; G_y \frac{\partial G(x,y)}{\partial y} \tag{3}$$

1) Global sharpness model approximation and calibration: A calibration routine is performed to approximate the mapping function between the sharpness and the displacement along z given a set of image frames and the corresponding ground truth z-translations of the piezo stage for each image frame. The calibration data is used to approximate the sharpness model as a function $\phi(z)$ of the overall object displacement z by fitting a sum of N Gaussian curves (Fig. 2 - top)

$$\phi(z) = \sum_{i=1}^{N} \alpha_i e^{-\frac{(z-b_i)^2}{{c_i}^2}} \tag{4}$$

where α_i the maximum amplitude of the i_{th} Gaussian term, b_i the depth at which the sharpness is maximum and c_i the standard deviation of the i_{th} Gaussian curve. An indicative set of calibration data is shown in Fig. 1, which include the depth measurements (orange), the image sequence and the corresponding sharpness values for each frame (blue). The calibration trajectory corresponds to scanning the microstructure along its height on an upwards and a downwards trajectory.

2) Global depth reconstruction: The global sharpness approximation as a function of the depth cannot be used directly to reconstruct the z-displacement as the Gaussian

functions are not invertible. In particular, it can be seen that two different depths can correspond to the same sharpness value (Fig. 2 - top). For this reason, the sinusoidal calibration trajectory can be divided in two sections to which a sum of one-sided Gaussian curves can be fitted (Fig 2 - Data sets 1 and 2 (top)). The corresponding inverted functions $z_i = f^{-1}(\phi)$ are depicted in Fig. 2 (bottom) (here $i = 1, 2$). The depth at which the data is divided corresponds to the depth z_s at which the approximated sharpness function is changing slope. The switching between the curves z_1 and z_2 (Fig. 2 - Fitted lines 1 and 2) is determined by a number of conditions. In particular, these switching conditions are defined by the sign of the rate of change of the focus measurement $\Delta\phi$, the sign of the update rate of the estimated depth Δz and the range at which the estimated depth belongs to (Data Set 1 or 2). Continuous switching can be ensured since a continuous z-trajectory is assumed and high alterations of the depth cannot be performed. Therefore, the switching conditions are given as follows

$$z_{est_{curr}} = \begin{cases} z_1, & \text{if } z_{est_{prev}} \leq z_s \text{ or } \Delta z < 0 \\ z_2, & \text{if } z_{est_{prev}} > z_s \text{ or } \Delta\phi_m < 0 \text{ and } \Delta z < 0 \end{cases} \tag{5}$$

$$\Delta\phi_m = \phi_{m_{curr}} - \phi_{m_{prev}}, \ \Delta z = z_{est_{curr}} - z_{est_{prev}} \tag{6}$$

where $z_{est_{curr}}$,$z_{est_{prev}}$ the current and previous estimated depth and $\phi_{m_{curr}}$,$\phi_{m_{prev}}$ the current and previous sharpness value respectively.

D. Depth estimation using local sharpness model approximation

The previous approach is sufficient for depth reconstruction when the sharpness function has a dominant peak that can describe the object displacement along the z-axis. Due to object transparency, the features below or above the focus plane contribute to the total sharpness measurement when using as input to the calculations the whole image frame. This is also indicated by the fact that the sharpness measure for the whole image is approximated by a sum of Gaussian terms. In the propose methodologies, it is assumed that each Gaussian curve corresponds to a dominant depth (or object feature). This implies that the contribution of the dominant features of the microstructure to the single sharpness measurement can correspond to different depths. In addition, during 3D rotation, different parts of the object become defocused, resulting to changes in the overall sharpness measurement. Hence, the overall sharpness measurement cannot identify whether defocusing occurs only locally at the object due to rotation or if it is uniform due to translation along the z-axis.

In order to get the corresponding depth of single features, smaller windows around single features of the object are obtained. The centres of the local windows are obtained by the projection of the registered 3D points of the CAD model on the image frame. The 3D orientation of the object is estimated as presented in [11]. Estimating the orientation and the 2D position of the object is necessary to obtain a window that contains only the feature of interest.

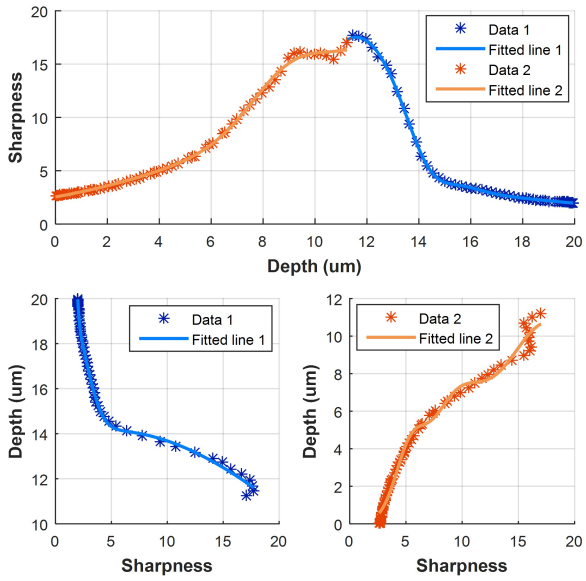

Fig. 2. Global sharpness model approximation from calibration (top). Inverse sharpness models (bottom)

1) Local sharpness model approximation and calibration: Similarly to the first method, a calibration routine is performed in order to establish the mapping between the sharpness and the depth of a single feature without interference by neighbouring features. It is shown that when the window is sufficiently small and contains single depth features, the sharpness model can be approximated with one Gaussian curve (Fig. 3). In this experiment, the features which are selected for the calibration are the corners of the teeth of the microgears. A square window of 15x15 pixels, which corresponds to 1.1 μ m^2, around the feature was created assuming that the interference from the neighbouring features is small (Fig. 3). It can be seen that there is a minor overlap between the corner and the edge features of the micro-gear teeth. However, the overlapping is sufficiently small so that the local sharpness can be approximated by a single Gaussian term. The focus measure which is used in this case is the Gray-Level Normalized Variance [13]. It is selected because a normalized measure is needed in order to have the same reference among all the fitted functions. In addition, the Gaussian Derivative is less sensitive to individual smaller peaks caused by the interfering object features.

2) Optimization based local depth reconstruction: After the local sharpness models are established, the estimated local depth $\hat{z}_i$ for the i_{th} feature is determined by minimizing the distances between the measured sharpness on the local window ϕ_{m_i} and the expected sharpness value from the derived Gaussian curve:

$$\hat{z}_i = \underset{z_i}{\operatorname{argmin}} \left\| \phi_{m_i} - \alpha_i \, e^{-\frac{(z_i - b_i)^2}{{c_i}^2}} \right\| \tag{7}$$

By using this optimization method, the individual curves don't need to be inverted. Moreover, by expanding the window more features are included and the right hand side

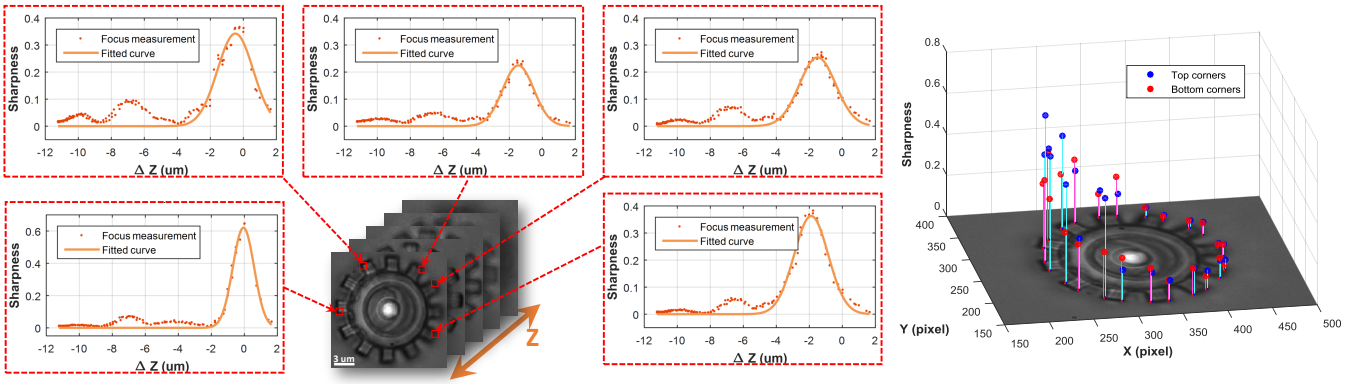

Fig. 3. Local sharpness model approximations using one Gaussian curve, the selected local windows and the corresponding object pose (left). The sharpness values as calculated for each of the selected features (right)

of the equation can be expressed as a sum of Gaussians, using the functions defined for the known features in the calibration procedure.

E. 2D position derivation and 3D model-based orientation estimation

As mentioned, the object position on the x-y plane and the 3D orientation are derived as presented in [11]. A brief overview of the methodology is described in this section.

1) Contour based 2D position derivation: The 2D position of the object is calculated using the extracted contour of the object. The centroid of the contour is used as the object 2D position for convex objects. Alternatively, for non-convex objects, the 2D position can be calculated with respect to another point within the extracted 2D contour.

2) Optimization based 2D to 3D registration: The microstructure 3D model is registered with its 2D projection on the image frame by manual selection of visible feature points. Since the initial point correspondences are known, the initial orientation R_0 and translation $\mathbf{t_0}$ of the object are derived by solving the following optimization problem:

$$R_0, \mathbf{t_0} = \underset{R,\mathbf{t}}{\operatorname{argmin}} \|\mathbf{P_{ci}} - K\,P\,(R\,\mathbf{P_{M_i}} + \mathbf{t})\| \qquad (8)$$
$$\text{s.t } R^T R = I_{3x3}$$

3) 3D model-based orientation estimation: The 3D orientation of the microstructure is estimated using online generated model-based templates. Model-based template matching is preferred because the initial point correspondences do not need to be tracked in every subsequent image frame. The templates are generated from a number of possible relative rotations by an angle step $\Delta\theta$ with respect to the current orientation of the microstructure. The generated templates are matched with the filled extracted object contour on the current image frame. Kalman filtering is also performed in addition to the method presented at [11].

III. EXPERIMENTAL RESULTS

The proposed methodologies are evaluated and compared to the ground truth data. The object motion is generated by translating the piezo stage along the z-axis and keeping the camera position fixed. The piezo stage trajectory is used as the ground truth data for the experimental validation.

A. Experimental Setup

The components of the hardware experimental setup include an Acoustic-Opto Deflector (AOD) Optical Tweezers (Elliot Scientific, UK) and an integrated 3D micromanipulation piezo stage which is used to simulate the object trajectories. The laser source is a 1070 nm fiber laser (Ytterbium Fiber Laser, IPG Photonics, USA) with 10 W maximum output power. A CCD camera (Basler AG, Germany) and an immersion oil lens optical microscope (Nikon Ti) with 100x magnification are used to capture the live feed during the experiment. The experimental set-up is shown in Fig. 4.

B. Microstructure fabrication

The structures used in the experiments are 3D-printed micro-gears. The height of the structure is 11.2 μm. The microstructures were printed using the Nanonsribe 3D-printer (Nanoscribe, Germany) and the IP-L photoresist. The printing power was set to 25 %. The structures were printed on a glass substrate, placed in deionized water and kept fixed on the glass substrate.

C. Depth estimation using global sharpness approximation model

1) Results and discussion: The depth of the 3D-printed microstructure in a different object trajectory is estimated by using the sharpness approximation model obtained from the calibration. In this case, a triangular trajectory was chosen but alternative motions could also be performed. The focus measurement of the new sequence is calculated and the resulting sharpness for each frame is shown in Fig. 5. The calculated sharpness values are used to estimate the depth using the inverse switching model as described in Section II-C. As shown in Fig.6, the estimated depth trajectory follows the ground truth stage trajectory. It is worth noting that the reconstruction error is similar for all three phases as the object trajectory is periodic.

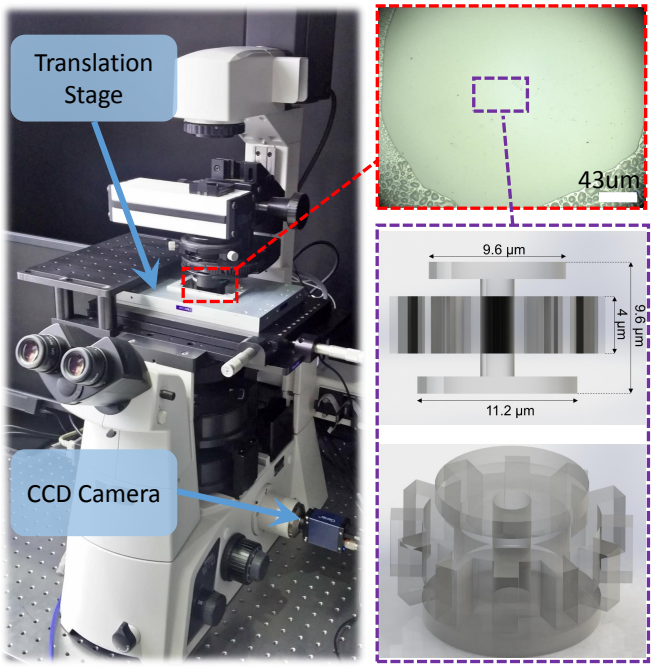

Fig. 4. The experimental set-up used in the experiments (left). The glass slide with the printed microstructures is placed above the microscope-camera system. The microstructures are printed in the middle of the glass slide (top right). The CAD model with the object dimensions are also displayed (bottom right)

D. Depth estimation using local sharpness approximation model

1) Results and discussion: To confirm that the proposed method can estimate the relative depth of the local features on the microstructure, the translational stage is moved by following a sine wave trajectory. For each local feature, a depth relative to the initial position is estimated by using the calibrated parameters of the 1D Gaussian model. As shown in Fig.7 (top), a set of local windows are cropped around the corner features in each image frame. The estimated relative depth follows the reference stage movement as long as the sharpness of the window changes. If the sharpness value does not change, which indicates the focus plane does not intersect with the microstructure, the proposed method assumes that the object is static. This could also be attributed to limitations of the used optimization method, which was the Quasi-Newton optimization algorithm. When the focus plane intersects the microstructure again (i.e. part of the object is in focus), the relative depth of the feature can be estimated, as shown in Fig.7.

It is worth noting that the calculated sharpness measures for each individual feature can also reveal information about the object orientation. In particular, Fig. 3 (right) depicts the selected features as projected on the image and their corresponding sharpness values. It is shown that the features on the left of the microgear are at lower depths compared to the features on the right. This can also be visually verified since the left features are more blurry than the ones on the right. This information can be fused with the results from

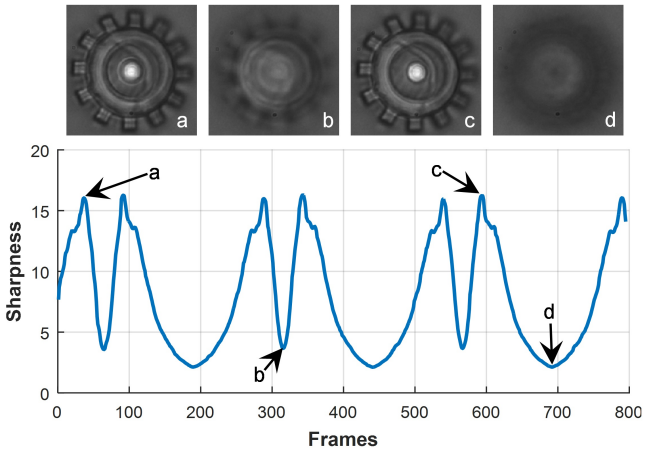

Fig. 5. Global sharpness measurement using the Gaussian Derivative and the corresponding images from the performed experiment

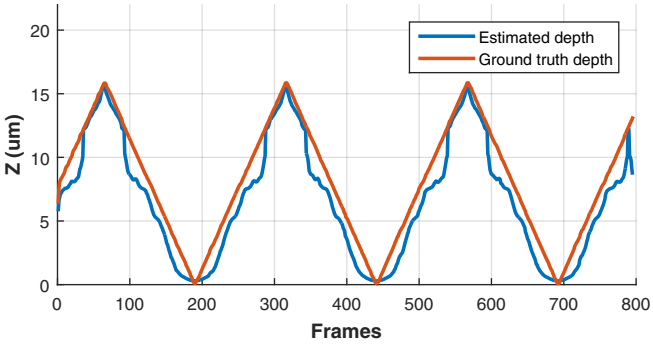

Fig. 6. The reconstructed trajectory (blue) and the ground truth trajectory of the piezo stage (orange)

the 3D estimation component of the framework to account for misalignments in the orientation or detect the out of focus parts of the microstructure. Therefore, the local depth estimation relates the local z-displacements of the object features to the estimated orientation.

E. 2D position and 3D model-based orientation estimation

1) Results and discussion: The corresponding orientation of the microstructure was estimated using the method described in Section II-E. As indicated by the local depth estimation, the 3D orientation of the object is not parallel to the focus plane but the object is slightly rotated around the X and Y axis of the global frame. The estimated orientation with respect to the global frame and the 2D position of the object are equal to:

$$R = R_z(-2.2081^\circ)\ R_y(2.1893^\circ)\ R_x(-0.6422^\circ) \quad (9)$$

$$t_x = 320.3724,\ \ t_y = 275.8510 \quad (10)$$

The corresponding matched template and the extracted contour are shown in Fig. 8.

IV. CONCLUSIONS

In this paper, two depth estimation methods are presented for transparent, laser-driven microrobots. A global and a local

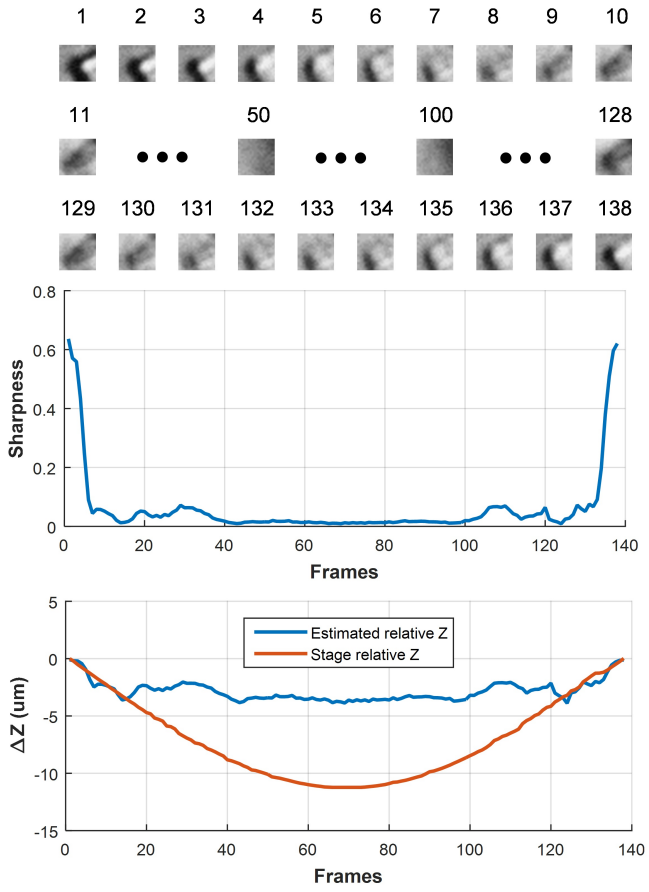

Fig. 7. An example of local feature depth estimation using the optimization based approach. A set of local windows located at a corner feature (top) are used to calculate the sharpness (middle) of the windows. The estimated and stage relative depth (bottom) shows that the proposed method can be used to estimate local depth information as long as the changes of sharpness can be detected.

method of depth estimation of the transparent microstructure are described. These methods aim to derive an estimation for the sharpness model with respect to the overall z-translation of the object and depth changes of individual features of the microstructure. The calibration methods presented for both methodologies are object and lighting condition specific and need to be performed prior to the depth estimation routine. The combination of these two methods aims at distinguishing the cause of image sharpness variations. These variations can be attributed either to object translation along the microscope z-axis detected by the global approach or to 3D rotation detected by the local approach. Hence, the local method is capable of identifying which parts of the microstructure are in and out of focus in combination with the 3D estimated orientation. The global sharpness can be approximated by a sum of Gaussian curves for interfering features at different depths due to transparency. The local sharpness model is approximated by one Gaussian curve for distinct object features. The local method can reveal information about the object orientation and can be potentially combined with the model-based approach to estimate 3D orientation. By

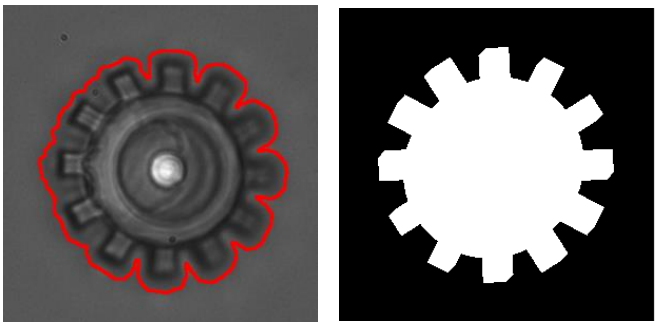

Fig. 8. The extracted contour and the corresponding matched templated generated from the orientation estimation component

combining the two methods, the framework can account either for errors that occur from the template matching or the local depth estimation.

REFERENCES

[1] Y. Shen and T. Fukuda, "State of the art: micro-nanorobotic manipulation in single cell analysis," *Robotics and Biomimetics*, vol. 1, pp. 1–13, 2014.

[2] K. Berk Yesin and B. Nelson, "Robust CAD model-based visual tracking for 3D microassembly using image space potentials," *IEEE Int. Conf. on Robotics and Automation, ICRA'04*, no. April, pp. 1868–1873, 2004.

[3] B. Tamadazte, E. Marchand, S. Dembele, and N. Le Fort-Piat, "CAD Model-based Tracking and 3D Visual-based Control for MEMS Microassembly," *The International Journal of Robotics Research*, vol. 29, no. 11, pp. 1416–1434, 2010.

[4] Z. Ni, A. Bolopion, J. Agnus, R. Benosman, and S. Régnier, "Asynchronous event-based visual shape tracking for stable haptic feedback in microrobotics," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1081–1089, 2012.

[5] C. Bergeles, B. E. Kratochvil, and B. J. Nelson, "Visually servoing magnetic intraocular microdevices," *IEEE Transactions on Robotics*, vol. 28, no. 4, pp. 798–809, 2012.

[6] A. Barbot, D. Decanini, and G. Hwang, "On-chip Microfluidic Multimodal Swimmer toward 3D Navigation," *Scientific Reports*, vol. 6, no. January, p. 19041, 2016.

[7] B. E. Kratochvil, L. Dong, and B. J. Nelson, "Real time Rigid-body Visual tracking in a Scanning Electron Microscope," *Int. J. Robotics Research*, vol. 28, no. 4, pp. 488–511, 2009.

[8] L. Cui, E. Marchand, S. Haliyo, and S. Regnier, "Three-Dimensional visual tracking and pose estimation in Scanning Electron Microscopes," *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5210–5215, 2016.

[9] N. Marturi, B. Tamadazte, S. Dembl, 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, Aug 2016.

[10] S. Fatikow, V. Eichhorn, 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.

[11] M. Grammatikopoulou and G.-Z. Yang, "Gaze Contingent Control for Optical Micromanipulation," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5989–5995, 2017.

[12] J.-M. Geusebroek, F. Cornelissen, A. W. Smeulders, and H. Geerts, "Robust autofocusing in microscopy," *Cytometry Part A*, vol. 39, no. 1, pp. 1–9, 2000.

[13] J. L. Pech-Pacheco, G. Cristóbal, J. Chamorro-Martinez, and J. Fernández-Valdivia, "Diatom autofocusing in brightfield microscopy: a comparative study," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 3. IEEE, 2000, pp. 314–317.