Scene-Adaptive Fusion of Visual and Motion Tracking for Vision-Guided Micromanipulation in Plant Cells

Scene-Adaptive Fusion of Visual and Motion Tracking for Vision-Guided Micromanipulation in Plant Cells

Ishara Paranawithana, *Member, IEEE*, Liangjing Yang, *Member, IEEE*, Zhong Chen, Kamal Youcef-Toumi, *Member, IEEE,* U-Xuan Tan, *Member, IEEE*

***Abstract*— This work proposes a fusion mechanism that overcomes the traditional limitations in vision-guided micromanipulation in plant cells. Despite the recent advancement in vision-guided micromanipulation, only a handful of research addressed the intrinsic issues related to micromanipulation in plant cells. Unlike single cell manipulation, the structural complexity of plant cells makes visual tracking extremely challenging. There is therefore a need to complement the visual tracking approach with trajectory data from the manipulator. Fusion of the two sources of data is done by combining the projected trajectory data to the image domain and template tracking data using a score-based weighted averaging approach. Similarity score reflecting the confidence of a particular localization result is used as the basis of the weighted average. As the projected trajectory data of the manipulator is not at all affected by the visual disturbances such as regional occlusion, fusing estimations from two sources leads to improved tracking performance. Experimental results suggest that fusion-based tracking mechanism maintains a mean error of 2.15 pixels whereas template tracking and projected trajectory data has a mean error of 2.49 and 2.61 pixels, respectively. Path B of the square trajectory demonstrated a significant improvement with a mean error of 1.11 pixels with 50% of the tracking ROI occluded by plant specimen. Under these conditions, both template tracking and projected trajectory data show similar performances with a mean error of 2.59 and 2.58 pixels, respectively. By addressing the limitations and unmet needs in the application of plant cell bio-manipulation, we hope to bridge the gap in the development of automatic vision-guided micromanipulation in plant cells.**

I. Introduction

Advancement of cell manipulation techniques has been greatly accelerated by the development of automated micromanipulation systems which equip the users with unprecedented performance and ease of operation [1]. While existing robotic micromanipulation systems are readily capable of single cell manipulation, the technology is not fully leveraged for micromanipulation application in plant cells. Therefore, most plant cell manipulation procedures still rely on manual control, thus the success rates highly depend on the operator's experience. Accuracy and repeatability of the study is hardly guaranteed under demanding working conditions. Hence, there remains a gap in extending current state-of-the-art robotic micromanipulation development to work on plant cells.

The unique structural properties of plant cells make micromanipulation task like microinjection a lot more challenging compared to the manipulation of single animal cell [2, 3]. Figure 1 depicts the images of the plant cell specimen and the tool in a common scene. As the tool interacts with the multiple layered array of plant cells, scene cluttering, and regional occlusion of tool tip make conventional visual tracking approach challenging, if not impossible. Difficulty in predicting such visual uncertainties further impose limitations on how cell manipulation is carried out. Development in vision-guided micromanipulation in plant cells is rarely explored because of these technical challenges. Nevertheless, bio-manipulation of plant cells has potentially important applications including the introduction of genetic material into plant cells to study plant transformation [4, 5].

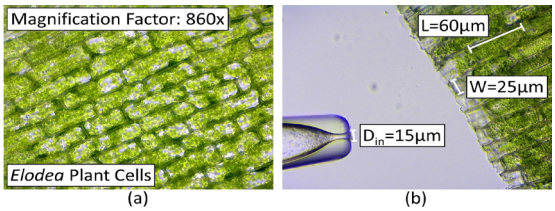

Figure 1. Microscopic image of (a) plant cell array (b) with a microholder.

To address the unmet needs, we proposed a fusion mechanism that combines trajectory and visual tracking data to overcome tracking failures in the complicated scene associated with multiple layers of plant cell arrays. The proposed fusion mechanism adjusts the influence of the trajectory data from the robotic task space based on a visual tracking score. The contribution is therefore a cooperative fusion method, which automatically adjusts the weightage of the trajectory data in event of poor confidence in visual tracking inferred from a low similarity score. Doing so overcomes the limitations of vision-guided micromanipulation for plant cells.

In the next section, we will begin with a literature review to survey relevant state-of-the-art development and limitations of existing practices to identify the gaps. Details of our proposed method will be discussed in Section III followed by a description of the experimental setup for evaluation in Section IV. Finally, results and discussion will be presented in Section V before concluding the paper by stating the significance and potential development in the future.

*Research supported by SUTD-MIT International Design Centre (IDC).

Ishara Paranawithana and U-Xuan Tan are with the Pillar of Engineering Product Development, Singapore University of Technology and Design, Singapore (phone: +65-6303-6600; e-mail: ishara_paranawithana@sutd.edu.sg, uxuan_tan@sutd.edu.sg).

Liangjing Yang is with the Zhejiang University/University of Illinois at Urbana-Champaign Institute (ZJU-UIUC Institute), China (e-mail: liangjingyang@intl.zju.edu.cn).

Zhong Chen is with the Natural Sciences and Science Education Academic Group, National Institute of Education, Nanyang Technological University, Singapore (e-mail: zhong.chen@nie.edu.sg).

Kamal Youcef-Toumi is with the Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (e-mail: youcef@mit.edu).

II. Literature Review

Existing vision-guided systems, including our previous work, for cell manipulation applications will be reviewed in this section. Emphasis of the discussion will be on the gap in existing systems to extend for plant cell manipulation. A relevant development [6] associated with plant cell manipulation applications will also be reviewed in this section.

Development in vision-guided micromanipulation has contributed greatly towards the advancement of automated cell manipulation [7-9]. Vision-guided approach for micromanipulation involves integrating the vision and the control of the microscope and micromanipulator, respectively. This integration procedure between the vision sensing and motion actuation is achieved by system calibration [10-16], which essentially establishes the mapping function between the coordinates of the motion actuation task space and the image sensing domain [17, 18].

In our previous work, we developed a portable and easy-to-deploy framework [19] that leverage uncalibrated [20] vision-guided micromanipulation with self-initialization [21] to achieve a self-contained manipulation system. The advantage of the self-contained system is that no tedious calibration procedure is required before the study. This offers great ease of integration and operation to the user which is essential for most onsite cell manipulation procedures.

Despite extensive research and development in vision-guided micromanipulation, including our previous work [20, 21], automated system for plant cell manipulation has not been substantially explored. There remains a gap in dealing with scene uncertainty and visual disturbances even for single cell procedures. Hence, we designed a method that identifies the appropriate tracking mode to adopt using prior knowledge of the scene. This is done based on the detected geometry and position of the cell specimen. The result is a self-reinitialization and -recovery method for uninterrupted visual tracking under tool-specimen interaction [22]. The recognition and localization steps of the cell are realized using previously developed cell detection method [23]. However, making use of known conditions to switch tracking mode, has limited the contribution towards automatic vision-guided micromanipulation. This calls for a more generalized workflow for fusion of multiple trackers.

Some interesting studies carried out on plant cell manipulation can be found in the existing literature [6, 16]. Han et al. developed a vision-based technique for plant cell microinjection using an autofocusing algorithm [6]. The main contribution of this work is to align the microneedle tip and target cell on the same imaging plane. The proposed injection strategy is not executed with a vision feedback driven track-servo mechanism. Hence, problems associated with visual tracking of the microneedle are not extensively discussed. In [16], the authors adopt an explicit calibration approach to establish a relationship between world coordinate system and image coordinate system before using modified 2D-to-2D SSD feature tracking method for small cell injection. Despite the development and interest shown in plant cell related applications, hardly any research addressed the challenges unique to visual track-servo in plant cell manipulation using an uncalibrated setup. Therefore, this work is motivated by the existing gaps and limitations identified in the current systems, including our previous work and the latent need to incorporate robust vision-guided control for plant cell manipulation.

III. Fusion of Trajectory Data and Visual Tracking

A. Conceptual Overview

The vision-guided manipulation for plant cells consists of 1) Visual track-servo framework, 2) Homography-based mapping of task space to image domain and 3) Score-based weighted averaging to combine the two sources of data. In essence, the concept is to combine tool tip trajectory executed by the micromanipulator with the visually tracked path of the tip in the microscope image.

A flowchart illustrating the workflow of the vision-guided manipulation is depicted in Figure 2. The fusion mechanism of the two data sources, namely, manipulator trajectory and tracked path of the tool in the image are combined by procedures demarcated by the dotted blue cluster. A projective homography is estimated using initial corresponding data between the trajectory and image coordinates. The estimated homography is then used to map the trajectory from the task space to the image domain. The projected trajectory is then combined with the tip path in image coordinates using a score-based weighted averaging approach.

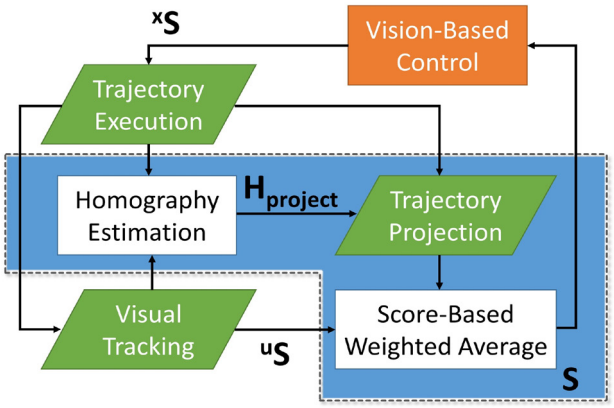

Figure 2. Workflow of the fusion mechanism; $^{x}\mathbf{S}$ and $^{u}\mathbf{S}$ denote step input in robot task space and image coordinates in camera frame, respectively.

1) *Visual Tracking in Image Domain*

Vision-guided micromanipulation is carried out using a track-servo framework and the DFTS workflow algorithm designed previously in [19, 21]. As introduced in Section II, the framework uses a practical approach to achieve 3D manipulation through 2D microscopy and image-based technique to automate initialization for the tracking of the tool tip. The 3D manipulation is achieved by template-based tracking and score-based depth compensation method detailed in Section III-B. The latter uses similarity score of the template to maintain the tool tip in the focal plane during the manipulation. This vision-based control is a self-reliant approach that is based purely on images of the scene. Through visual tracking the path $^{u}\mathbf{S}$ of the tip manipulation can be obtained.

As the feedback is solely based on vision, any form of visual occlusion or scene disturbance will have detrimental

influence on the vision-based control. There is a need to incorporate trajectory data from the manipulator to enhance tracking of the tool tip. However, it is impractical, if not impossible, to localize the tool tip based on the forward kinematics of the multi-body systems as the trajectory information from the manipulator is in the order of micrometers. Physical disturbances like micropipette deflection and vibration are easily amplified in the microscale level. Therefore, we explore the option of fusing trajectory information with our previously developed track-servo method.

2) *Mapping Trajectory to Image Domain*

To address the limitations in pure vision-based tracking during micromanipulation application in plant cells, we incorporate trajectory data from the manipulator by mapping them to the image domain. Instead of performing a robot-camera calibration and the forward kinematics of the manipulator system, we use a projective homography to map the trajectory to the image domain. Doing so works around the influence of large uncertainty associated with performing forward kinematics of the entire multibody system of the manipulator.

The homography matrix projects a set of points to a particular plane. Since the micromanipulator is mounted rigidly with respect to the microscope camera, there exists a homography that project the planar trajectory to the image plane [24]. Our goal is hence to estimate a homography that best describes the relationship between the two planes using known point correspondences. This estimated homography can be used to map subsequent trajectory data to the same image plane during the process of vision-guided manipulation as explained in Section III-C. Projecting these data points to a common image plane where visual tracking is based on, allows comparison and data fusion for better estimates. However, physical disturbance from the environment can compromise the initial assumption about the existence of single homography, hence there is a need for fusion with visual information.

3) *Fusion of Tracking Data*

Finally, the template matching score discussed in the visual tracking section is used to infer the certainty of the template tracking data. The score acts as a weight that adjusts the linear combination of the two sources of estimation. When there is partial occlusion or visual disturbance in the vicinity of the tool tip, the score will fall leading to diminishing influence on the final estimate. The trajectory data on the other hand will receive higher weightage. This mechanism is therefore scene sensitive and naturally shift the weighting of estimation sources based on the quality of the visual tracking outcome. The result is an adaptively adjusted linear combination of the tracking sources.

B. Visual Track-Servo Framework

The visual track-servo framework developed in our previous work will be presented concisely to ensure a self-contained discussion. Figure 3 gives a general overview of the framework configuration and the workflow. Interested readers may refer to our previous work on self-initializing and unified track-servo mechanism in the DFTS framework [21].

1) *Template Match*

A region of interest surrounding the focused tool tip is registered as a template for visual tracking during the manipulation. For a given image frame during the operation, the template is compared against patches of neighborhood pixels. For a specific patch with center pixel (u,v), a score w is computed from the normalized cross-correlation of the template g with the associated patch f. The cross-correlation $w_{cc}(u,v)$ at image coordinates (u,v) of a template patch $g(p,q)$ and the image $f(p,q)$ is expressed as

$$w_{cc}(u,v)=\sum_{p=0}^{P}\sum_{q=0}^{Q} g(p,q)f(p+u,q+v) \qquad (1)$$

for a P x Q patch and U x V image. To further account for the sensitivity of cross-correlation index due to intensity variation, (1) is re-expressed as the normalized cross-correlation coefficient

$$w(u,v)=\left(\sum_{p=0}^{P}\sum_{q=0}^{Q}\langle G\rangle\langle F\rangle\right)\Big/\left[\left(\sum_{p=0}^{P}\sum_{q=0}^{Q}\langle G\rangle^2\right)\left(\sum_{p=0}^{P}\sum_{q=0}^{Q}\langle F\rangle^2\right)\right]^{0.5}, \qquad (2)$$

where $\langle G\rangle=(g(p,q)-\bar{g})$ and $\langle F\rangle=f(p+u,q+v)-\bar{f}(u,v)$. Notation $\bar{g}$ and $\bar{f}$ represent the mean intensity value in the template and the overlapping region, respectively.

The tracked location of the tool tip is at the pixel coordinates associated with the maximum score. The template-based matching method is iteratively carried out over the sequence of images to localize the tool tip position and tracks its motion. However, this step only localizes the tip position in 2D image coordinates. As we do not require the micromanipulator to be planarly aligned with the image plane, there is a need to manipulate the motion in 3D space such that the trajectory is maintained within the focal plane. This requires motion in the z-axis which can be carried out using the method to be described in the next section.

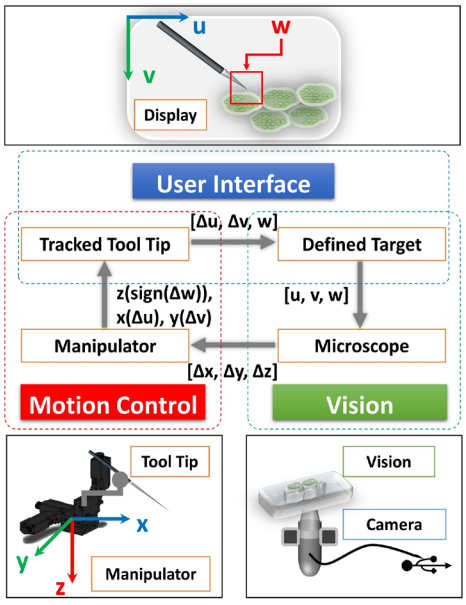

Figure 3. General overview of the visual track-servo framework.

2) *Similarity-Score-Based Depth Compensation*

Our vision-guided control uses the similarity score to compensate the depth of the tip keeping it in focus. As the x- and y-axis manipulates the tip to its target, z-axis of the manipulator compensates for the deviation from focal plane concurrently. The aim is to maximize the similarity-score while actuating the manipulator joints. Doing so ensures that the tool tip is manipulated in the focal plane because the similarity between the template and a detected patch is maximum when the tip is in focus. Similar assumption regarding similarity score-maximizing has been used in existing systems [7, 8] to focus tool tip. These systems, however, do so before manipulating the tool tip i.e. the tip focusing is done as a separate step from visual servo. In a gradient ascending fashion, the z-position of the manipulator is adjusted online during manipulation such that the score is maintained at maximum. As shown in the pseudocode in Table I, the tool tip is adjusted in the z-direction by Δz such that change in Δw reduces to a specified tolerance *tol*.

TABLE I. PSEUDO- CODE FOR ADAPTIVE COMPENSATION

Gradient Ascending Updating Algorithm

```
1. initialize
2.     Δz:= Δw:=tol;
3.     w:= [a]compute_score(g,f);
4. loop while Δw ≥ tol
5.     Δw:= w- compute_score(g,f);
6.     Δz:= Δz* [b]sgn(Δw);
7.     w:= compute_score(g,f);
8. end loop
```

a. **compute_score**() denotes a function that implements Equation 2 b. **sgn**() extracts the sign

C. Homography-Based Projection of Trajectory Data

Our proposed method estimates a projective homography $\mathbf{H}_{\text{project}}$ to map the trajectory of the micromanipulator in the task space to the image domain. This is based on the assumption that there exists a projective transformation which projects the planar trajectory data ${}^{x}\mathbf{S}_{\text{robot}}$ from the task plane $\prod_x$ to the image plane $\prod_u$ such that the projected data ${}^{u}\mathbf{S}_{\text{robot}}$ align with the visual tracked trajectory ${}^{u}\mathbf{S}_{\text{camera}}$ as illustrated in Figure 4.

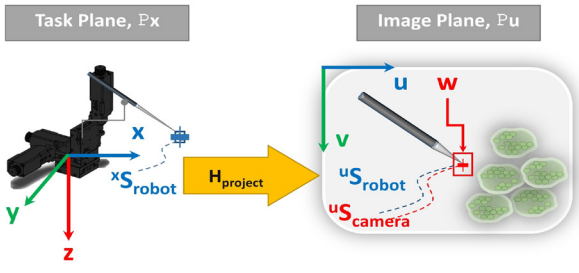

Figure 4. Mapping of robot trajectory in task plane to image coordinates in camera frame.

The objective is therefore to estimate the homography. This can be done by maximizing the inliers using the RANSAC algorithm for a projective transformation relationship.

For N randomly selected samples, ${}^{u}\mathbf{S}_{\text{camera}}$ and ${}^{x}\mathbf{S}_{\text{robot}}$ are removed if the algebraic distance

$$\text{norm}\left({}^{u}\mathbf{s}_{camera},\ \prod\left({}^{x}\mathbf{s}_{robot}:\mathbf{H}_{\text{proj}}\right)\right)=\sqrt{(i-k\cdot u)^2-(j-k\cdot v)^2} \quad (3)$$

is greater than the threshold $\boldsymbol{\tau}$. As the visual servo loop terminates when the error is less than $\pm$ 3 pixels, $\boldsymbol{\tau}$ is assigned to be 6 pixels in order to be comparable to the control precision. Notation Π denotes the projection of ${}^{x}\mathbf{S}_{\text{robot}}$ via an estimated 3 by 3 homography matrix $\mathbf{H}_{\text{proj}}$ [24, 25]. Variables i, j and k are the unnormalized image projection coordinates of the tool tip such that

$$\left(i/k \quad j/k\right)^{\mathrm{T}}=\mathbf{H}_{\text{proj}}\,{}^{x}\mathbf{s}_{robot}. \quad (4)$$

The value of N is updated dynamically based on the proportion of identified inliers, according to the original RANSAC algorithm [26, 27]. By iteratively updating the estimated projective matrix $\mathbf{H}_{\text{proj}}$ until the distance metric $\mathbf{M}$ of n points,

$$\mathbf{M}=\sum^{n}\text{threshold}\left(\text{norm}\left({}^{u}\mathbf{s}_{camera\ m},\prod\left({}^{x}\mathbf{s}_{robot\ m}:\mathbf{H}_{\text{proj}}\right)\right),\tau\right) \quad (5)$$

converges, an optimal homography producing the maximum number of inliers is obtained. Finally, the estimated $\mathbf{H}_{\text{project}}$ is used to transform ${}^{x}\mathbf{S}_{\text{robot}}$ to ${}^{u}\mathbf{S}_{\text{robot}}$ for subsequent cooperative form of fusion as will be discussed in the next section.

D. Template Match Score-Based Data Fusion

The position of the tool tip is estimated using the weighted average of the visual tracking data in the image domain with the projected trajectory data. As the similarity score w measures how much, a neighbourhood of pixels matches a template of the tool tip, it can therefore be used to indicate the confidence of the visual tracking data point.

The score of the template match is a good indication of how well a specific region of interest is likely to be the tracked object. However, unlike single cell microinjection, the background is not homogenous but cluttered and occluded by cells as the tool approaches and interacts with the specimen as seen in Figure 1. Our previous method of motion cue and block template-based tracking could not work well in such situations. Despite the advantage of a unified visual track-servo framework [20, 21] that can manipulate the tool in 3D, the introduction of multiple layered arrays of plant cells undermines the accuracy of visual tracking. As it is non-trivial to detect the structural geometry of the specimen to perform, previous method of self-initialization and recovery tracking [22] is not suitable. Pure vision-based tracking has limitation in tool tracking with plant cell specimen in the scene. There is a need to fuse trajectory information from the robot coordinate system which is mapped to the image domain using the previously estimated projective homography.

For a pair of corresponding estimates $({}^{u}\mathbf{S}_{\text{robot}}, {}^{u}\mathbf{S}_{\text{camera}})$, the normalized weighted average of the estimates is expressed as

$${}^{u}\mathbf{s}=(1-w)\left(\mathbf{H}_{project}\,{}^{x}\mathbf{s}_{robot}\right)+w\,{}^{u}\mathbf{s}_{camera}, \quad (6)$$

where w is the similarity score derived in (2). The final estimate ${}^{u}\mathbf{S}$ is subsequently used as the feedback for the track-servo mechanism in our plant cell micromanipulation application. This fusion of complementary data sources combines trajectory and visual information. The former and

latter are robust individually against visual disturbances and mechanical uncertainties, respectively. Therefore, it is expected to minimize the influence from both kinds of uncertainties by combining them. Probabilistic fusion methods like Kalman Filter requires historical temporal data to make estimates which makes them slow in responding to unforeseen visual disturbances. The simplistic fusion method we proposed neither requires additional information nor makes assumptions based on previous data, thus provides robust visual tracking for micromanipulation.

IV. Experimental Setup

Experiments were carried out using the portable micromanipulation platform developed in our previous work [19] as shown in Figure 5. This setup essentially demonstrates the feasibility of our long-term research goal of bringing vision-guided micromanipulation outside laboratory environment. Furthermore, it facilitates our proposed method for plant cell manipulation under more challenging working conditions.

Microscopic imaging of the plant specimen is acquired at 30 frames-per-second using a portable digital USB microscope (AM4515T8 Dino-Lite Edge Series, AnMo Corp., Taiwan) which has a 1.3 Megapixel CMOS color image sensor. It comes with a continuous magnification range of 700x to 900x and built-in LED illumination source. For the experiments, a fixed magnification factor of 860x is used to acquire clear images of both plant cells and micro-holder. 3-axis motorized Cartesian micro-stages (8MT173; Standa Ltd., Lithuania) are used for manipulation. Each micro-stage has a resolution of 1.25µm/step and a workspace of 20mm. A dedicated multi-axis controller (8SMC4; Standa Ltd., Lithuania) which interfaces to workstation computer through USB is used to control the actuation of micro-stages. The vision-guided micromanipulation is carried out at a speed of 6.25µm/s. An aquatic plant named *Elodea* which contains two layers of cell arrays is used as the plant cell specimen for the feasibility test. Each cell of the specimen is approximately 20-30µm wide and 60-130µm long.

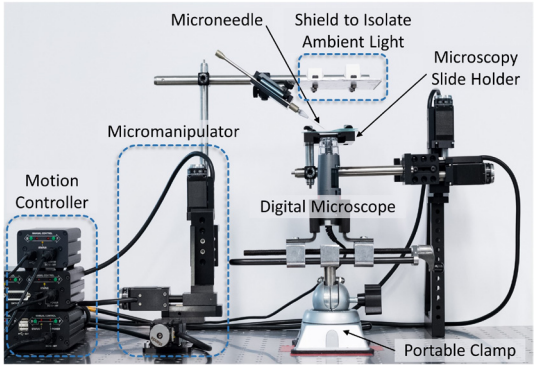

Figure 5. Experimental system and setup.

V. Results and Discussion

The experiments were designed to evaluate performance of fusion-based tracking mechanism on qualitative observations and quantitative evaluation. The objective of the experiments is to demonstrate the feasibility of the proposed method under the presence of plant cells with several practical scenarios including different levels of partial occlusion and non-occlusion of the ROI. The structural complexity of the plant cell array complicates the scene by creating partial occlusion of the tracking ROI. Therefore, this makes visual tracking more challenging where most of the conventional visual tracking methods would fail. The fusion-based tracking mechanism is compared against score-based template tracking and projected trajectory of the manipulator in the next two subsections. Details about video demonstration of the tracking workflow is also presented in the latter part of the discussion.

A. Qualitative Observations

Discrepancies in the localization results of different tracking methods can be identified through visual inspection of the tracked tool tip. Therefore, a detailed discussion is provided highlighting the reasons for differences in visual observations.

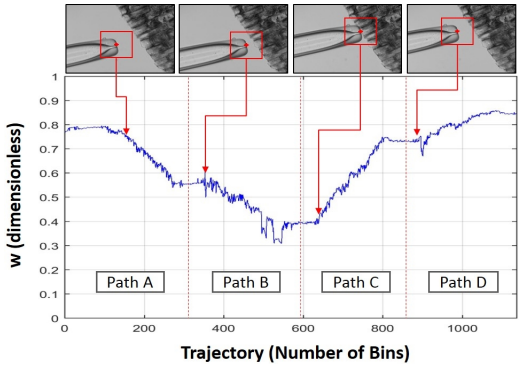

Figure 6. Similarity score profile of the base template along square trajectory; score drops from 0.6 to 0.3 along linear path B.

A square trajectory consisting of 4 linear segments is executed with vision guided micromanipulation under the presence of plant cells. Figure 6 shows the similarity score profile of the base template depicting the variations as the tool tip moves along the square path. It can be observed that similarity score falls below 0.6 at the start of path B and further drops closer to 0.3 as the tool tip reaches the point of interaction with the specimen. This clearly illustrates the adverse effect on score-based template tracking as a result of partial occlusion of the tracking ROI. However, this approach is essential for our unified track-servo method [20] to perform 3D servoing based only on 2D microscopic imaging information. Hence, the goal of the fusion-based method is not to completely outperform traditional template tracking but to leverage our previous work while overcoming the limitations of template tracking.

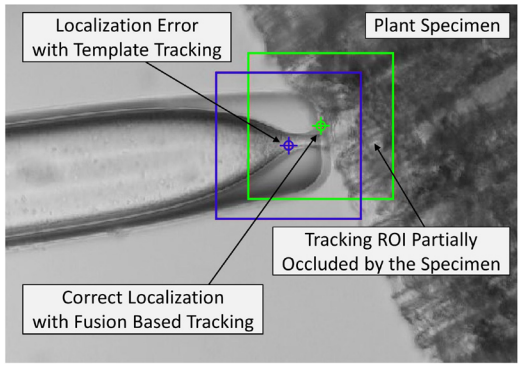

Figure 7. Superimposed microscope images during visual tracking.

The problem of tool tip localization error which typically leads to tracking failure is demonstrated in Figure 7. It can be observed that score-based template tracking resulted in an incorrect localization of the tracked tool tip in the vicinity of the plant cells. This is due to relatively low similarity score inferred as low as 0.3, when tracking ROI is partially occluded by the specimen. Based on this observation, it suggests that similarity score is a good confidence measure to identify the failing tracker. Since micromanipulator trajectory is not affected by the scene uncertainties, it can be exploited to minimize the localization error. Fusion-based tracking mechanism continues to place a favourable weight on trajectory information under occlusion condition and make an accurate localization estimate as seen in Figure 7.

The tracking performance of the tool tip under the presence of plant specimen is illustrated in Figure 8. Both template tracking and projected manipulator trajectory failed to demonstrate accurate localization in path B. Inaccuracy in template tracking in path B is expected because it is the closest path (out of the four segments) to the plant cell structure which introduces substantial amount of partial occlusion to the tracking ROI. It can also be seen that template tracking completely failed at the end of path B where tool comes to an interaction point with the specimen. Moreover, the innate uncertainty of the micromanipulator contributes to the tracking errors in the projected trajectory. However, fusion-based approach demonstrates its ability to overcome such limitations in template tracking and leverage trajectory information to estimate an accurate localization. This observation suggests that fusion-based method can effectively reduce the risk of tracking failures and improve the robustness of tracking under challenging scenes. Figure 8 further shows that tracking error of our proposed method is confined within that of original two sources of tracking data. Tracking performance for paths A, C and D, in which tracking ROI is not as badly affected as path B, will continue to perform better than the poorest performing tracker.

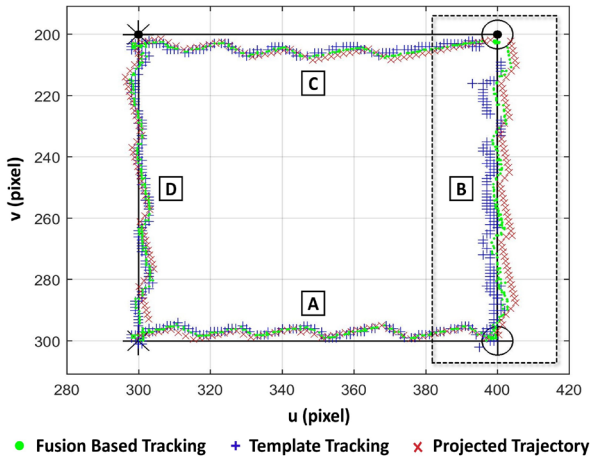

Figure 8. Tracking performance of the square trajectory; Error of path B is confined within that of orignal two sources of tracking data.

B. *Quantitative Evaluation*

In this subsection, a quantitative evaluation is presented by analyzing the tracking error along four straight line paths which forms a pre-specified square trajectory in the locality of plant cells as shown in Figure 9. The black square (in solid lines) represents the desired trajectory while color-coded data points denote the visually tracked tool tip in three tracking modes namely, template tracking, projected manipulator trajectory and fusion-based tracking. The four linear paths consist of 1137 frames in total are labelled as Path A (308 frames), Path B (273 frames), Path C (265 frames) and Path D (291 frames). The perpendicular pixel distance between tracked data points and corresponding straight-line path (geometric error) is used to quantify the tracking error for each tracking method.

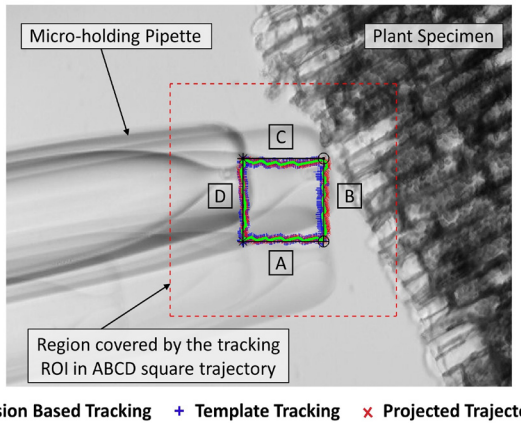

Figure 9. Overlay of images in square tracjectory in three tracking methods under the presence of plant specimen.

Mean geometric error for all four segments of the trajectory with three tracking methods discussed earlier is reported in Figure 10. In general, fusion-based tracking performed better than both template tracking and projected manipulator trajectory by maintaining the lowest mean error of 2.15 pixels (=0.7525µm). Template tracking and projected manipulator trajectory has a mean error of 2.49 pixels (=0.8715µm) and 2.61 pixels (=0.9135µm) respectively. Each pixel represents 0.35µm by 0.35µm square in manipulator task space. The reported mean error is reasonable as the micromanipulator has a resolution of 1.25µm (=3.57 pixels) and vision-based control was executed with a tolerance of 3 pixels. It should be noted that mean error of path B using template tracking and projected manipulator trajectory have similar performances with 2.59 pixels and 2.58 pixels, respectively. However, fusion-based approach succeeded in significantly reducing the mean error of path B to 1.11 pixels.

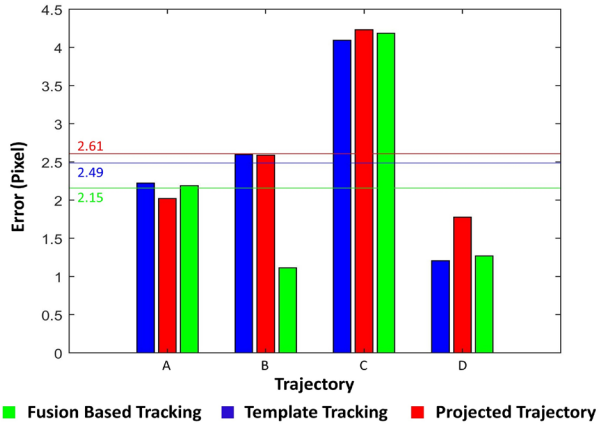

Figure 10. Geometric error of four linear segments; The weighted mean error of each tracking method is marked by color-coded straight line.

C. *Demonstration of the Tracking Workflow*

A video demonstration of the proposed fusion-based tracking workflow for plant cell manipulation application is presented in the attached media file. Firstly, the problem of visual tracking failure as a result of partial occlusion of the tracking ROI by cell specimen is illustrated. Secondly, the tracking performance of fusion mechanism is demonstrated by executing a desired square trajectory in the vicinity of plant cells.

VI. Conclusion and Future Work

In this paper, we proposed a fusion mechanism to combine two sources tracking data namely, template tracking and projected manipulator trajectory, using score-based weighted averaging. The main contribution of this work is to apply vision-guided micromanipulation in more challenging and fairly less explored area of plant cell applications by exploiting previously developed uncalibrated, self-initializing portable micromanipulation setup. Fusion-based tracking approach demonstrated the capability of overcoming limitations in traditional vision-based and calibration-based approaches and bringing the best of both worlds. By proposing scene adaptive fusion of vision and motion tracking, we hope to bridge the gap in extending vision-guided manipulation for plant cell studies. In general, it would allow cell biology research to be performed onsite rather than constraining the study to specific laboratory conditions.

Our future work will focus on carrying out more application specific and realistic experiments in different types of plant cells with the presence of full occlusion. This includes but not limited to applications relating to transfer of genetic materials to/from plant cells. By doing so, we envision to realize our long-term research goal of extending vision-guided micromanipulation for plant cells beyond conventional laboratory environment thus positively impacting the way plant cell studies are being performed.

References

[1] J. P. Desai, A. Pillarisetti, and A. D. Brooks, "Engineering approaches to biomanipulation," *Annu Rev Biomed Eng,* vol. 9, pp. 35-53, 2007.

[2] P. Barbier de Reuille, I. Bohn-Courseau, C. Godin, and J. Traas, "A protocol to analyse cellular dynamics during plant development," *The Plant Journal,* vol. 44, pp. 1045-1053, 2005.

[3] T. Kunkel. (2015, 1/24/18). *Microinjection into plant cells of etiolated seedlings*. Available: http://www.labonline.com.au/content/life-scientist/article/microinjection-into-plant-cells-of-etiolated-seedlings-981244058

[4] G. Neuhaus and G. Spangenberg, "Plant transformation by microinjection techniques," *Physiologia Plantarum,* vol. 79, pp. 213-217, 1990.

[5] M. Y. A. Masani, G. A. Noll, G. K. A. Parveez, R. Sambanthamurthi, and D. Prüfer, "Efficient transformation of oil palm protoplasts by PEG-mediated transfection and DNA microinjection," *PLoS One,* vol. 9, p. e96831, 2014.

[6] M. Han, Y. Zhang, C. Y. Shee, T. F. Chia, and W. T. Ang, "Plant cell injection based on autofocusing algorithm," in *Robotics, Automation and Mechatronics, 2008 IEEE Conference on*, 2008, pp. 439-443.

[7] Y. Sun and B. J. Nelson, "Biological cell injection using an autonomous microrobotic system," *The International Journal of Robotics Research,* vol. 21, pp. 861-868, 2002.

[8] Y. Sun and B. J. Nelson, "Microrobotic cell injection," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, 2001, pp. 620-625.

[9] W. Wang, X. Liu, D. Gelinas, B. Ciruna, and Y. Sun, "A fully automated robotic system for microinjection of zebrafish embryos," *PloS one,* vol. 2, p. e862, 2007.

[10] L. S. Mattos and D. G. Caldwell, "A fast and precise micropipette positioning system based on continuous camera-robot recalibration and visual servoing," in *Automation Science and Engineering, 2009. CASE 2009. IEEE International Conference on*, 2009, pp. 609-614.

[11] J. Bert, S. Dembélé, and N. Lefort-Piat, "Performing weak calibration at the microscale. Application to micromanipulation," in *IEEE International Conference on Robotics and Automation, ICRA'2007.*, 2007, pp. 4937-4992.

[12] Y. Zhou and B. J. Nelson, "Calibration of a parametric model of an optical microscope," *Optical Engineering,* vol. 38, pp. 1989-1995, 1999.

[13] M. Ammi, V. Frémont, and A. Ferreira, "Automatic camera-based microscope calibration for a telemicromanipulation system using a virtual pattern," *Robotics, IEEE Transactions on,* vol. 25, pp. 184-191, 2009.

[14] M. Ammi, V. Fremont, and A. Ferreira, "Flexible microscope calibration using virtual pattern for 3-d telemicromanipulation," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, 2005, pp. 3888-3893.

[15] G. Li and N. Xi, "Calibration of a micromanipulation system," in *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, 2002, pp. 1742-1747.

[16] Y. Zhang, M. Han, C. Y. Shee, and W. T. Ang, "Automatic vision guided small cell injection: Feature detection, positioning, penetration and injection," in *Mechatronics and Automation, 2007. ICMA 2007. International Conference on*, 2007, pp. 2518-2523.

[17] C. C. Wang, "Extrinsic calibration of a vision sensor mounted on a robot," *IEEE Transactions on Robotics and Automation,* vol. 8, pp. 161-175, 1992.

[18] R. Horaud and F. Dornaika, "Hand-eye calibration," *The international journal of robotics research,* vol. 14, pp. 195-210, 1995.

[19] L. Yang, I. Paranawithana, K. Youcef-Toumi, and U.-X. Tan, "Automatic Vision-Guided Micromanipulation for Versatile Deployment and Portable Setup," *IEEE Transactions on Automation Science and Engineering,* 2017.

[20] L. Yang, K. Youcef-Toumi, and U.-X. Tan, "Towards automatic robot-assisted microscopy: An uncalibrated approach for robotic vision-guided micromanipulation," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, 2016, pp. 5527-5532.

[21] L. Yang, K. Youcef-Toumi, and U.-X. Tan, "Detect-Focus-Track-Servo (DFTS): A vision-based workflow algorithm for robotic image-guided micromanipulation," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, 2017, pp. 5403-5408.

[22] L. Yang, I. Paranawithana, K. Youcef-Toumi, and U.-X. Tan, "Self-Initialization and Recovery for Uninterrupted Tracking in Vision-Guided Micromanipulation," presented at the *IEEE International Conference on Intelligent Robots and Systems, IROS'2017*, Vancouver, Canada, 2017.

[23] I. Paranawithana, W. X. Yang, and U. X. Tan, "Tracking extraction of blastomere for embryo biopsy," in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2015, pp. 380-384.

[24] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision in C++ with the OpenCV Library*: O'Reilly Media, Inc., 2013.

[25] I. The MathWorks, Massachusetts, United States. (10/01/2018). *Estimate Geometric Transformation (R2017b ed.)*. Available: https://www.mathworks.com/help/vision/ref/estimategeometrictransformation.html

[26] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in computer vision*, ed: Elsevier, 1987, pp. 726-740.

[27] P. H. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Computer vision and image understanding,* vol. 78, pp. 138-156, 2000.