Confidence-Based Hybrid Tracking to Overcome Visual Tracking Failures in Calibration-Less Vision-Guided Micromanipulation

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

Abstract—This article proposes a confidence-based approach for combining two visual tracking techniques to minimize the influence of unforeseen visual tracking failures to achieve uninterrupted vision-based control. Despite research efforts in vision-guided micromanipulation, existing systems are not designed to overcome visual tracking failures, such as inconsistent illumination condition, regional occlusion, unknown structures, and nonhomogenous background scene. There remains a gap in expanding current procedures beyond the laboratory environment for practical deployment of vision-guided micromanipulation system. A hybrid tracking method, which combines motion-cue feature detection and score-based template matching, is incorporated in an uncalibrated vision-guided workflow capable of self-initializing and recovery during the micromanipulation. Weighted average, based on the respective confidence indices of the motion-cue feature localization and template-based trackers, is inferred from the statistical accuracy of feature locations and the similarity score-based template matches. Results suggest improvement of the tracking performance using hybrid tracking under the conditions. The mean errors of hybrid tracking are maintained at subpixel level under adverse experimental conditions while the original template matching approach has mean errors of 1.53, 1.73, and 2.08 pixels. The method is also demonstrated to be robust in the nonhomogeneous scene with an array of plant cells. By proposing a self-contained fusion method that overcomes unforeseen visual tracking failures using pure vision approach, we demonstrated the robustness in our developed low-cost micromanipulation platform.

Note to Practitioners—Cell manipulation is traditionally done in highly specialized facilities and controlled environment. Existing vision-based methods do not readily fulfill the need for the unique requirements in cell manipulation including prospective plant cell-related applications. There is a need for robust visual tracking to overcome visual tracking failure during the automated vision-guided micromanipulation. To address the gap in maintaining continuous tracking for vision-guided micromanipulation under unforeseen visual tracking failures, we proposed a purely visual data-driven hybrid tracking approach. Our proposed confidence-based approach combines two tracking techniques to minimize the influence of scene uncertainties, hence, achieving uninterrupted vision-based control. Because of its readily deployable design, the method can be generalized for a wide range of vision-guided micromanipulation applications. This method has the potential to significantly expand the capability of cell manipulation technology to even include prospective applications associated with plant cells, which are yet to be explored.

Index Terms—Cell manipulation, robot vision systems.

I. INTRODUCTION

THE importance of robotic micromanipulation system is well evidenced in its contribution toward the advancement of micromanipulation technology. Apart from more common industrial applications in microassembly and fabrication [1]–[3], the field of cell manipulation also benefits from the development of robotic micromanipulation system [4], [5]. Robotic micromanipulation benefited cell manipulation applications with speed, repeatability, and ease of the operation.

The vision-based control is an effective approach for robotic cell manipulation leveraging visual sensing from the microscope to control the manipulator. This vision-guided manipulation approach combines pure visual sensing and servoing to automate cell manipulation tasks. The advantage is that it is a self-contained framework without the need for an external active source of sensing other than the visual information from the microscope imaging system, which is a typical component in cell study. Such an approach may also catalyze a promising breakthrough in contactless visual servo control for force regulation in microinjection [6].

Apart from the importance of robotic vision-guided micromanipulation, this work is also motived by the existing gap in addressing visual tracking failure. The development of vision-based control in cell manipulation for general deployment is mainly restricted by the need for specialized equipment and system calibration beyond the user level. In addition, susceptibility to unforeseen visual tracking failures restricts

general deployment outside the laboratory environment. All these factors result in a gap in the development of robotic vision-guided micromanipulation including unprecedented applications such as plant cell manipulation [7]–[11].

With the advancements of plant micropropagation, there is increased popularity for plant cell manipulation. Plant genetic transfer has been widely adopted in the areas of plant improvement, disease elimination, and production of secondary metabolites [12]. Microinjection is considered to be one of the most efficient techniques for genetic transformation. It has also shown potential in the applications, such as isolation of high-yielding plant genotypes with stress tolerance and disease resistance capability [7]. The production of novel recombinant proteins influenced by the genetic plant transform has opened up new horizons in the pharmaceutical industry. Plant-derived medicinal substances are becoming increasingly popular due to lower production costs, absence of human pathogens, and the ability to assemble complex proteins precisely [13].

Despite research efforts [14]–[16], including our own work [17], in developing a robust vision-guided system for cell manipulation, many of the methods are designed to work under a controlled environment. Unlike processes in microassembly, typical cell manipulation procedures rely highly on manual operation. Although automated procedures and microma- nipulators exist, they usually require a tediously calibrated setting and controlled imaging environment. It is challenging to model the general tasks for autonomous execution in cell manipulation due to the procedure-specific complexity, which includes unstable lighting conditions, regional occlusion of the tracking region-of-interest (ROI), artifacts, unknown specimen geometries, and unpredictable tool-specimen interaction.

To overcome the problem of the unforeseen visual tracking failures, this work proposes a hybrid tracking method based on confidence measures derived from the selected visual trackers. In this work, we leverage previously developed motion-cue and template-based tracking workflow to maintain an uncali- brated [18], self-initializing [19], and self-recovery [20] track- servo approach while complementing the development with the ability to combat unforeseen visual tracking failures, which is not yet addressed in our previous work.

The contribution of this work is a confidence-based hybrid tracking method combining motion-cue feature detection and similarity score-based template matching. Unlike the existing work, there is conceptually no need for excessive specification of imaging conditions, tracking requirements, or prior mod- els of the physical setup through calibration. This approach overcomes problems in conventional visual tracking asso- ciated with the mentioned unforeseen visual uncertainties. Uninterrupted visual tracking can, therefore, achieve visual servo without prior scene assumptions. This concept can also be generalized for other vision-guided systems with little or no assumption about the imaging scene. While our long-term research goal works toward the holistic development of a self-contained automatic vision-guided system, this work focuses on the aspect of overcoming visual tracking failures. Fig. 1 shows the embodiment of our low-cost calibration-less vision-guided micromanipulation systems.
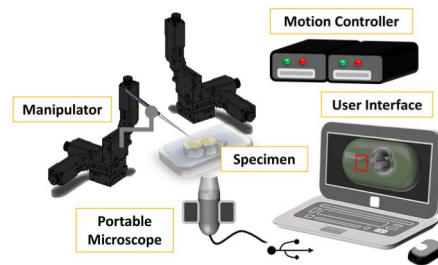

Fig. 1.   Low-cost portable vision-guided micromanipulation system.

In Section II, we review the related work in vision- guided micromanipulation systems to identify the limitations in existing methods. A detailed discussion of the proposed confidence-based hybrid tracking is presented in Section III followed by the experimental setup and evaluation procedures in Section IV. Section V covers the results and discussion of the experiment. Finally, the article concludes by reiterating the contribution and some remarks on the future work in Section VI.

II. Related Work

This section reviews related work in vision-guided micro- manipulation systems discussing the limitations of existing methods in the context of cell manipulation. The discussion justifies the motivation of this work by expounding on the gap in existing state-of-the-art development for vision-guided micromanipulation.

While there has been extensive research and development in vision-guided micromanipulation [14]–[16] over the past decade including highly sophisticated automatic end-effector tip locating method [21], current practices for cell manip- ulation continue to rely highly on manual operations [22]. This could be due to the challenges involved in integrat- ing the microscopic vision and micromanipulator control. A typical integration procedure between the microscope vision sensing and manipulator motion actuation involves calibra- tion [23]–[29]. This is a procedure that establishes the map- ping function between the coordinates of the imaging sensor domain and the actuation task space [30], [31].

To facilitate versatile deployment and portable setup of vision-guided micromanipulation, we have been developing a self-contained framework [17] that exploits the uncal- ibrated [18] vision-guided manipulation to achieve self- initialization [19] and ease of deployment. It is self-contained in a sense that there is no need for an external active source of sensing other than the visual information from the microscope, which is already in place for most cell manipulation proce- dures. However, existing visual tracking techniques are still susceptible to tracking disruption during visual servo, hence, motivating the exploration of a robust visual tracking method.

Our earliest vision-guided manipulation system uses tem- plate matching as a unified approach to achieving 2-D tracking and 3-D servoing of the tool tip [18]. The deviation from the focal plane is inferred from the similarity score change in the matching process. The similarity score-based depth compensation controls the depth of the tip from the camera.
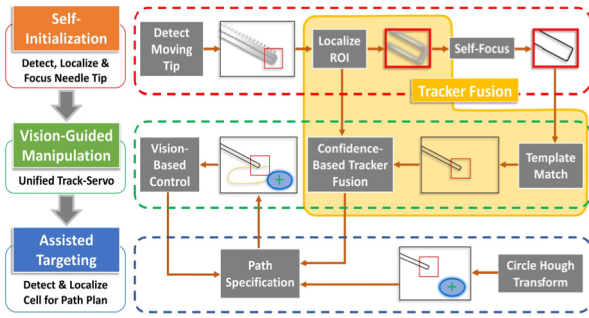
Fig. 2. Incorporating tracker fusion in the workflow of our autonomous vision-guided micromanipulation.

Without prior calibration, this mechanism maintains the tip in the focal plane as it moves toward a target.

Our second work further improves on the automation workflow by integrating the detect-focus-track-servo (DFTS) algorithm [19]. This is a self-initializing workflow that detects and focuses the tool tip automatically. This framework relieves users from the tedious process of manually locating and focusing the tool tip. The previously mentioned unified track-servo framework then performs vision-guided manipulation. This workflow is graphically summarized by the first two layers of the flowchart shown in Fig. 2. Without the self-initializing function, the operation requires manual localization and focusing before a base template can be obtained. The second layer essentially represents the unified track-servo method that tracks the tip in 2-D images and performs 3-D manipulation under the microscope.

To complete the automated workflow, we combine visual tracking and servoing of the tool with automatic image-based detection of specimens. The third layer of the flowchart shown in Fig. 2 depicts this component. The detection and recognition of specimen is an interesting topic for automated cell micromanipulation. We have previously demonstrated image-based detection and tracking of the micropipette and blastomere for preimplantation genetic diagnosis (PGD) [32]. The image-based recognition of the embryo structure for automatic microinjection on immobilized zebrafish embryo is also demonstrated by Wang *et al.* [16].

Despite the development in vision-guided robotic micro-manipulation systems, including [18] and [19], a bottleneck remains in the issue of visual uncertainties during cell imaging. Existing visual tracking methods [33], [34], which use low-level features, do not directly solve the problem of unforeseen visual disturbance for micromanipulation. Recently, advances in supervised learning approaches demonstrate an effective solution to combat visual tracking failure [35]–[37]. However, visual servo applications require low computational cost less computationally intensive tracking of low-level features. Hence, we designed a self-reinitialization and self-recovery method for uninterrupted visual tracking under tool-specimen interaction [20]. This method uses specially designed heuristics to identify the appropriate tracking mode to use based on the detected geometry and location of the specimen using our cell detection method [32]. However, using known conditions to switch tracking mode has limited contribution

toward the robust vision-guided manipulation. This approach limits the method from extending to more general scenarios, including deformable scene uncertainties and scene cluttering especially in applications that are associated with multiple cells or array of plant cells [8]–[11]. Our recent work on homography-based self-calibration for micromanipulation in plant cells attempts to address the latter problem [38], [39]. While the method is effective against visual disturbance, it assumes a homography transformation relationship between the image and manipulation plane. The performance may be subjected to adverse effects of mechanical uncertainty, especially when cell imaging is done at high magnification. A more generalized observation-based approach from a fusion of visual trackers is needed.

By formalizing a combined hybrid tracking approach based on the consistency of estimated potential locations, tracking is more robust against unforeseen uncertainties including artifacts and cluttered nonhomogenous background scene. The formalized method addresses the gap identified in the above discussion while maintaining the relevance to existing vision-based control methods. This is especially relevant to biomedical applications, such as embryo biopsy, blastomere isolation or PGD [40]–[42], and prospective plant cell manipulation applications [8]–[11], where the imaged scene could be challenging for visual tracking due to the nonhomogenous scene with an array of irregular cell dimensions. To build the provision for autonomous vision-guided robotic cell micromanipulation, there is a need to overcome unforeseen visual tracking failures.

III. CONFIDENCE-BASED HYBRID TRACKING

A. Conceptual Overview

The heart of our proposed solution toward overcoming unforeseen tracking failures is a confidence-based hybrid tracking method. This method encompasses the concept of confidence measures and the weighted averaging of the trackers. The selected tracking methods include motion-cue feature detect and similarity score-based template match, as introduced in Section II. As the two trackers are complementary in their sensitivity toward the spatial and temporal variability, fusing estimations from the two trackers provides robustness against unforeseen visual tracking failures, such as regional occlusion of tracking target and nonstatic background due to the unstable illumination. In essence, a normalized weight vector $(\hat{w}_v \ \hat{w}_u)$, comprising the confidence measures associated with the respective trackers, is used to obtain the weighted average of the estimates $(\mathbf{x}_v \ \mathbf{x}_u)$ where $\mathbf{x}$ is a vector $(x y)^T$ in image coordinates. Subscript v and u are used to associate motion-cue feature detect and template match, respectively. The selected confidence measures are expressed as a function of the statistical precision of the potential locations estimated by the two trackers. The weight vector is finally normalized such that the elements sum up to unity.

1) Leveraging Existing Development: The trackers are chosen to leverage our previous development of a self-initializing track-servo workflow in an automatic vision-guided micromanipulation system. The motion-cue feature detection estimates
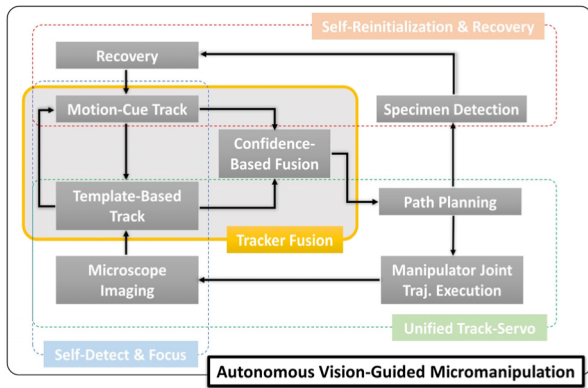
Fig. 3. Role of the hybrid tracking for fusing motion-cue and template-based trackers in autonomous vision-guided manipulation.

the position of a moving tool tip during the initialization by extracting low-level features from a difference image of two temporally adjacent image frames without the need for manual indication of a template. Subsequent tracking is done by score-based template matching that provides concurrent depth compensation for visual servoing in the 3-D workspace. While this article is self-contained in discussing the proposed hybrid tracking method, interested readers may refer to our previous work for more details on how the motion-cue feature detection achieves self-initialization and the novel design of the template-based unified track-servo algorithm in the DFTS workflow [19].

2) Enhancing Vision-Guided Micromanipulation: The proposed hybrid tracking technique integrates seamlessly into our automated vision-guided micromanipulation system [17] including recent work on self-reinitialization and recovery [20]. The role of the hybrid tracking is illustrated in the flowchart, as shown in Fig. 3.

The hybrid tracking technique utilizes motion-cue tracking in initialization and recovery phase while combining the template-based tracking for the track-servo task via a confidence-based fusion. This is achieved through the vectorization of the tracking certainty that can be mathematically incorporated into the automated workflow. The mathematical representations of the confidence measures and the fusion procedure using weighted averaging are further explained in Sections III-B and III-C).

B. Confidence Measure From Statistical Precision

The confidence measures are represented by the reciprocal variance of potential tool tip locations based on: 1) the motion-cue feature detection and 2) the score-based template matching. The statistical precision indicates how certain the trackers are about their estimates. The reciprocal of variance in potential positions of the tool tip is an attribute that is associated with the spatial consistency by comparing images at adjacent temporal intervals via two tracking approaches.

1) Reciprocal Variance of Low-Level Feature Locations: The prenormalized confidence measure w_v of the motion-cue feature detection is represented by the reciprocal variance,

i.e., statistical precision

$$w_v = \frac{1}{\sigma_v^2} = N \Big/ \left[\sum_i^N (x_i - \bar{x})^2 + \sum_i^N (y_i - \bar{y})^2 \right] \quad (1)$$

of N feature locations (x, y) with variance σ_v^2 in a given difference image

$$I_\Delta(x, y) = I_k(x, y) - I_{k-1}(x, y) \quad (2)$$

between temporally adjacent image frames, I_k and I_{k-1}. Subsequent operation of motion-related features will be processed in the difference image I_Δ.

Potential locations of the tracked tool tip can be represented by a feature in the form of an interest point that exhibits strong intensity gradient in more than one direction. These interest points can be detected using the Harris corner detector [43] with the corner response function written as follows:

$$C = \det(S) - \alpha \cdot \text{trace}^2(S) \quad (3)$$

where α is a tunable parameter while S is the structure tensor of the sum-of-square-difference (SSD)

$$D(x, y) = \sum_p^P \sum_q^Q [I_\Delta(p, q) - I_\Delta(p + x, q + y)]^2 \quad (4)$$

of the patch I centered on (p, q) and itself when shifted (x, y). Although only the furthest feature from the origin is taken as the tip position, the N points associated with the top values of C are considered potential locations. The closer these points cluster, the more likely they are associated with the moving tip in the difference image I_Δ. Hence, the confidence measure of the motion-cue tracker can be obtained using these points with (1).

2) Reciprocal Variance of Template Matches: In a similar fashion, the prenormalized confidence measure w_u of the template match tracking can be inferred from the variance in locations of the potential template matches. These potential locations are determined by the top N similarity scores N of patches compared to a reference base template. For an image I_k in Frame k, the similarity between a base template $g(p, q)$ and the regional image $f(p, q)$ is represented by the normalized cross-correlation coefficient

$$U = \sum_p^P \sum_q^Q \langle G \rangle \langle F \rangle \Big/ \left[\left(\sum_p^P \sum_q^Q \langle G \rangle^2 \right) \left(\sum_p^P \sum_q^Q \langle F \rangle^2 \right) \right]^{0.5} \quad (5)$$

where $\langle G \rangle = (g(p, q) - \bar{g})$ and $\langle F \rangle = f(p + x, q + y) - \bar{f}(x, y)$. Notation $\bar{g}$ and $\bar{f}$ are the mean intensity in the $P \times Q$ template and the overlapping patch, respectively.

The similarity score is an important measurement in our existing vision-guided workflow. It provides the feedback signal for depth compensation in the 3-D workspace and a means to determine the image coordinates of the tool tip that is associated with neighborhood intensity of high U values. It plays an important role in the vision-based control of our robotic micromanipulation platform [17]. The image coordinates with neighboring pixels associated with the highest U is

TABLE I
PSEUDOCODE FOR ADAPTIVE COMPENSATION

Algorithm: Score Maximizing for Z-Axis Servo

```
// Initialize
1.    Δz:= ΔU:=tol;
2.    w:= ᵃ U(g,f);
// Visual Servo in Z-direction
3. while ΔU≥tol || Suspend_Z==false
4.    (Δw, Δz):= (w- U(g,f), Δz* ᵇsgn(Δw));
5.       w:= U(g,f);
6. end
```

a. **U**() denotes a routine that implements Equation 5 b. **sgn**() extracts the sign

the estimated location obtained by template matching, which is subsequently used to update the visual servo loop for x- and y-axes. The z-axis of the manipulator is adjusted concurrently in a gradient ascending fashion during the manipulation to maximize U during visual servoing.

Table I is a pseudocode showing the tip motion compensated in the z-direction by Δz resulting in a change in score ΔU to converge to a preset tolerance tol. This depth compensation method has then been demonstrated effectively in previously proposed workflow [19] and portable micromanipulation platform [17]. The difference here is an end-loop condition, "Suspend_Z==True" when the tool and cell interact [20].

This work further utilizes the similarity score U to obtain potential template matches. The reciprocal variance of N potential matches is used to infer the confidence measure of the estimates in this tracking process. As U reflects the similarity between a given patch in the image compared with the template to be tracked, we express the prenormalized confidence measure

$$w_{\mathrm{u}} = \frac{U}{\sum U_k}\frac{1}{\sigma_{\mathrm{u}}^2} = \hat{U}\frac{1}{\sigma_{\mathrm{u}}^2} \qquad (6)$$

that is, the product of a normalized score $\hat{U}$ and the reciprocal variance. where ΣU_k is the summation of scores of the N potential template locations. Together with w_{v}, this confidence measure forms the prenormalized weight vector for the fusion process to be discussed in Section III-C.

C. Fusion via Normalized Weighted Averaging

Our hybrid tracking method is essentially a weighted averaging of the two spatial localization estimates derived from the motion-cue feature detection and score-based template matching in each temporal image frame. For a pair of estimates $(\mathbf{x}_{\mathrm{v}}, \mathbf{x}_{\mathrm{u}})$, the normalized weighted average of the estimates is expressed as follows:

$$\hat{\mathbf{x}}_{\mathrm{tip}} = \begin{pmatrix} \hat{w}_{\mathrm{v}} & \hat{w}_{\mathrm{u}} \end{pmatrix}\begin{pmatrix} \mathbf{x}_{\mathrm{v}} & \mathbf{x}_{\mathrm{u}} \end{pmatrix}^T. \qquad (7)$$

The normalized weight vector in terms of the variance is

$$\begin{pmatrix} \hat{w}_{\mathrm{v}} & \hat{w}_{\mathrm{u}} \end{pmatrix} = \begin{pmatrix} \dfrac{\sigma_{\mathrm{u}}^2/\hat{U}}{\sigma_{\mathrm{v}}^2 + \sigma_{\mathrm{u}}^2/\hat{U}} & \dfrac{\sigma_{\mathrm{v}}^2}{\sigma_{\mathrm{v}}^2 + \sigma_{\mathrm{u}}^2/\hat{U}} \end{pmatrix}. \qquad (8)$$
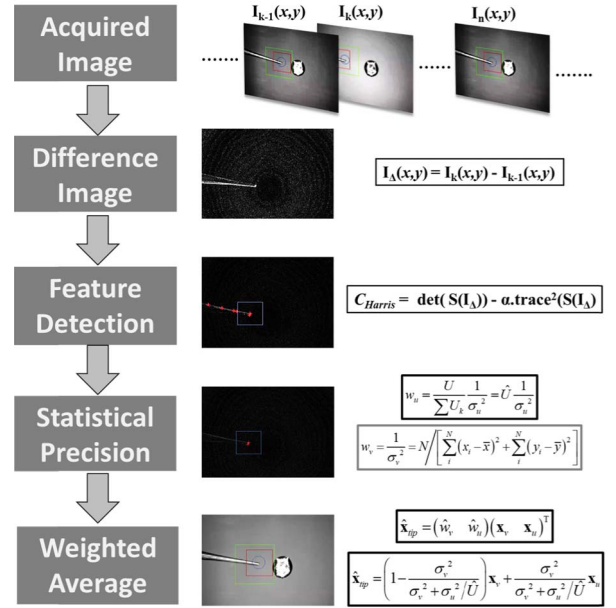

Fig. 4. Image processing techniques and fusion process of hybrid tracking for autonomous vision-guided manipulation.

Hence, the weighted average is expressed as follows:

$$\hat{\mathbf{x}}_{\mathrm{tip}} = \left(1 - \frac{\sigma_{\mathrm{v}}^2}{\sigma_{\mathrm{v}}^2 + \sigma_{\mathrm{u}}^2/\hat{U}}\right)\mathbf{x}_{\mathrm{v}} + \frac{\sigma_{\mathrm{v}}^2}{\sigma_{\mathrm{v}}^2 + \sigma_{\mathrm{u}}^2/\hat{U}}\mathbf{x}_{\mathrm{u}}. \qquad (9)$$

For the above equation to be valid, the variances of both the trackers cannot be zero at the same time. To deal with such ambiguity of choice in the trackers, we incorporate a logic check to only perform fusion when the main tracker produces variance greater than a predetermined value, σ_{min}. Hybrid tracking will not be present in instances where both trackers have variances that are close to zero as can be observed in some parts of the attached media file. As our vision-guided manipulation uses template matching for 3-D vision-based control, motion-cue tracking is regarded as the auxiliary tracker that refine the estimation when $\sigma_{\mathrm{u}} \geq \sigma_{\mathrm{min}}$. The value σ_{min} can be determined by the user depending on the imaging condition. Since the minimum division of the measurement is a pixel, setting $\sigma_{\mathrm{min}} = 1$ would generally give the two trackers almost equal priority.

The weighted average used as a feedback to execute motion command defined by the planned trajectory. Unlike existing fusion techniques [44], [45], our proposed fusion process requires no referencing of past frames or assumptions of motion. This makes the localization estimate in each frame independent of previous error and less susceptible to unforeseen visual tracking failures. Fig. 4 shows the fusion process and image processing techniques for the hybrid tracking.

D. Overcoming Visual Tracking Failures

The proposed hybrid tracking method provides a formal framework for the fusion of trackers to overcome scene- or event-related tracking failures. The former could be due to the unstable illumination, partial occlusion of artifacts,

nonhomogenous scene, and irregularity in target geometries. A solution for the latter problem has been demonstrated previously [20] using an event-based approach to logically toggles between trackers for self-reinitialization and recovery. However, this previous method assumes known information about the workflow and target geometry. The current proposed hybrid tracking method is a generalized approach for overcoming visual tracking failure due to the following two factors.

1) Failure in Event-Related Visual Disturbance: For event-related disturbance such as tool-cell interaction and deformation of the cell, the hybrid tracking method, based on statistical precision, naturally rectify the erroneous motion tracker in favor of template-based tracking. This complementary fusion approach is in contrast to our previous self-reinitialization and recovery method, which uses competitive fusion based on known events and prior knowledge of the cell geometry. By integrating the current proposed method, which is purely observation-driven, tracking failures could be avoided even when geometrical and foreseeable event-based prior knowledge are absent.

In [20], the location and dimension of a cell specimen can be automatically computed to predict occlusion and deformation to switch between the tracking modes. This approach addresses the problem of tool-cell interaction, which gives rise to a specific visual disturbance during cell manipulation but not for general unforeseen uncertainties.

In a tracking application for embryo biopsy [32], we detect and localize circular embryonic specimen using circle Hough transformation [46]. The Cartesian coordinates of potential feature points extracted using Canny edge detection [47] are mapped to their respective loci in the Hough space. The radius and location of the circle are determined based on overlapping counts of the loci. The square of the nominal radius of a circle is expressed as follows:

$$R_{cell}^2 = (i - x_{cell})^2 + (j - y_{cell})^2 \qquad (10)$$

where (i, j) and (x_{cell}, y_{cell}) are the image coordinates of the potential feature points and the image coordinates of the specimen center, respectively. A conical surface locus can be formed by plotting all the possible values of $(x_{cell}, y_{cell}, R_{cell})$ associated with a particular feature point (i, j). By further discretizing the Hough space with voxels, the number of counts they coincide with a locus indicates the vote from the potential feature points on a particular Hough space coordinates, i.e., circle's parameters. For a specimen of nominal radius R_{cell} and centered at (x_{cell}, y_{cell}), the extent of occlusion on a ROI centered at (x_{roi}, y_{roi}) by the specimen can be expressed as follows:

$$O_x = (R_{cell} + R_{roi}) - \left\| (x_{cell}, y_{cell})' - (x_{roi}, y_{roi})' \right\| \qquad (11)$$

where R_{roi} is the radius of the circle that circumscribes the ROI. Cell deformation is indicated by the condition

$$(R_{cell} + \Delta) - \left\| (x_{cell}, y_{cell})' - (x_{roi}, y_{roi})' \right\| > 0 \qquad (12)$$

where Δ is a contact margin that is introduced to control the sensitivity in deformation detection. The observation of the geometrical parameters (R_{cell}, R_{roi}) is shown in Fig. 5.
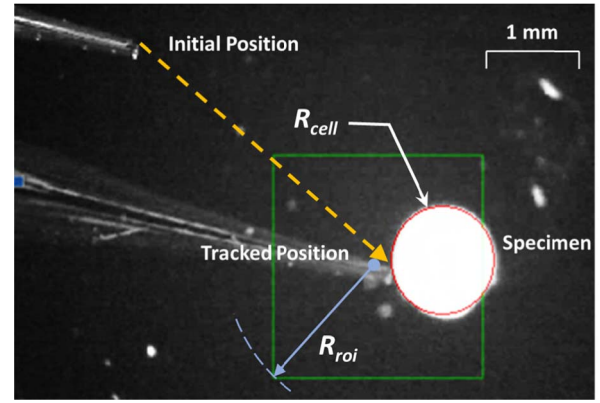

Fig. 5. Parameters associated with the assumed geometrical models.

The error in localization when using a fixed template and tracked using motion cues detection without recovery is 47 (=0.59 mm) and 12 pixels (=0.15 mm), respectively. When self-reinitialization and recovery are used, the error is 9 pixels (=0.11 mm), is reduced by more than fivefold. This is a reduction from more than 50% to less than 10% of the specimen size. For a more detailed discussion, readers may refer to the self-reinitialization and recovery approach proposed previously [20] under known target and an assumed event sequence.

The unification of the statistical hybrid tracking method and the previous event-based conditional approach of known cell geometry does not affect the operation during the micromanipulation procedure. They can be implemented in a common vision-based control system for robust visual tracking. In the event when assumed geometrical models in the scene like the dimension of the cell are detected, imminent tool-cell interaction and its extent could be predicted for a competitive fusion of the trackers based on logical conditions.

In the absence of known geometries detected, as in the case of cluttered arrays of plant cells, complementary fusion based on the statistical precision sets in for the tracking operation. The case of unknown geometry and nonhomogenous scene is further expounded in Section III-D2.

2) Failure in Nonhomogeneous Scene: Visual tracking is problematic in tracking objects in cluttered or nonhomogeneous background. It is very common for plant cells to be in arrays of multiple layers where boundaries are not distinctively defined. These unknown irregular geometries make it very difficult to predict tool-cell interaction, as shown in Fig. 6. Because the cells come in varying sizes, these irregular geometries are challenging for visual tracking.

The hybrid tracking method can automatically weight each tracker even when the scene involves the presence of an array of a plant cell as demonstrated in Section V. Our previous method [20], which is based on known geometries and imaging condition, cannot handle such an application. Nevertheless, the hybrid tracking method can be combined with prior information about the manipulation task seamlessly.

By formalizing a fusion technique with our proposed concept of confidence measures based on the observation data,
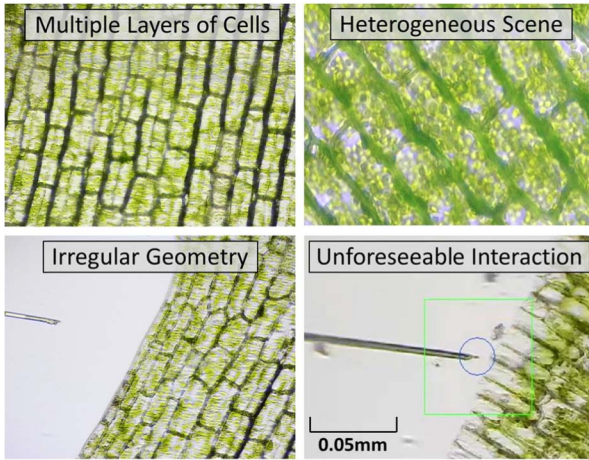
Fig. 6. Visual uncertainties in images of *Elodea* plant cells.

Fig. 7. Experimental setup and system with a portable digital microscope. Mode A (left) with a magnification $20\times$–$50\times$ for the animal cell. Mode B (right) with a magnification of $900\times$ for the plant cell.

we improve the decision-making process in the weighted averaging of the trackers, uninterrupted by unforeseen disturbance. This approach facilitates the extension of a more generalized context including plant cell application as will be demonstrated in experiments in Section V. Interested readers may also refer to a previous work related to the plant cell application using homography-based approach for online self-calibration [38], [39].

IV. EXPERIMENTAL STUDY

A. Conditions

The experiments were conducted to evaluate the tracking performance of motion-cue feature detect, score-based template match, and hybrid tracking under different conditions. These working conditions include unstable illumination and the presence of regional artifacts. A control study is also carried out to observe the tracking performance of a trajectory free of the two mentioned sources of influence. The validation on the nonhomogenous scene with plant cell application is also demonstrated with the complete visual track-servo workflow. The goal of the experiments is to evaluate and demonstrate the ability of the hybrid tracking to overcome unforeseen visual tracking failures using our developed vision-guided micromanipulation platform.

B. Setup

The experiments are performed using a portable micromanipulation platform developed in [17]. Fig. 7 shows the physical setup of the portable micromanipulation platform for two different orders of magnification. For experiments on plant cells at $900\times$ magnification, a second micromanipulator arm for the slide holder is installed to automatically focus the specimen. The setup is an implementation of the embodiment of our research vision toward a self-contained portable low-cost micromanipulation system. Because of the self-contained portable nature of the design, experiment on a wide range of dimensions can be readily set up.

Micromanipulation is executed using an actuated Cartesian micromanipulator (8MT173; Standa Ltd., Vilnius, Lithuania)
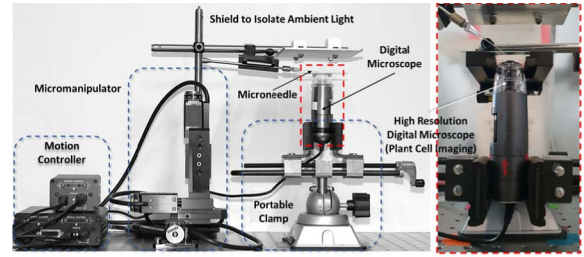
with a resolution of 1.25 μm/step and a working range of 20 mm in each axis. The motion control is performed using a dedicated controller (8SMC4; Standa Ltd) that interfaces with a host computer through USB. The manipulation speeds for the experiments are 62.5 μm/s unless specified otherwise. A USB digital microscope (AM7013MZT Dino-Lite, AnMo Corporation, New Taipei City, Taiwan) is used to acquire microscope images at a rate of 30 frames/s. This portable USB microscope camera has a 5-Megapixel CMOS image sensor. Its magnification options consist of a continuous range of $20\times$–$50\times$ and a separate option for $200\times$. The portable microscope also comes with built-in illumination with an adjustable polarizer.

To demonstrate the plant cell application, a microscope with a continuous magnification range of $700\times$–$900\times$ (AM4515T8 Dino-Lite Edge Series, AnMo Corporation) is used. For the experiment on plant cells, a fixed magnification factor of $900\times$ is used to acquire clear images of both plant cells and microneedle. The same motion control system is used with a speed of 12.5 μm/s for vision-guided micromanipulation because of the higher order of magnification.

V. RESULTS AND DISCUSSION

The performance of hybrid tracking, in comparison with the two original trackers, is presented and discussed in this section. The extent of adverse influence on the performance of visual tracking is observed by intentionally varying the illumination conditions and including regional artifacts. The former results in the unpredictable intensity distribution in the scene while the latter leads to an unforeseen regional occlusion. Both conditions are common in the deployment of the portable microscope in an uncontrolled environment, which is in alignment with our research vision toward ubiquitous micromanipulation beyond the laboratory setting. Our analysis covers qualitative observations as well as quantitative evaluations of the tracking data. A demonstration of the hybrid tracking method for automated vision-guided micromanipulation on nonhomogenous plant cell array is also presented to demonstrate this potentially useful application, which has yet to be investigated.

A. Qualitative Observation

Based on visual observation, we can identify discrepancies in the localization results of the respective tracking methods.
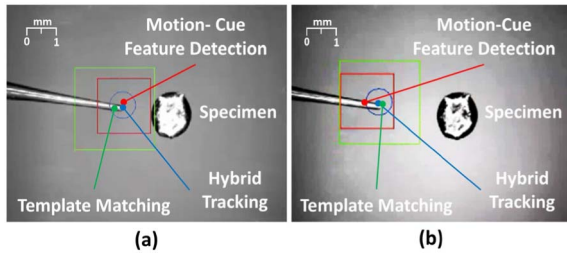
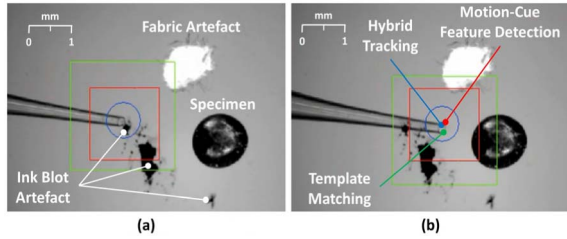Fig. 8. Robust against unstable lighting with hybrid tracking. (a) Under illuminated. (b) Over illuminated.

Fig. 9. Hybrid tracking robust against artifact. (a) Tool approaching artifacts. (b) Surrounded by artifacts.
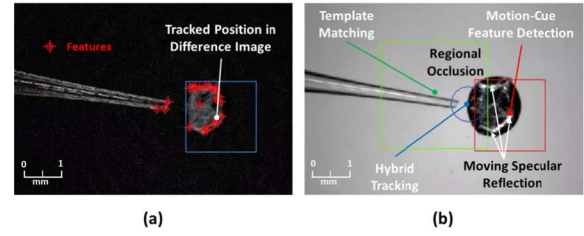

Fig. 10. Unstable illumination with the presence of regional occlusion. (a) Difference image for feature detect. (b) Microscope image with tool tip tracked.
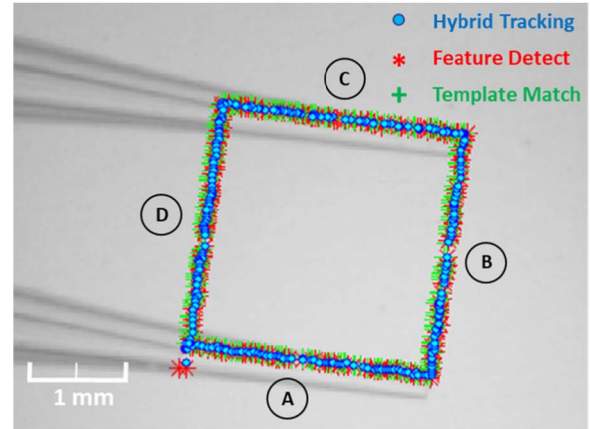

Fig. 11. Tracking linear paths on the order of ABCD; all three tracking methods produce the same measurement.

Qualitative discussion based on the visual inspection is presented in this section to justify the observations.

1) Unstable Scene Illumination: Fig. 8 shows the annotated screen captures of microscope images with the tool tip tracked during micromanipulation. The built-in illumination source of the microscope is continuously varied to artificially replicate the unstable lighting condition. It can be observed that the template match and motion-cue feature detect fail to localize the tip accurately shown in Fig. 8(a) and (b), respectively.

Hybrid tracking reduces the error as can be seen in both screen captures. As the two basic trackers are susceptible to a different type of visual tracking failure, hybrid tracking weighs the nonfailing tracker selectively to produce an effective accurate estimate.

2) Regional Occlusion of ROI: The regional occlusion occurs when tracking ROI got partially disturbed by artifacts. In this study, fabric and inkblot artifacts are intentionally introduced to the scene, as shown in Fig. 9. These artifacts act as regional occlusions that undermine the tracking performance of template matching. All three tracking approaches can tolerate slight partial occlusion of ROI, as shown in Fig. 9(a). However, the template match is susceptible to occlusion in the vicinity of the tracked tip, as shown in Fig. 9(b). The performance of motion-cue feature detection (red ROI), as shown in Fig. 9, is not at all affected by the regional occlusion as these artifacts are stationary in the scene. Unless the artifact occludes the tracked tip completely, tracking remains reliable using motion-cue feature detect. Under this occlusion condition, hybrid tracking again demonstrated the ability to rely highly on the right tracking method as observed.

3) Unstable Illumination and Regional Occlusion: Fig. 10 shows a situation when both motion-cue feature detect and template match were subjected to adverse influence concurrently. The moving specular lighting on the specimen in the scene negatively influenced the motion-cue feature detection and at the same time, partially occludes the ROI for the template match. Hybrid tracking was, however, able to reduce the error to a better estimate.

Based on a conservative evaluation, even at times when both trackers perform poorly, hybrid tracking will not be worse than the poorer performing tracker. However, determining the weaker tracker in unforeseen disturbance is nontrivial. There is, therefore, a need for a statistical method like the proposed hybrid tracking to autonomously weight each tracker's influence based on their confidence measures. This operation does not require camera calibration or robot-camera hand-eye calibration.

B. Quantitative Evaluation

In this section, we quantitatively evaluate the tracking performances based on a specified trajectory that consists of four linear path segments under three different imaging conditions, namely: 1) controlled condition; 2) unstable illumination by varying lighting; and 3) regional occlusion with artifacts. To analyze the tracking performance isolated from the servo task, visual tracking is performed offline.

1) Controlled Condition: In the controlled condition, the built-in lighting of the portable microscope is kept constant with ambient lighting unpredictability isolated using our test rig shown in Fig. 7. All three tracking methods are tested in this controlled scene condition. The tracked paths produced by the three approaches are similar, as observed in Fig. 11.

The tracking error is quantified by measuring the geometric error of the estimated positions in image coordinates.
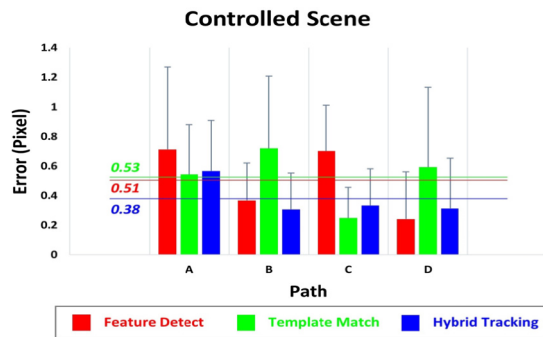
Fig. 12.   Tracking error under controlled scene; the mean errors of each method are marked by corresponding color lines on the chart.
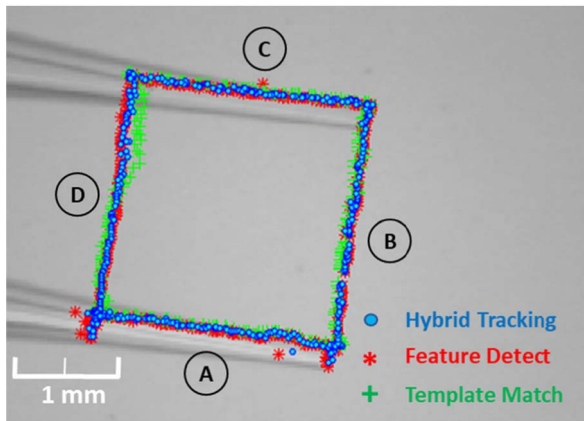

Fig. 14.   Tracking error under unstable illumination; the mean errors of each method are marked by corresponding color lines on the chart.

Fig. 13.   Tracking linear paths (ABCD) with unstable illumination; template match failed mainly at D; feature detect has erratic errors.

Fig. 15.   Tracking linear paths (ABCD) with regional occlusion; template match failed obviously at path B.

This geometric error is the shortest distance of the estimated position to the known linear path in pixels executed by a programmed linear trajectory of the manipulator. The straight-line path can be derived from the path joining the centroid of the stationary initial points read over multiple frames and that of its final stationary positions.

Fig. 12 shows the error using the three different tracking approaches for all four segments of the square trajectory. The hybrid tracking has the lowest mean error (0.38 pixels). Both features, detect and template matching, have similar performances with a mean error of 0.51 pixels and 0.53 pixels, respectively. This error is insignificant compared to our vision-based control tolerance of 1 pixel. All tracking methods produced mean error within subpixel in the controlled scene.

2) Unstable Illumination: The unstable illumination adversely influences the tracking performance of both motion-cue feature detect and template matching, as shown in Fig. 13. However, due to the different nature of the two tracking methods, they failed in different instances and seldom together, as shown in the plot. Mistakes in tracking by template match are mainly continuous and concentrated at path D while most of the errors from feature detect are erratic and sparse. These errors can be treated as outliers and removed readily. As this study focuses on overcoming unforeseen visual tracking failures, outlier removal methods are not considered. This further provides a conservative evaluation of the accuracy results.
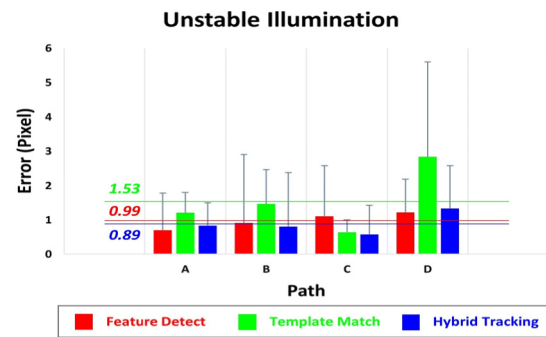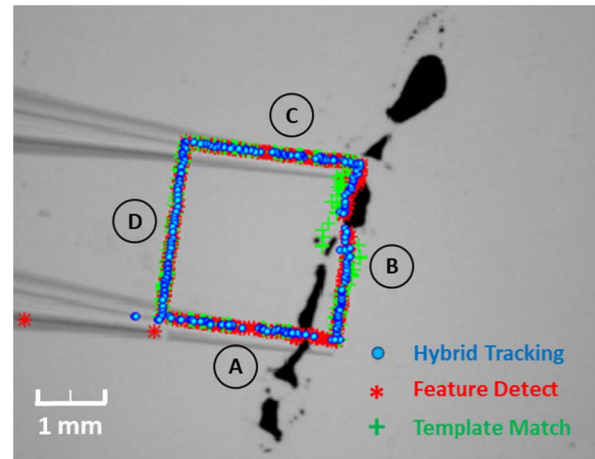
Hybrid tracking alleviates the uncertainty by combining the estimates using the confidence measures and weighted averaging. The failing tracker will naturally have a lower confidence measure, as explained earlier. Lower confidence measure leads to the reduced influence on the effective localization of that particular tracker.

Despite the unstable illumination, hybrid tracking continues to localize within subpixel accuracy and outperformed the two individual trackers. Template matching produced a mean error of 1.53 pixels while motion-cue feature detect results in a mean error of 0.99 pixels under unstable illumination. However, on the basis of visual inspection of the error trajectory shown in Fig. 13, the tracking performance of feature detect is more inconsistent with erratic deviations from the linear paths. Even though the cumulative error is smaller relative to that of template match results, several frames are deviating far out of the actual path. Fig. 14 suggests that hybrid tracking is associated with the lowest error (0.89 pixels).

3) Regional Occlusion: The presence of scene artifact heavily interrupts template matching, as shown in Fig. 15. The region at path B displayed misalignment in the tracking results of template matching with the actual linear path. This is a reasonable observation as regional occlusion disrupts the template matching process between the base template and potential match patches directly. It can also be seen that
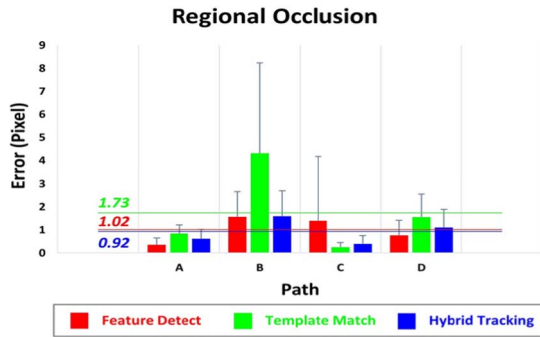
Fig. 16. Tracking error under regional occlusion. The mean errors of each method are marked by corresponding color lines on the chart.
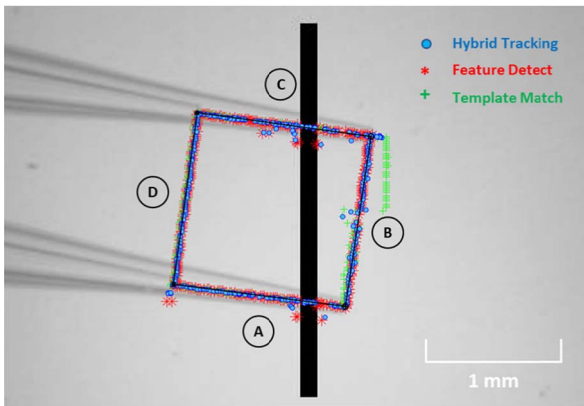

Fig. 17. Tracking linear paths (ABCD) under regional occlusion with digitally inserted artifacts showing similar tracking error as Fig. 15.

Fig. 18. Image sequence of tracking outcome under unforeseen adverse visual uncertainties.

template match was not able to track continuously when the tool tip is completely blocked by the inkblot at the ending segment in path A. We can infer that template match is highly sensitive to regional occlusion.

The hybrid tracking, in this condition, complements the shortcoming of template match using the confidence-weighted fusion approach. It can be observed that hybrid tracking spontaneously shifted its effective localization toward an estimate highly-weighted on the motion-cue feature detect method. This observation demonstrates the ability of the method to recognize weak measurement from a tracker and place favorable weightage on the other tracker.

Under the influence of regional occlusion, hybrid tracking continues to maintain a subpixel tracking mean error of 0.92 pixels while the two basic trackers have mean errors of 1.02 pixels and 1.73 pixels, respectively, as shown in Fig. 16. It is obvious that the regional occlusion represented by background cluttering of ink blob caused a prominent spike in the error.

To further validate that the observed results for regional occlusion due to the artifact is repeatable for analysis and validation, we digitally overlay an artifact on the video data of a controlled environment to qualitative investigate if the influence is consistent. Fig. 17 shows the imaging scene with tracked paths. It can be seen that paths A, B, and C are all adversely affected by the artifact.
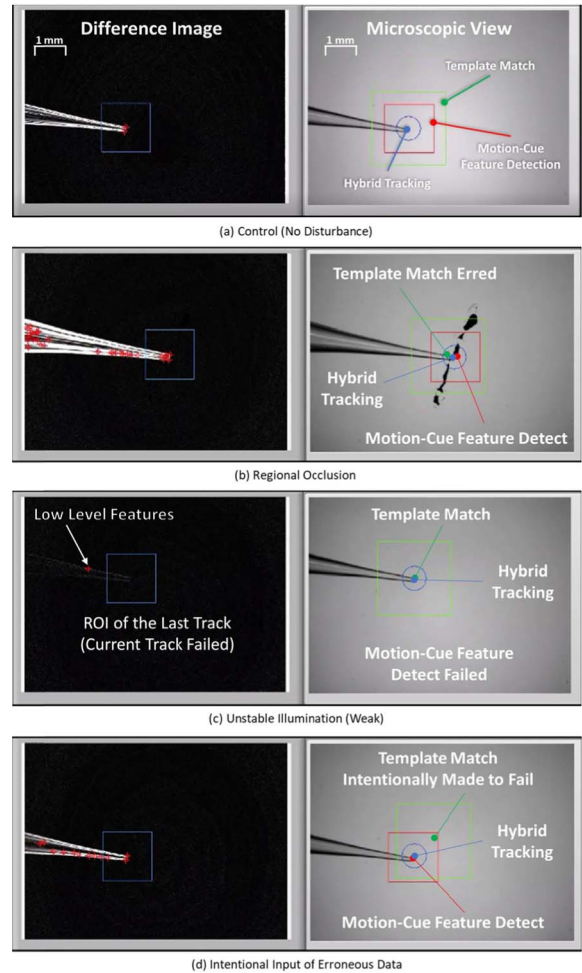
C. Demonstrating Robustness in Hybrid Tracking

The hybrid tracking in the automated workflow of vision-guided micromanipulation for the procedures described in the experiments is demonstrated in the attached media file. All three tracking methods, namely: 1) motion-cue feature detection; 2) score-based template matching; and 3) the hybrid tracking, are performed in the various conditions, as presented in the video of the microscopic view during the micromanipulation. These conditions are investigated in the scene with: 1) controlled condition; 2) regional occlusion; 3) unstable illumination; and 4) intentional disturbance to template match. Hybrid tracking always performs better than the poorest performing tracker and performs the best on average as demonstrated in by having the lowest mean errors in all imaging conditions. This shows the robustness of the method against tracking failures.

Although hybrid tracking may not produce the most accurate localization in certain cases, it is more robust against tracking failures. This can be shown by the various failure examples in Fig. 18.

Fig. 18(a) shows the controlled condition where all trackers are consistent. Regional occlusion shown in Fig. 18(b) resulting in the failure of the template-based tracker. This error is,

however, rectified by hybrid tracking. Fig. 18(c) shows the image of unstable scene illumination. There is tracking failure in both template match and motion-cue feature detection. The tracking is, nevertheless, rectified by hybrid tracking. The latter only experienced failure in motion-cue detection tracker. This error is rectified by the hybrid tracking with the tracked location in compliance with the template match. When an error is intentionally planted to the template match tracking data, the method was able to correct it based on autonomously increasing the weighting for motion-cue detection tracker, as shown in Fig. 18(c).

The tracking is done at a typical computation speed of about 20 frames/s. The computational load is dependent on several conditions including requirements in image resolution, control precision, error tolerance, etc. These parameters can be adjusted to lower computational load based on the user's specifications accordingly. Combining all the imaging frames ($N = 1360$) in the uncontrolled randomized scenes of visual disturbances, as discussed in Sections V-B2 and V-B3), we compare the observed errors associated with hybrid tracking against that of template matching and feature detection approach individually. Tracking improvement is observed by taking the difference between the errors of hybrid tracking and that of the other two individual trackers. A paired-sample t-test is performed for each comparison to test if the improvement of our proposed hybrid tracking is supported by the observation data statistically. The null hypothesis that there is no improvement or negative improvement (i.e., improvement ≤ 0) for template match and feature detection approaches are rejected at 5% significance interval with $p = 0.0059$ and $p = 1.41 \times 10^{-27}$, respectively. Hence, improvements in hybrid tracking over template matching and feature detection are supported by the observed data.

D. Validation on Plant Cell Applications

We demonstrate the use of the hybrid tracking on an array of plant cells to overcome the problem of unforeseen visual tracking failures in applications where geometries in the scene are highly irregular and uncertain. Our previous method [20], which is based on known geometries and imaging condition, is not able to handle such an application. The haphazard motion of the micro-organelles including chloroplasts can lead to the regional occlusion. For validation purpose, the study of the plant cell application will not involve controlled conditions. Visual observation and video demonstration (see attached media) are used for this validation study.

Experiments were conducted to demonstrate the performance of the hybrid tracking for onsite plant cell manipulation studies. A sample of waterweed (*Elodea)* was used as the plant cell specimen for the experiments which introduced substantial regional occlusion with a nonhomogeneous background in the visual scene. The microscopic images of the scene with the *Elodea* cells are shown in Fig. 19. *Elodea* aquatic plant often used for aquarium vegetation is rich in chloroplast and has two layers of cell arrays. Multi-layered cellular structures and moving micro organelles (i.e., chloroplast) of the *Elodea* plant specimen produce a highly complex visual scene.
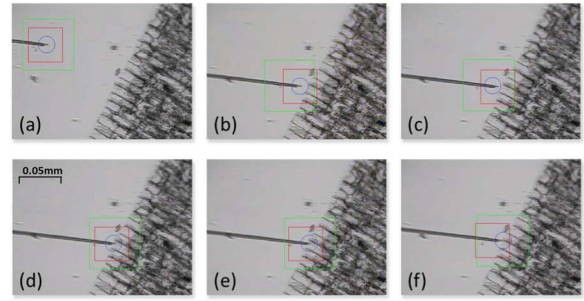

Fig. 19.   (a) and (d) Hybrid tracker consistent with both trackers. (b) and (c) Hybrid tracker rectifies error in template-based tracking. (e) and (f) Hybrid tracker rectifies the error in motion-cue tracking.

The irregular geometry and physical uncertainties, as shown in Fig. 19, limits the use of prior knowledge for modeling. Unforeseeable interaction makes predicting the dynamics of plant cells difficult as observed. Hence, micromanipulation tasks in plant cells deal with more challenging working conditions compared to the manipulation of an isolated single cell, which is usually the norm for animal/human cell manipulation. Before being in proximity to the plant cell shown in Fig. 19(a), the trackers are consistent with one another similar to the situation of a controlled environment. It can be observed in Fig. 19(b)–(f) that the hybrid tracking approach represented by the blue round ROI can rectify tracking errors from template-based and motion-cue feature detection approach represented by big green and small red ROI, respectively.

It is worth noting that, like any tracking method, it is very difficult, if not impossible, to work under all possible conditions exhaustively. There are specific operating conditions where both trackers might fail. An extreme case, for example, is the total occlusion. Although beyond the scope of this work, we investigated the use of motion trajectory and a homography-based self-calibrating micromanipulation approach in a separate study [38], [39]. Evaluation of currently proposed method shows that vision-guided micromanipulation is viable with our proposed approach, despite using only visual data. The goal of the proposed method is to contribute toward existing state-of-the-art. This validation experiments on plant cells complement our long-term research goal of making cell manipulation ubiquitous and extending vision-guided micromanipulation to be deployed on-the-fly for plant cell studies.

VI. Conclusion

In this work, we leverage previously developed motion-cue feature detect and template-match tracking workflow to maintain an uncalibrated, self-initializing, and self-recovery track-servo approach while addressing the problem of unforeseen visual disturbance. This work focuses on instantaneous estimation of the trackers, which, unlike existing fusion techniques, makes no assumption about motion models and noise in the system.
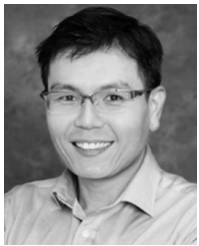
By proposing a hybrid tracking method to enhance robustness in our self-contained micromanipulation platform, we hope to realize a research goal for autonomous vision-guided cell micromanipulation. While the proposed method

uses only visual observation for tracking and automated micromanipulation, it is a crucial provision to prospective sophisticated micromanipulation applications. Our approach can be generalized to a much broader application including dealing with a complicated scene of uncertain physical models and geometries, such as plant cells, as demonstrated in the experimental studies. Because our method uses low-level features, the approach may eventually expand to tracking of different targets for broader application scope. The future work will investigate the feasibility of other fusion techniques [48]–[50] that statistically infer better estimates from past observations and multiple data sources other than pure visual data.

REFERENCES

[1] B. Tamadazte, N. L.-F. Piat, and S. Dembélé, "Robotic micromanipulation and microassembly using monoview and multiscale visual servoing," *IEEE/ASME Trans. Mechatronics*, vol. 16, no. 2, pp. 277–287, Apr. 2011.

[2] A. Sulzmann and J. Jacot, "3D computer vision for microassembly stations and microfabrication," *Proc. SPIE*, vol. 3202, pp. 42–51, Jan. 1998.

[3] M. Wang, X. Lv, and X. Huang, "Self-optimizing visual servoing control for microassembly robotic depth motion," in *Proc. IEEE Int. Conf. Inf. Acquisition*, Aug. 2007, pp. 482–486.

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

[5] Z. Zhang, X. Wang, J. Liu, C. Dai, and Y. Sun, "Robotic micromanipulation: Fundamentals and applications," *Annu. Rev. Control, Robot., Auto. Syst.*, vol. 2, no. 1, pp. 181–203, 2019.

[6] Q. Xu, "Visual servo control with force regulation for microinjection," in *Micromachines for Biological Micromanipulation*. Cham, Switzerland: Springer, 2018, pp. 209–223.

[7] H. Daniell, C.-S. Lin, M. Yu, and W.-J. Chang, "Chloroplast genomes: Diversity, evolution, and applications in genetic engineering," *Genome Biol.*, vol. 17, no. 1, 2016, Art. no. 134.

[8] M. Han, Y. Zhang, C. Y. Shee, T. F. Chia, and W. T. Ang, "Plant cell injection based on autofocusing algorithm," in *Proc. IEEE Conf. Robot., Autom. Mechatronics*, Sep. 2008, pp. 439–443.

[9] Y. Zhang, M. Han, C. Y. Shee, T. F. Chia, and W. T. Ang, "Automatic vision guided small cell injection: Feature detection, positioning, penetration, and injection," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2007, pp. 2518–2523.

[10] G. Neuhaus and G. Spangenberg, "Plant transformation by microinjection techniques," *Physiol. Plantarum*, vol. 79, no. 1, pp. 213–217, 1990.

[11] T. Kunkel. (2015). *Microinjection into Plant Cells of Etiolated Seedlings*. Accessed: Jan. 24, 2018. [Online]. Available: http://www.labonline.com.au/content/life-scientist/article/microinjection-into-plant-cells-of-etiolated-seedlings-981244058

[12] A. Hussain, I. A. Qarshi, H. Nazir, and I. Ullah, "Plant tissue culture: Current status and opportunities," in *Recent Advances in Plant in Vitro Culture*. Rijeka, Croatia: InTech, 2012.

[13] A. L. Rivera, M. Gómez-Lim, F. Fernández, and A. M. Loske, "Physical methods for genetic plant transformation," *Phys. Life Rev.*, vol. 9, no. 3, pp. 308–345, 2012.

[14] S. Yu and B. J. Nelson, "Microrobotic cell injection," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, vol. 1, May 2001, pp. 620–625.

[15] Y. Sun and B. J. Nelson, "Biological cell injection using an autonomous microrobotic system," *Int. J. Robot. Res.*, vol. 21, nos. 10–11, pp. 861–868, 2002.

[16] 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, no. 9, p. e862, 2007.

[17] L. Yang, I. Paranawithana, K. Youcef-Toumi, and U.-X. Tan, "Automatic vision-guided micromanipulation for versatile deployment and portable setup," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1609–1620, Oct. 2018.

[18] L. Yang, K. Youcef-Toumi, and U.-X. Tan, "Towards automatic robot-assisted microscopy: An uncalibrated approach for robotic vision-guided micromanipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 5527–5532.

[19] 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 *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2017, pp. 5403–5408.

[20] L. Yang, I. Paranawithana, K. Youcef-Toumi, and U.-X. Tan, "Self-initialization and recovery for uninterrupted tracking in vision-guided micromanipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 1127–1133.

[21] J. Liu *et al.*, "Locating end-effector tips in robotic micromanipulation," *IEEE Trans. Robot.*, vol. 30, no. 1, pp. 125–130, Feb. 2014.

[22] M. Savia and H. N. Koivo, "Contact micromanipulation—Survey of strategies," *IEEE/ASME Trans. Mechatronics*, vol. 14, no. 4, pp. 504–514, Aug. 2009.

[23] L. S. Mattos and D. G. Caldwell, "A fast and precise micropipette positioning system based on continuous camera-robot recalibration and visual servoing," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2009, pp. 609–614.

[24] J. Bert, S. Dembele, and N. Lefort-Piat, "Performing weak calibration at the microscale, application to micromanipulation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Apr. 2007, pp. 4937–4992.

[25] Y. Zhou and B. J. Nelson, "Calibration of a parametric model of an optical microscope," *Opt. Eng.*, vol. 38, no. 12, pp. 1989–1995, 1999.

[26] M. Ammi, V. Fremont, and A. Ferreira, "Automatic camera-based microscope calibration for a telemicromanipulation system using a virtual pattern," *IEEE Trans. Robot.*, vol. 25, no. 1, pp. 184–191, Feb. 2009.

[27] M. Ammi, V. Fremont, and A. Ferreira, "Flexible microscope calibration using virtual pattern for 3-D telemicromanipulation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Apr. 2005, pp. 3888–3893.

[28] G. Li and N. Xi, "Calibration of a micromanipulation system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 2, Sep./Oct. 2002, pp. 1742–1747.

[29] J. A. M. Rodríguez, "Microscope self-calibration based on micro laser line imaging and soft computing algorithms," *Opt. Lasers Eng.*, vol. 105, pp. 75–85, Jun. 2018.

[30] C.-C. Wang, "Extrinsic calibration of a vision sensor mounted on a robot," *IEEE Trans. Robot. Autom.*, vol. 8, no. 2, pp. 161–175, Apr. 1992.

[31] R. Horaud and F. Dornaika, "Hand-eye calibration," *Int. J. Robot. Res.*, vol. 14, no. 3, pp. 195–210, 1995.

[32] I. Paranawithana, W.-X. Yang, and U.-X. Tan, "Tracking extraction of blastomere for embryo biopsy," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2015, pp. 380–384.

[33] S. Gauglitz, T. Hoellerer, and M. Turk, "Evaluation of interest point detectors and feature descriptors for visual tracking," *Int. J. Comput. Vis.*, vol. 94, no. 3, pp. 335–360, Sep. 2011.

[34] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

[35] S. Krebs, B. Duraisamy, and F. Flohr, "A survey on leveraging deep neural networks for object tracking," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 411–418.

[36] G. Ning *et al.*, "Spatially supervised recurrent convolutional neural networks for visual object tracking," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2017, pp. 1–4.

[37] X. Jiang, P. Li, X. Zhen, and X. Cao, "Model-free tracking with deep appearance and motion features integration," 2018, *arXiv:1812.06418*. [Online]. Available: https://arxiv.org/abs/1812.06418

[38] I. Paranawithana, L. Yang, Z. Chen, K. Youcef-Toumi, and U.-X. Tan, "Scene-adaptive fusion of visual and motion tracking for vision-guided micromanipulation in plant cells," in *Proc. IEEE 14th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2018, pp. 1434–1440.

[39] I. Paranawithana, Z. H. Chau, L. Yang, Z. Chen, K. Youcef-Toumi, and U.-X. Tan, "Automatic cell segmentation with robust scene-adaptive tracking for vision-guided micromanipulation in plant cells," presented at the IEEE Int. Conf. Robot. Autom., Montreal, QC, Canada, 2019.

[40] V. C. Y. Lee, J. F. C. Chow, W. S. B. Yeung, and P. C. Ho, "Preimplantation genetic diagnosis for monogenic diseases," *Best Pract. Res. Clin. Obstetrics Gynaecol.*, vol. 44, pp. 68–75, Oct. 2017.

[41] K. Sermon, A. Van Steirteghem, and I. Liebaers, "Preimplantation genetic diagnosis," *Lancet*, vol. 363, no. 9421, pp. 1633–1641, 2004.

[42] J. L. Simpson, "Preimplantation diagnosis and other modern methods for prenatal diagnosis," in *Genetic Steroid Disorders*, O. Lekarev, A. Parsa, T. T. Yuen, B. W. O'Malley, and G. D. Hammer, Eds. San Diego, CA, USA: Academic, 2014, ch. 13C, pp. 343–350.

[43] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, vol. 15, 1988, p. 50.

[44] X. Gu, H. Wang, L. Wang, and C. Pan, "Adaptive multi-cue fusion for visual target tracking based on uncertainly measure," in *Proc. IEEE 25th Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, Nov. 2010, pp. 1–8.

[45] P. Brasnett, L. Mihaylova, D. Bull, and N. Canagarajah, "Sequential Monte Carlo tracking by fusing multiple cues in video sequences," *Image Vis. Comput.*, vol. 25, no. 8, pp. 1217–1227, 2007.

[46] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognit.*, vol. 13, no. 2, pp. 111–122, 1981.

[47] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[48] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[49] P. Pérez, J. Vermaak, and A. Blake, "Data fusion for visual tracking with particles," *Proc. IEEE*, vol. 92, no. 3, pp. 495–513, Mar. 2004.

[50] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.

Ishara Paranawithana (M'18) received the bachelor's degree in electronic and telecommunication engineering from the University of Moratuwa, Moratuwa, Sri Lanka, in 2016.

He joined the Singapore University of Technology and Design, Singapore, as a Research Assistant. His research interests include computer vision-based motion control, ultrasound image-guided percutaneous interventions, biomedical signal processing for assistive, and rehabilitation technologies.

Kamal Youcef-Toumi (M'09) received the B.S. degree in mechanical engineering from the University of Cincinnati, Cincinnati, OH, USA, in 1979, and the M.S. and Sc.D. degrees in mechanical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1981 and 1985, respectively.

He joined the MIT Mechanical Engineering Department faculty in January 1986 as an Assistant Professor. Prior to his faculty appointment, he served as a Research Associate and Lecturer in this department. His research has focused primarily on modeling, design, instrumentation, and control systems theory and their applications to dynamic systems.

Liangjing Yang (M'11) received the B.Eng. and M.Eng. degrees in mechanical engineering from the Department of Mechanical Engineering, National University of Singapore, Singapore, in 2008 and 2011, respectively, and the D.Eng. degree from the University of Tokyo, Tokyo, Japan, in 2014, working with the Bio-Medical Precision Engineering Laboratory.

He is currently an Assistant Professor with the Zhejiang University/University of Illinois at Urbana–Champaign Institute (ZJU-UIUC Institute), Haining, China. His research interests are in robotics and computer vision focusing on surgical robots, image-guided interventions, and vision-based control.

U-Xuan Tan (S'07–M'10) received the B.Eng. and Ph.D. degrees from Nanyang Technological University, Singapore, in 2005 and 2010, respectively.

From 2009 to 2011, he was a Post-Doctoral Fellow with the University of Maryland, College Park, MD, USA. From 2012 to 2014, he was a Lecturer with the Singapore University of Technology and Design, Singapore, where he took up a research intensive role in 2014 and is currently an Assistant Professor. His research interests include mechatronics, sensing, control, and onsite precision applications.