Title

A Conditional Random Field Model For Tracking In Densely Packed Cell Structures - A Technical Report

Authors

Chakraborty, Anirban
Roy-Chowdhury, Amit

Publication Date

2014

A Conditional Random Field Model For Tracking In Densely Packed Cell Structures - A Technical Report

Anirban Chakraborty, and Amit K. Roy-Chowdhury, *Senior Member, IEEE*

***Abstract*—Modern live imaging technique enables us to observe the internal part of a tissue over time by generating serial optical images containing spatio-temporal slices of hundreds of tightly packed cells. Automated tracking of plant and animal cells from such time lapse live-imaging datasets of a developing multicellular tissue is required for quantitative, high throughput analysis of cell division, migration and cell growth. In this paper, we present a novel cell tracking method that exploits the tight spatial topology of neighboring cells in a multicellular field as contextual information and combines it with physical features of individual cells for generating reliable cell lineages. The 2D image slices of multicellular tissues are modeled as a Conditional Random Field and spatio-temporal cell to cell correspondences are obtained by performing inference on this CRF using loopy belief propagation. We present results on (3D+t) confocal image stacks of Arabidopsis shoot meristem and show that the method is capable of handling many visual analysis challenges associated with such cell tracking problems, viz. poor feature quality of individual cells, low SNR in parts of images, variable number of cells across slices and cell division detection.**

I. Introduction

In developmental biology, the causal relationship between cell growth patterns and gene expression dynamics has been one of the major topics of interest. A proper quantitative analysis of the cell growth and division patterns in both the plant and the animal tissues has remained mostly elusive so far. Information such as rates and patterns of cell expansion and cell division play a critical role in understanding morphogenesis in a tissue. The need for quantifying the cellular parameters such as average rate of cell divisions, cell cycle lengths, cell growth rates etc. and observing their time evolution is, therefore, extremely important.

Towards this goal, with the advancements in microscopy and other imaging techniques, time lapse videos are being collected to quantify the behavior of hundreds of cells in a tissue over multiple days. For visualizing the cells over time within a densely packed multilayer tissue, one such in-vitro time-lapse microscopy technique is confocal laser scanning microscopy (CLSM) based *Live Cell Imaging*. With this technique, optical cross sections of the cells in the tissue are taken over multiple observational time points to generate spatio-temporal image stacks. For high-throughput analysis of these large volumes of image data, development of fully automated image analysis pipelines are becoming necessities, thereby giving rise to many new automated visual analysis challenges.

Automated cell tracking with cell division detection is one of the major components of all such pipelines (such as [1]). The computational challenges related to a robust design of cell tracker come from multiple sources such as variable number of cells in the field of view (FoV), deformation of cell shapes, complex topologies of cell clusters, low SNR in the images, etc. In this paper, we present an automated visual tracker for cells tightly packed in developing multilayer tissues. This calls for developing strategies for *temporal* associations of the cells. Moreover, since at every time point of observation a cell could be imaged across multiple spatial images, the tracking method must be capable of finding correspondences in the *spatial* direction as well. Beyond these, the tracker has to be able to detect cell divisions, detect new cells as the deeper layers of the tissues are imaged, differentiate between cells in a close neighborhood sharing similar physical features and generate correct matches in presence of low SNR. These challenges are evident in the sample CLSM image stack of a live Arabidopsis shoot meristem, as shown in Fig. 1.

A. Related Work and Our Contributions

There has been some work on automated tracking and segmentation of cells in time-lapse images, for both plants and animals. One of the well-known approaches for segmenting and tracking cells is based on evolution of active contours [2], [3], [4], [5], [6]. However, this method is not suitable for tracking where all the cells are in close contact with each other and share very similar physical features, nor is there any reported result on spatial correspondence. In fact, in spatio-temporal image stacks where the cells are arranged in compact multilayer structure, slice of a new cell can legitimately appear at the exact same spatial location as that of a different cell located in the layer just above it. This characteristic, along with the fact that these tightly packed cells are mostly stationary can force the active contour based tracker to generate false spatial tracks.

The Softassign method uses the information on point location to simultaneously solve both the problem of global correspondence as well as the problem of affine transformation between two time instants iteratively [7], [8]. However, these methods are more suitable for aligning global features than finding correspondences between non-uniformly growing individual cells. Although [8] present a sample result on SAM

A. Chakraborty and A. K. Roy-Chowdhury are with the Department of Electrical Engineering, University of California, Riverside, CA, 92521 USA e-mail: amitrc@ee.ucr.edu

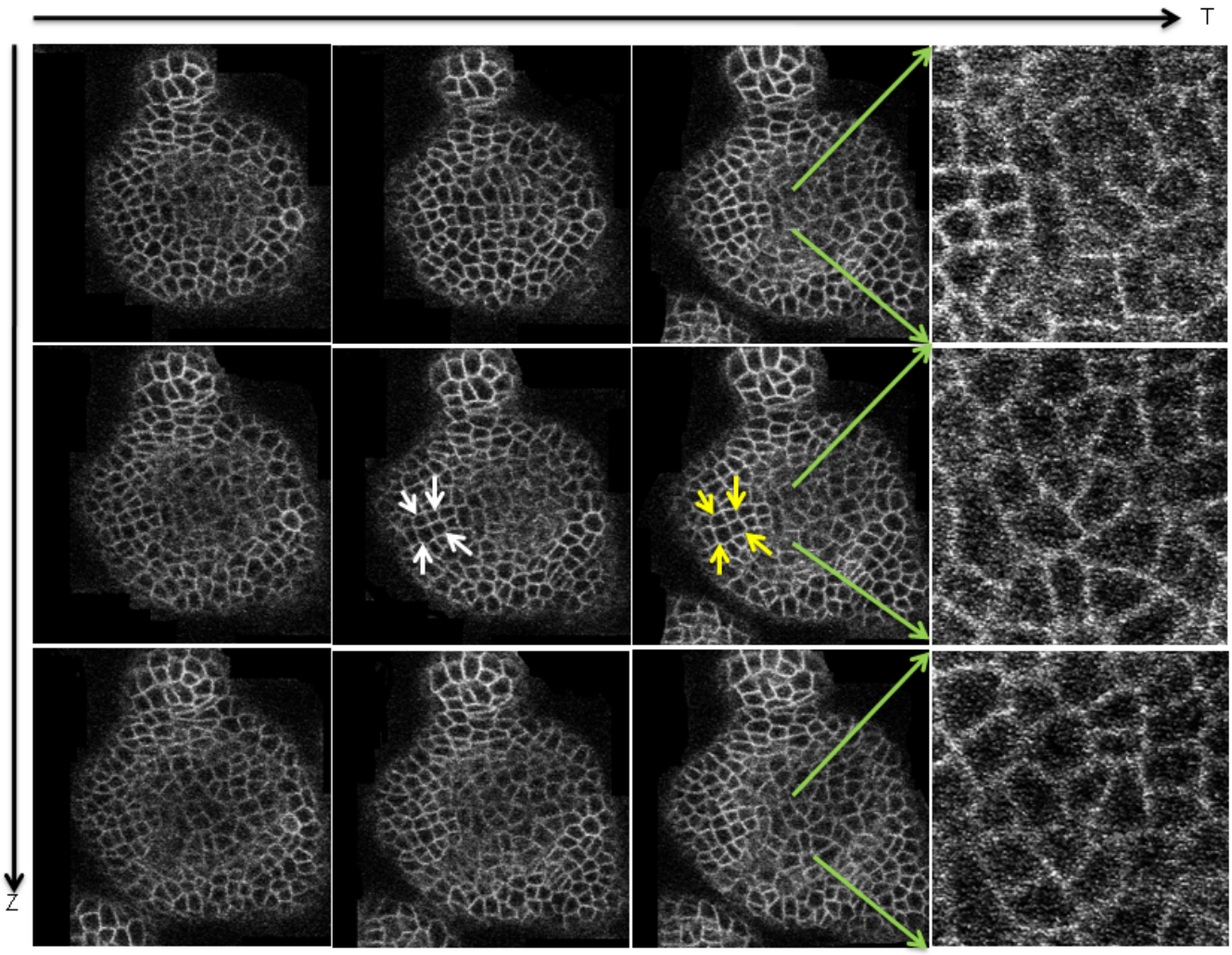

Fig. 1. A typical 4D (X-Y-Z-T) live-imaging data. A live Arabidopsis shoot meristem tissue is imaged using a confocal laser scanning microscope at multiple time points. The plasma membranes of the cells are stained with fluorescent proteins and that is why the cell walls are the only visible parts. Each of the first three columns of images presents Z stack of image slices, i.e the cross sections of the tissue imaged at various depths of it. When such images are collected over time to capture the growth of the tissue along with that of individual cells in it, it forms a 4D image stack. As can be seen from the figure, there are various challenges associated with the problem, viz. growth/deformation of the cells in the tissue, stereotypical cell shapes in the tissue and hence less discriminative physical features (as an example, 4 cells from a close neighborhood are marked with white and yellow arrows respectively in two consecutive time points which have very similar shapes and sizes), minor shifts between images and low SNRs in the central regions of the tissue. We have zoomed into these low SNR regions in the 4^{th} column of the figure. As seen, it is really difficult to even manually mark the boundaries of a number of cells in these regions.

shoot meristem without validating against ground truth, it is not enough to evaluate the accuracy of this method on a typical 4D confocal data.

Besides the aforementioned approaches, tracking based on association between detections such as [9], [10] has shown good performance on time-lapse images. In [11], the authors proposed a cell tracking method on phase contrast time-lapse images that performs a global association of tracklets generated by frame-by-frame detection based tracking. However, these methods perform well when the feature quality or the underlying motion model is reliable. We are looking at a more challenging problem, where the features extracted from each cell may not be reliable enough for accurate data association. As an example, in this paper the experiments are performed on confocal time lapse image stacks of plant shoot apical meristem, where hundreds of cells are tightly clustered in a multi-layered architecture and only the boundary of each cell is visible. Thus the features extracted for each cell could only be the shape and area, which could often be non-discriminative between cells even from a local neighbourhood. Moreover, as in most confocal live-imaging datasets, the cells in the deeper central regions of the tissue have poor image quality because of the light absorption in the tissue, thereby making the local cell level features even less reliable for these cells. In Fig. 1, we have tried to bring out these challenges by showing a small spatio-temporal confocal substack of Arabidopsis SAM. Examples of stereotype in cell shapes are shown for four cells in a close neighborhood (marked by arrows) and the low SNR

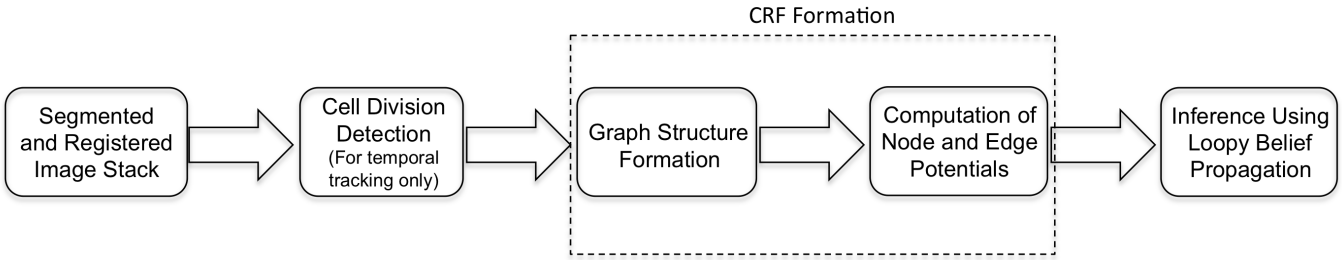

Fig. 2. Proposed cell tracking framework - different sequential components in the proposed method. The input to the method is a Watershed segmented and registered 3D or 4D image stack. For temporal tracking only, the next stage is detection of possible cell division events. The tracking is done sequentially on pairs of spatially or temporally consecutive slices. For any of such pairs, once the cell divisions are detected, we remove the parent and children cells from the respective segmented images and build a graph on one of the images of the pair based on neighborhood structure around each cell with individual cells as nodes in the graph. The candidate matches for each cell is found from the other image in the pair under consideration (for details, see Sec. III-A). The graph is then represented as an CRF. The node and edge potentials are computed using methods described in Sec. III-E and Sec. III-F and finally inference on this CRF is drawn using loopy belief propagation. The inferred state for each node (cell) gives us the correct correspondence between these pairs of images.

in the central regions of the tissue is highlighted by zooming into these regions in the 4^{th} column of the figure.

In such cases, for tracking and data associations in absence of very reliable features one can use the states of objects or points other than the target, that have strong spatio-temporal correlations with the states of the target. These correlations are utilized to rectify/estimate the target states in absence of reliable measurements for the target. Such secondary information are often termed as the 'contextual information' in the visual tracking literature and have resulted in significant improvements in tracking accuracy. For example, in [12], feature points from the scene that are not on the target but have strong motion correlation with the target are used to estimate the target state under occlusion.

In [13], [14], a spatio-temporal tracking algorithm for Arabidopsis SAM was proposed, where relative positional information of neighbouring cells was used to generate unique features for each cell. The best cell pair in two different image slices across space or time is found based on the computed features and the correspondence is grown sequentially outwards from these 'seed cell points' using a local search mechanism to find match between the rest of the cells. However, the location of this spatial search window depends on the position of the last tracked cell and hence this method tends to accumulate error that can throw the tracker off for cells spatially distant from the 'seed'.

In this work, we propose to solve the spatio-temporal tracking problem as a graph inference problem. To track cells between two image slices consecutive in time or space, we build a graph on one of the images with individual cells as the nodes and neighboring nodes sharing an undirected edge between them. We further define a Conditional Random Field (CRF) on the graph, the probable states of each node being the candidate cell correspondences from the next image. A distance defined on the physical features extracted from a cell and that of each of its candidate matches is used to constitute the node potential. The spatial context is modeled on each of the edges based on the relative location of the cell and its neighbors by utilizing the tight spatial topology of the cell clusters. We obtain the correspondences by maximizing the marginal distribution computed at each node (cell). The approximate marginals are obtained by a Loopy Belief Propagation scheme. The overall tracking pipeline is shown in Fig. 2.

B. Organization

The rest of the paper is organized as follows. An overview of the method is given in Sec. II. The mathematical and algorithmic details of the different components of the proposed method are provided in Sec. III. The experimental results are presented in Sec. IV followed by concluding discussion and future research directions in Sec. V.

II. Overview of The Proposed Method

As mentioned earlier, many animal and plant tissues (such as the shoot meristem of a plant, epithelial tissues in animals) are a collection of tightly packed small cells arranged in clonally different layers forming a solid 3D structure. To visualize the internal parts of these 3D structures we employ imaging techniques such as Confocal Laser Scanning Microscopy (CLSM) that generates serial optical cross sections of the tissue at various focal planes, thereby generating a 3D stack of images, each containing tightly packed 2D cross sections of 3D cells. In case of a time-lapse 'live cell imaging', the same tissue is imaged at successive time points resulting in a collection of a number of such 3D stacks. 2D segmentation techniques (such as Watershed) are employed to segment out individual 2D cell cross sections on each of the confocal slices. The problem of finding correspondences between such 2D cell slices along the depth of the tissue is called 'spatial tracking', analogous to the 'temporal tracking' problem where such correspondences are estimated between slices of the same cell at successive observational time points.

The 2D slices of the cells in the tissue are already registered in one 3D stack. We can also register cell slices across time between any two image slices (Fig. 1) of the tissue. This ability to register cell slices across space and time, along with the fact that the relative positions of the centroids of two neighboring cells in the tissue does not vary substantially across both time

and space motivate us to pose the problems of spatio-temporal cell tracking as a graphical inference problem.

A. Graph Structure

As can be seen in Fig. 3, a graph can be built on top of every slice image in the tissue. The nodes of the graph would be the cells (or more precisely, the centres of the cells) and each of the immediately neighboring cells would share a link/edge between them. In spatial tracking, each of these cells can either have a correspondence to one of the cell slices in the next z-slice or they can have no correspondence - in case the cell ends in the present image slice and not imaged in the deeper slice. Also, in case of temporal tracking, a cell might be out of Field-of-View (FoV) or not detected because of noise in the image. Thus, a candidate set of cell slices from the subsequent slice can be estimated for each cell in the image slice on which the current graph is built and this candidate set can be considered as the set of all possible states/labels for a certain node in the graph. An additional state, corresponding to the case that the cell is not imaged in the next confocal plane or next time point needs to be included in this set. Additionally, for temporal tracking, we first detect the cell division events across the two images (Fig. 4) and then build the graph with the rest of the cell slices in the first image as the nodes. The details on how the graph is formed and the set of states/labels for each node is ascertained are given in Sec. III-A.

B. Computation of Potential Functions

As we have mentioned earlier, the relative positions of the centroids of the neighbouring cell slices do not vary a lot in short time intervals or along z and hence the knowledge of the most probable state for any cell can substantially aid in estimating the maximum likely state for its neighbouring cells. We consider the graph just formed as a 'Conditional Random Field' (CRF) and draw inference from the joint distribution of the states of all the nodes by maximizing a potential function defined over all the nodes and edges (see Sec. III-E and Sec. III-F) of the graph. The node potentials in the CRF are computed based on the shape similarity between a cell slice and each of its candidates along with their relative centroid locations. The edge potentials are obtained based on similarity between the relative positions of two neighbouring nodes and that of their any two candidate cells' centroids in the successive slice or time point. The details of the edge and node potential computations are described in Sec. III-E and Sec. III-F.

C. Graph Inference

Once the graph is formed and the necessary potential functions computed, the next step is to design an inference strategy on this graph to estimate the maximum likely states for each of the nodes.. We employ a 'Loopy Belief Propagation' (LBP) based on the well known 'Sum-Product' algorithm [15] for this purpose. In Sec. III-G, we show the iterative parallel inter-node message updation strategy in the traditional sum-product scheme and perform the inference.

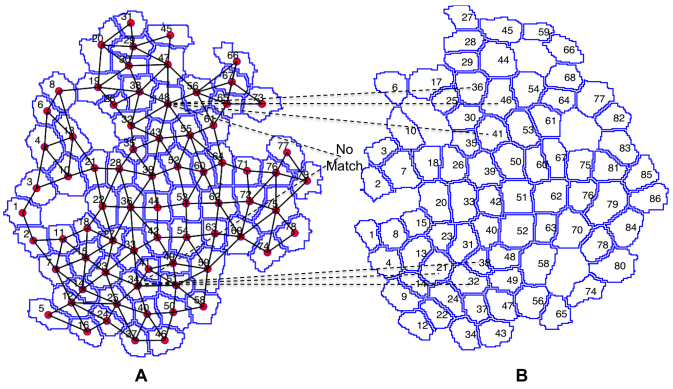

Fig. 3. Graph Structure. (A) For tracking cells between two spatially and temporally consecutive image slices, a graph is built on one of the images, where the nodes of the graph are the rest of the segmented cells and two neighbouring cells share an edge between them. For temporal tracking, the cells undergoing division are aet aside before constructing the graph. (B) From the next image slice, the candidate matches for each cell in A are estimated. Again, for temporal tracking, the children cells after division are also removed from the image and the candidate set of best 'K' states for each node in A is estimated through a search in B in a spatial window around the location of each of the nodes in A. A 'K+1'th state is added to each of the candidate sets corresponding to the case that the cell is not imaged or poorly imaged in B, referred to as the 'No Match' state in the figure. Now the graph is expressed as a CRF, where the node potentials are computed based on feature distances between each node and its candidates (see Sec. III-E)and the edge potentials are computed based on the relative locations of the neighbouring nodes in A and the same between any two cells in B from within their respective candidate states in B (Sec. III-F).

III. Graphical Model Design and Inference

A. Graph Formation on 2D Segmentations

Let us define the problem to be to find correspondences between the cells in two segmented confocal image slices I_G and I_M. The Watershed segmentation of I_G and I_M produces two sets of cell segments Ω_G and Ω_M respectively. Thus, the set of observations is given as

$$\mathcal{O} \;=\; \Omega_G \cup \Omega_M, \tag{1}$$

which comprises of 2D Watershed segmentations of both I_G and I_M. However, for temporal tracking, we first detect if some cells form I_G have divided into pairs of cells in I_M following the method described in III-C and remove the parent cells that has undergone division from Ω_G and the divided children from Ω_M. The graph and the candidate states of each node of the graph are thereafter formed using the remaining subsets of cells V_G and V_M containing N_G and N_M cells respectively, i.e. the remaining cells

$$\begin{aligned} v_G^1, v_G^2, \cdots v_G^{N_G} \;&\in\; V_G \subseteq \Omega_G \\ v_M^1, v_M^2, \cdots v_M^{N_M} \;&\in\; V_M \subseteq \Omega_M \end{aligned} \tag{2}$$

The graph is built on I_G and the set of nodes V_G is same as the set of segmented cells. Any two nodes v_G^i and v_G^j will have an edge between them if v_G^i and v_G^j are spatial neighbours. For tightly packed cluster of cells, v_G^i and v_G^j are neighbours if they share a common boundary and thus the set of all neighbours of a cell v_G^i would be

$$N(v_G^i) \;=\; \{v_G^j \; s.t. \; v_G^i \text{ and } v_G^j \text{ share common boundary}\}. \tag{3}$$

For other datasets, where the cells are generally not compactly arranged, this set can be represented as

$$N(v_G^i) \;=\; \{v_G^j \; s.t. \; ||\mathbf{c}_G^i \; - \mathbf{c}_G^j||_2 \leq th\}, \tag{4}$$

where c_G^i and c_G^j are the centroids of v_G^i and v_G^j respectively.

Thus, we can represent the graph g_G on I_G as an adjacency matrix A_G between the nodes,

$$\begin{aligned} A_G\left(i,j\right) &= 1 \text{ iff } v_G^j \in N\left(v_G^i\right), \\ &= 0, \text{ otherwise} \end{aligned} \tag{5}$$

B. Determination of Candidate States For Every Node

For finding correspondences between cells across two segmented slices I_G and I_M, the graph is built on the slice I_G following Sec. III-A. Each node in the graph, corresponding to each cell slice v_G^i represents a random variable x^i that can take a label from the set S_G^i which is the set of K closest segments in the slice I_M around the point $\mathbf{c}_G^i$, the centroid of v_G^i on I_G. Therefore,

$$S_G^i \;=\; \{s_1^i, s_2^i, \cdots s_K^i\} \tag{6}$$

where, $s_k^i \in V_M \;\forall k = 1, 2, \cdots K$ and

$$\begin{aligned} ||\mathbf{c}_M^{s_1^i} - \mathbf{c}_G^i|| \leq ||\mathbf{c}_M^{s_2^i} - \mathbf{c}_G^i|| \cdots ||\mathbf{c}_M^{s_K^i} - \mathbf{c}_G^i|| \\ \leq ||\mathbf{c}_M^{s_j^i} - \mathbf{c}_G^i|| \;\forall\; j \in \{1, 2, \cdots N_M\},\; j \notin S_G^i \end{aligned} \tag{7}$$

We can safely assume that the actual tracked cell slice in I_M would be amongst the K closest cells, as I_G and I_M are already registered.

Now, we add an additional label s_0^i to the candidate set S_G^i that represents the case where the cell slice v_G^i is not imaged in the slice I_M. Thus, the complete set of candidate states becomes

$$S_G^i \;=\; \{s_0^i, s_1^i, \cdots s_K^i\}. \tag{8}$$

C. Cell Division Detection

To detect cell divisions before forming the graph g_G in temporal tracking, we first compute the candidate sets C_G^i in I_M for a segmented cell slice $\omega_G^i \in \Omega_G$ following similar method as in Eqn. 7. Next we form all possible pairs of the candidate cells from C_G^i that share a boundary as in

$$D_G^i \;=\; \left\{(cd_p^i,\; cd_q^i) \; s.t. \; cd_p^i \in N(cd_q^i) \text{ and } cd_p^i, cd_q^i \in C_G^i\right\}. \tag{9}$$

Now, if the cell ω_G^i has divided into two children cells cd_p^i and cd_q^i, then ideally the shape of ω_G^i should be very similar to the combined shape of cd_p^i and cd_q^i, taken together (i.e. to the shape of $cd_p^i \cup cd_q^i$) and each of cd_p^i and cd_q^i would be approximately half the size of ω_G^i. Motivated by this physical property associated with cell division, we compute a Modified Hausdorff Distance (MHD) metric to estimate the shape similarity between $b(\omega_G^i)$ and $b(cd_p^i \cup cd_q^i)$, where b is the set of boundary points on a shape, when the point coordinates

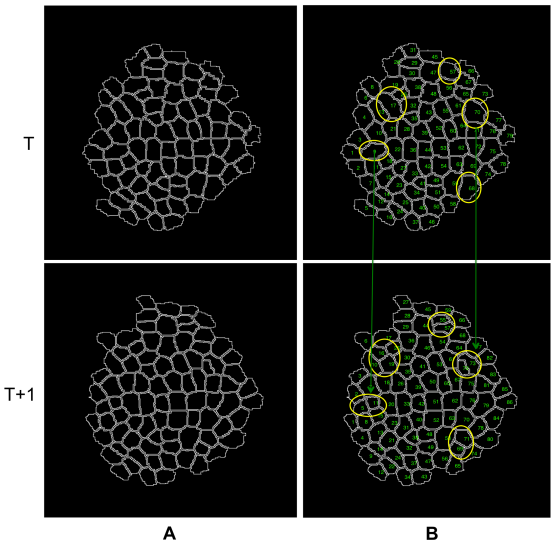

Fig. 4. Cell Division Detection. (A) Two segmented image slices one time point apart. (B) The ellipses in image at T mark the parent cell that have undergone divisions between time points T and T+1 and those at time point T+1 mark the children cells after division.

are recomputed with respect to the shape centroid. With these, we compute a set of distances as

$$\begin{aligned} d(\omega_G^i,\; D_G^i) \;=\; \frac{1}{t_1} MHD(b(\omega_G^i), b(cd_p^i \cup cd_q^i)) \;+\; \frac{1}{t_2}\left[\left|\frac{1}{2}\right.\right. \\ \left.\left. - \frac{area(cd_p^i)}{area(\omega_G^i)}\right| \;+\; \left|\frac{1}{2} - \frac{area(cd_q^i)}{area(\omega_G^i)}\right|\right] . \end{aligned} \tag{10}$$

If $\min d(\omega_G^i,\; D_G^i) \;\leq\; 1$, then it is inferred that the cell ω_G^i has divided into a pair of cells (cd_p^i, cd_q^i) for which this minimum is obtained. The values of the parameters t_1 and t_2 are learnt from a small training image set and the details of parameter learning is described in Sec. IV-D.

Once the cell division events are detected for one or more cells in I_G, the graph g_G is constructed using the methods described in III-A and III-B after eliminating the parents undergoing division and the divided children cells from Ω_G and Ω_M respectively and forming V_G and V_M.

D. Conditional Random Field Modeling

Let the set of random variables associated with v_G^i be $X \;=\; \{x^1, x^2, \cdots x^{N_G}\}$, which are to be estimated given the observation I_M. These random variables correspond to the state of each node in the graph and the support for each of these variables is the candidate set as discussed in Sec. III-B.

Then the overall CRF is expressed as

$$\begin{aligned} P(X;\mathcal{O}) &= \exp\left(-E(X;\mathcal{O})\right)/Z \\ &= \exp\left\{-\sum_{c \in clq(X)} E_c(X;\mathcal{O})\right\}/Z, \end{aligned} \tag{11}$$

where Z is the partition function and E is the energy function defined on all the cliques of the graph, which can be further split into individual nodes and edges as

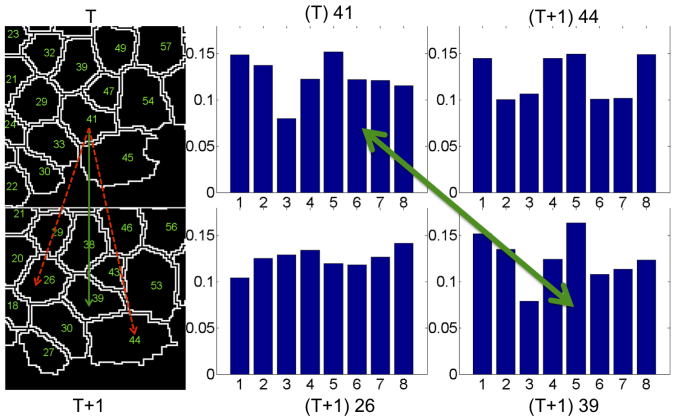

Fig. 5. Shape descriptor for individual cells. Cell 41 at time point T has cells 26, 39 and 44 as candidates for correspondence at time T+1. The correct correspondence for 41 at T is 39 at T+1. In the left column of figure, the correct correspondence is shown in green arrow whereas the the incorrect ones are shown in red. Shape histogram descriptors are computed for each cell following the method described in Sec. III-E. As expected, the histogram for 41 at T is very similar to that of 39 at T+1 and the descriptors for the other two candidate cells are very different.

$$E(X;\mathcal{O}) = \sum_{i=1}^{N_G} E_i\left(x_i;\mathcal{O}\right) + \sum_{i=1}^{N_G} \sum_{j:v_G^j \in N(v_G^i)} E_{ij}(x_i,x_j;\mathcal{O}). \tag{12}$$

Then,

$$\begin{aligned} P(X;\mathcal{O}) &= \frac{1}{Z}\prod_{i=1}^{N_G} \exp\left(-E_i(x_i;\mathcal{O})\right) \cdot \\ &\quad \prod_{\substack{(i,j) \\ : v_G^j \in N(v_G^i)}} \exp\left(-E_{i,j}(x_i,x_j;\mathcal{O})\right) \\ &= \frac{1}{Z}\prod_{i=1}^{N_G} \phi_i(x_i;\mathcal{O}) \cdot \\ &\quad \prod_{\substack{(i,j) \\ : v_G^j \in N(v_G^i)}} \psi_{i,j}(x_i,x_j;\mathcal{O}) \end{aligned} \tag{13}$$

Here ϕ_i represents the node potential of any node v_G^i in g_G, and ψ_{ij} is the edge potential from node v_G^i to node v_G^j. To estimate the optimal states for every node, we have to maximize $P(X;\mathcal{O})$. Towards that objective, we first estimate the approximate marginal distributions $P(x_i;\mathcal{O})$ at each node using belief-propagation scheme as described later. The optimal states that maximize the posterior distribution could be then estimated by maximizing the marginals independently.

E. Computation of Observation/Node Potential:

The node potential is defined on every node of the graph, which is the likelihood on the label taken by a node belonging to V_G, given the observation $\mathcal{O}$. It is analogous to the probability distribution of any node v_G^i being assigned to each of its candidate states. This distribution is computed independently for each node based on its shape similarities and proximities in location of its centroid from each of its candidates.

For measuring similarities between cell shapes, we generate a shape histogram descriptor for each of the cells, which is very similar to one of the methods described in [16]. First we recompute co-ordinates of a cell's peripheral points by shifting the origin to the cell's centroid. Next, we partition the x-y plane into 8 angular sectors centred at the origin and compute the mean Euclidean distances of the peripheral points falling into each of these partitions from the origin. The set of this distances forms a 8 bin histogram descriptor for the shape of a cell, the angular sectors being sorted counter clockwise from x-axis. Note that, unlike the classical shape histograms of this sort [16], we compute mean distances from each sector instead of counting the number of points, as the latter gives us scale invariance and may lead to a high match score between a legitimate cell and a small region generated by over-segmentation on noisy images. Some sample descriptors of a cell and its candidates for correspondence are given in Fig. 5.

Let the shape histogram associated with the cell slice v_G^i be h_G^i and that with the candidate slice s_j^i be h_M^j (as $s_j^i \in V_M$). We computed the K-L divergence (KLD) between h_G^i and h_M^j which gives us a distance measure between these two cell slices and suppose it is represented as $d_1^i\left(v_G^i, s_j^i\right)$,

$$d_1^i\left(v_G^i,\ s_j^i\right) = \text{KLD}\left(h_G^i,\ h_M^j\right). \tag{14}$$

We also compute the distances between the centroids of a cell slice in I_G and each of its candidates in I_M and the distance is given by,

$$d_2^i\left(v_G^i,\ s_j^i\right) = \|\mathbf{c}_M^{s_j^i} - \mathbf{c}_G^i\|_2 . \tag{15}$$

Hence, the overall distance between a cell slice v_G^i and one of its candidates $v_M^{s_j^i}$ is expressed as a combination of normalized d_1 and d_2 as

$$d^i\left(v_G^i,\ s_j^i\right) = w\frac{d_1^i}{\lambda_1} + (1-w)\frac{d_2^i}{\lambda_2},\quad 0 \le w \le 1 . \tag{16}$$

The corresponding node potential for each node is

$$\phi_i\left(x_i = s_j^i;\mathcal{O}\right) = \exp\left(-d^i(v_G^i,\ s_j^i)\right)\ \forall j = 1,2\cdots K \tag{17}$$

and

$$\phi_i\left(x_i = s_0^i;\mathcal{O}\right) = 1 - \max_j \left\{\phi_i\left(x_i = s_j^i;\mathcal{O}\right),\ j = 1,2,\cdots K\right\} \tag{18}$$

The normalization parameters λ_1 and λ_2 (in Eqn. 16) are learnt from a training dataset. See Sec. IV-D for details on parameter estimation.

F. Computation of Spatial Context/Edge Potential:

This potential function is defined on edges connecting pairs of neighbouring nodes and is representative of the conditional distribution $P(x_j|x_i,\mathcal{O})$. The computation of the potential function depends on the fact that if two neighbouring cells v_G^i and v_G^j are tracked to two cell slices v_M^p and v_M^q, then the relative position of v_G^j with respect to v_G^i should be very similar to that of v_M^q and v_M^p. As a result, if v_G^i is tracked

to v_M^p then the probability that v_G^j corresponds to v_M^q gets boosted if

$$\mathbf{c}_G^j - \mathbf{c}_G^i \;\approx\; \mathbf{c}_M^q - \mathbf{c}_M^p \;, \tag{19}$$

where $\mathbf{c}_G^i, \mathbf{c}_G^j, \mathbf{c}_M^p, \mathbf{c}_M^q$ be the centroids of $v_G^i, v_G^j, v_M^p, v_M^q$ respectively.

Clearly, the additional evidences for matching two cell slices in I_G and I_M comes in the form of local neighbourhood structure based contextual information.

Thus, the contextual transition potentials between any two nodes v_G^i and v_G^j taking non-zero states can be expressed as a function of the shift between the relative positions of those nodes

$$\begin{aligned}&\psi_{i,j}\left(x_i = s_p^i, x_j = s_q^j; \mathcal{O}\right)\\ &\quad= \exp\left\{-\gamma\|(\mathbf{c}_G^j - \mathbf{c}_G^i) - (\mathbf{c}_M^{s_q^j} - \mathbf{c}_M^{s_p^i})\|_2\right\}\end{aligned} \tag{20}$$

$\forall p, q = 1, 2, \cdots K$, where $s_p^i \in S_G^i, s_q^j \in S_G^j$ and $i, j \neq 0$.

Now, in both spatial and temporal tracking, there is one more state s_0^i for every node i that corresponds to the case that the particular cell is not imaged in the successive slice (spatial or temporal). Thus, the transition potentials must also incorporate the case where one of the cells is not tracked and its neighbouring cell is matched to one of the cells in the next slice or not matched to any cell and vice versa. Incorporating these values, the complete edge potential function between any two neighbouring nodes v_G^i and v_G^j would be

$$\psi_{i,j}\left(x_i = s_0^i,\; x_j = s_q^j; \mathcal{O}\right) \;=\; \frac{1}{K+1}\;\forall q = 0, 1, \cdots K\;. \tag{21}$$

This corresponds to the case when v_G^i is not matched to any cell in I_M. When both the cells v_G^i and v_G^j have correspondences in the subsequent spatial or temporal image I_M,

$$\begin{aligned}&\psi_{i,j}\left(x_i = s_p^i,\; x_j = s_q^j; \mathcal{O}\right)\\ &\;=\; \exp\left\{-\gamma\|(\mathbf{c}_G^j - \mathbf{c}_G^i) \;-\; (\mathbf{c}_M^{s_q^j} - \mathbf{c}_M^{s_p^i})\|_2\right\}\;,\end{aligned} \tag{22}$$

for $p, q \neq 0$.

Finally, when v_G^i has a match in the next spatial or temporal image slice I_M, but its neighbour v_G^j does not, then the corresponding edge potential entries become

$$\begin{aligned}&\psi_{i,j}\left(x_i = s_p^i,\; x_j = s_0^j; \mathcal{O}\right)\\ &= 1 - \max_q\left\{\psi_{i,j}\left(x_i = s_p^i, x_j = s_q^j; \mathcal{O}\right),\; q = 1, 2, \cdots K\right\}\end{aligned} \tag{23}$$

for $p \neq 0$.

G. Inference: Loopy Belief Propagation

The next step is to do the inference on the CRF, which involves the computation of the marginal probability distributions for the states x_i of each node $v_G^i \in V_G$, given the observations $\mathcal{O}$. For computation of the marginals at each node, we choose to use a very popular local *message-passing* algorithm known as *Belief Propagation* (BP) [17]. Since there are many loops or cycles in our graph, the algorithm is called a *Loopy Belief Propagation* (LBP). This is an iterative algorithm and at l^{th} iteration, each node v_G^i computes a message to be sent to each of its neighbours and the message sent to $v_G^j \in N(v_G^i)$, according to the popular *Sum-Product* algorithm [15], is,

$$\begin{aligned}m_{i,j}^{(l)}(x_j) = \alpha \sum_{x_i}\Bigg\{&\psi_{i,j}(x_i, x_j; \mathcal{O})\phi_i(x_i;\\ &\mathcal{O}) \prod_{x_k : v_G^k \in N(v_G^i)\setminus v_G^j} m_{k,i}^{(l-1)}(x_i)\Bigg\}\end{aligned}$$

where α is a normalizing constant. Note that the updation strategy employed here is parallel, i.e. all the edges in the CRF are updated simultaneously in each iteration.

Also, at each iteration l, each node v_G^i produces an approximate marginal distribution

$$P^{(l)}(x_i; \mathcal{O}) \;=\; \alpha\phi_i(x_i; \mathcal{O}) \prod_{x_j : v_G^j \in N(v_G^i)} m_{j,i}^{(l)}(x_i) \tag{24}$$

For a tree type graph, these approximate marginal distributions are guaranteed to converge to the true marginals, but for a graph as ours that contains multiple loops there is no guarantee of convergence of the LBP [18]. However, in literature, such as [19], LBP has shown very good empirical performance and in most of our experiments the method converged very quickly. If the convergence is reached at iteration L, the MAP estimates for the most likely states is computed for each marginal (i.e. for each node) as,

$$\hat{x_i} \;=\; \arg_{x_i} \max\; P^{(L)}(x_i; \mathcal{O}) \tag{25}$$

and this optimum state corresponds to either the 'no-match' case or a specific cell in I_M.

IV. Experimental Results

A. Data Collection and Preprocessing

For the experiments performed in the present study, the 3D structure of the tissues are imaged using single-photon confocal laser scanning microscope and we have specially dealt with the 'Shoot Apical Meristem' (SAM) of the plants that showcase all the challenges associated with any spatio-temporal cell tracking problem in a tightly packed multilayer tissue. The SAM of *Arabidopsis Thaliana* consists of approximately 500 cells and they are organized into multiple cell layers that are clonally distinct from one another. By changing the depth of the focal plane, CLSM can provide in-focus images from various depths of the specimen. To make the cells visible under laser, fluorescent dyes are used. The set of images, thus obtained at each time point, constitute a 3-D stack, also known as the 'Z-stack'. Each Z-stack is imaged at a certain time interval (e.g. 3-6 hours between successive observations) and it is comprised of a series of optical cross sections of SAMs that are separated by approx. 1.5-2 μm. A standard shoot apical meristematic cell has a diameter of about 5 - 6 μm and hence in most cases, a single cell is not visible in more than 3-4 slices when the tissue is sparsely imaged at the

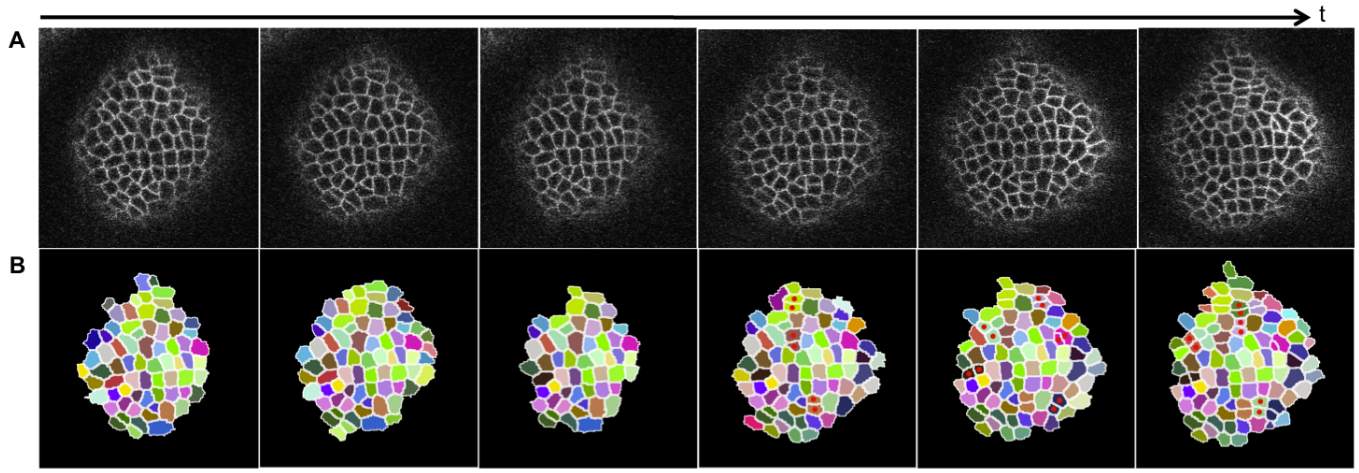

Fig. 6. Results on the temporal tracking on Arabidopsis SAM live imaging dataset with time resolution of 3 hours. (A) Raw confocal image slices at 3 μm deep into the tissue imaged every 3 hours from 3^{rd} hour of observation to 18^{th} hour. (B) Temporal tracking result shown by color coding the cells. The same cells are marked with the same color. After cell division, the children cells are marked with the same color as their parent, also a red dot is put at the center of each of the children.

aforementioned z-resolution to avoid photodynamic damage to the cells.

Each 2D image slice in the 4D confocal image stack is further segmented into individual cell slices. The choice of the 2D segmentation algorithm is largely data-specific. For our experiments on the SAM tissues, we use an adaptive Watershed segmentation method [20] that learns the 'h-minima' threshold directly from the image data so that a uniformity in cell sizes is maintained as a result of the segmentation. This method works satisfactorily for SAM cells as, in general, all SAM cells on a 2D confocal slice have similar sizes. This 2D segmentation method is also robust to over and under-segmentation errors to a large extent.

The image slices in one single 3D confocal stack is already registered because of minimal movement of the tissue specimen during imaging at any given time point of observation. However, during successive observations the specimen is moved in and out of the imaging setup which causes rotation and shift of the imaged 3D stack from that at the previous time point. Thus, the image slices in successive time points have to be registered prior to the cell-tracking. The dataset-1 was given to us preregistered and for experimental dataset-2, we register the cell slices across time between any two confocal images of the tissue using our 'local graph' based registration technique [21]. This is a fully automated landmark based registration method that finds out correspondences between the two image slices and utilizes these correspondences (landmarks) to register one image to the other.

B. Tracking Results and Analysis

We have tested our proposed cell tracking method on two 4D confocal stacks of Arabidopsis SAM. The details of CLSM imaging for generating the raw data is described in Sec. IV-A. The first dataset contains 3D stacks of Images observed every 3 hours and in the second dataset, the 3D image-stacks are taken every 6 hours. In both the datasets, the z-resolution in each 3D stack is 1.5μm.

For spatial tracking across a 3D stack or to track cells in one SAM slice observed at multiple time points, we apply our tracker sequentially, as our tracker finds correspondences between cells from one slice to another. Thus to generate cell lineages between two time points 't' and 't+4', we run our tracker 4 times on successive pairs of images ((t, t+1), (t+1, t+2), (t+2, t+3), (t+3, t+4)).

1) Temporal Tracking on Dataset 1: Fig. 6 shows a typically obtained result for temporal tracking in the Dataset 1. Fig. 6A shows raw confocal image slices at a depth of 3 μm from the tip of SAM through 6 consecutive time points (3^{rd} to 18^{th} hours). Although the images are registered, because of the growth of the cells in the tissue there are local shifts in the cells' positions, which makes the task of temporal tracking more challenging. The segmentation and tracking results for these slices are shown in Fig. 6B, where the slices of one cell across different time points are marked with the same color. We have also marked the 12 cell division events detected by the tracker on the same images. The children cells are marked with red dots and they share the same color with their parent cell. The result portrays the typical high value of accuracy we obtain through our tracker to generate temporal cell lineages.

2) Combined Spatio-Temporal Tracking On Dataset-1: We perform spatial cell tracking across the depth of the 3D confocal image stacks and combine them with temporal tracking of the same cells across time in Fig. 7. We sample three consecutive spatial slices from confocal stacks at 4 different time points (at $12^{th}, 15^{th}, 18^{th}$ and 21^{st} hours of observation) and the tracking result for them are shown. The 2D slices coming from the same 3D cell are correctly tracked for all the cells across 4 different time instants and are marked with the same color. It can be observed that slices of new cells appear as we go deeper into the tissue and as expected, they are not matched to any cell from the slice above. Again, because of the growth in tissue over time, some new cells become visible in the chosen focal planes and the tracks are initialized. In such cases, these cell slices or tracks are initialized with a

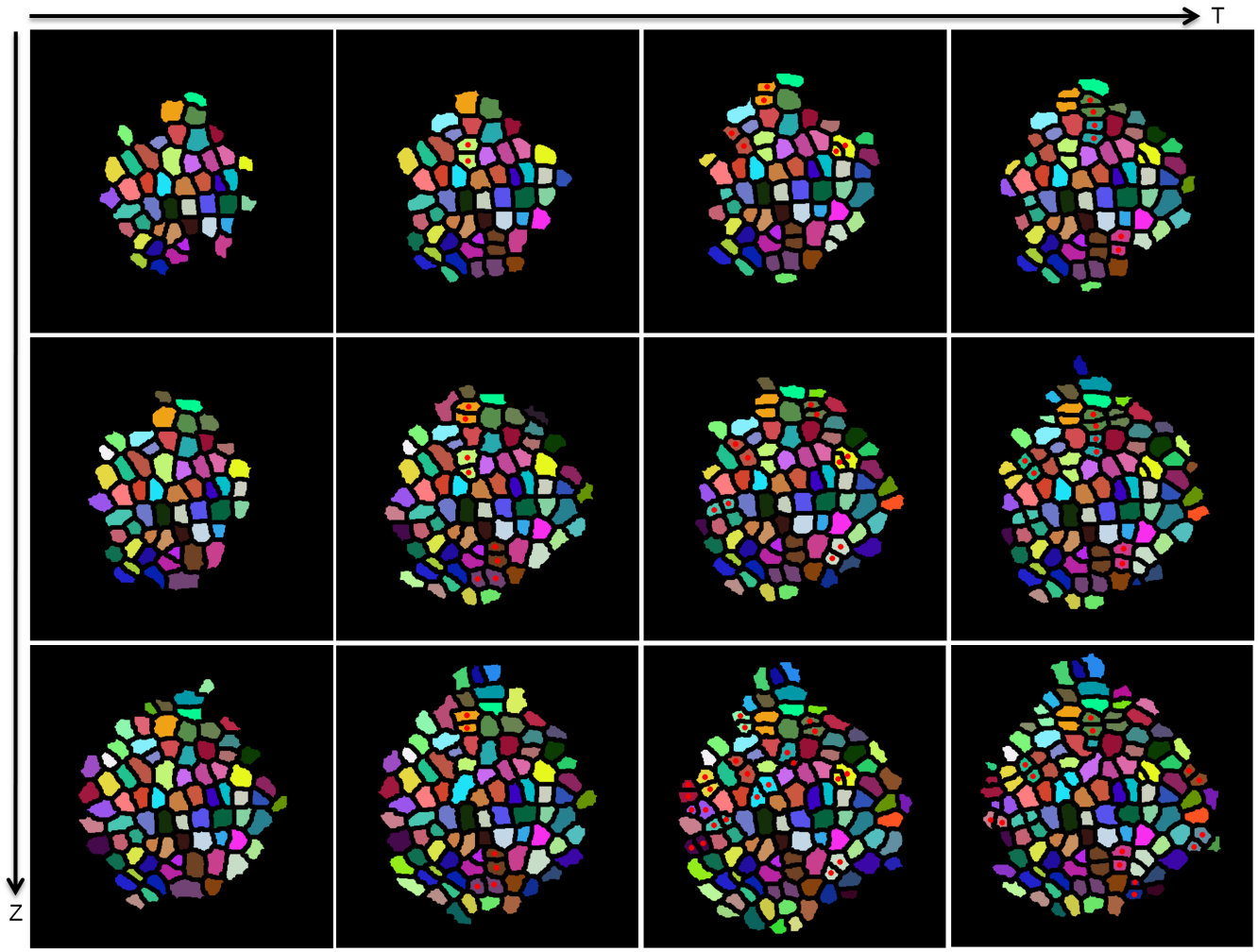

Fig. 7. Results showing combined spatio-temporal tracking on Arabidopsis SAM dataset 1. A number of cells are tracked across four time points of observation ($12^{th}, 15^{th}, 18^{th}$ and 21^{st} hours). Three image slices are sampled from the 3D stack at each time point (at 3 μm, 4.5 μm and 6 μm respectively). Cell slices corresponding to the same cell across space and time are marked with the same color. Cell divisions are also detected and the children cells having the same colors as their parents are marked with red dots.

new label (a random color as in this figure) and each of their correspondences are searched in the next deeper slice and in the next observational time points .

The complete tracking result on dataset 1 (7 slices and 12 time points) is summarized quantitatively in Table I. We split the results in four different classes, True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). TP corresponds to the cases where two cell slices are correctly matched either in space or time. When cell slices from two different cells are incorrectly matched together, it falls under FP. When the tracker fails to pick up a correct correspondence, it is represented by FN and its opposite case is tabulated under TN. From Table I, it can be seen that for spatial tracking, the correctly tracked cases (TP and TN combined) is as high as 98% whereas FP and FN cases are merely around 2%. Similar accuracy is observed for temporal tracking too, where the accuracy of the tracker is more than 97.5%. The tracker can successfully detect cell divisions as out of 33 cell division events in 36 hours, 31 events ere correctly picked up by the tracker and there is no False Positive.

TABLE I
TRACKING RESULT SUMMARY: DATASET 1

	TP	FP	TN	FN
Spatial	86%	0.25%	12.13%	1.62%
Temporal	83%	0%	14.66%	2.34%
Division (L1)	31/33	0	-	2/33

3) Tracking With Larger Time Gaps - Dataset 2: To evaluate the accuracy of the tracker in the situations where the temporal resolution is small, i.e., the 3D stacks are imaged after large time gaps we tested the proposed tracker on a second dataset. In this dataset (dataset 2), the imaging is done every 6 hours (compared to 3 hours for dataset 1). With a longer gap between observations, the deformation of the cells in the tissue is even more visible, which results in larger shifts between centroid locations of the same cell in successive time points. Moreover, there are more number of cell division

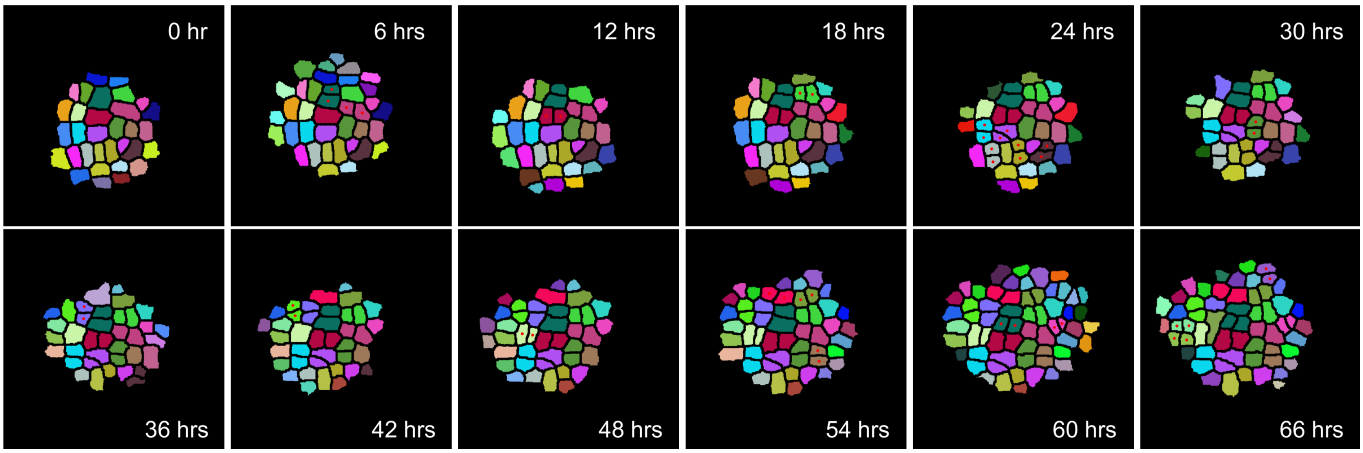

Fig. 8. Results on the temporal tracking on Arabidopsis SAM live imaging dataset where the images are collected through three days with time resolution of 6 hours between successive observations. (A) Raw confocal image slices from L1 layer of the tissue imaged every 6 hours from 0^{th} hour of observation to 66^{th} hour. (B) Temporal tracking result shown by color coding the cells. The same cells are marked with the same color. The cell division events are also detected with perfect accuracy and are displayed on this figure the same way as Fig. 6.

events which makes the temporal tracking problem even more challenging.

Because of the robustness of the proposed method, we obtain highly accurate temporal tracking results on dataset-2 as seen in Fig. 8 with the same set of CRF parameters that was used for dataset-1. We also show that our method is capable of maintaining tracks for long durations (66 hours as shown in the figure) and it detected all legitimate cell division events. The results on spatial tracking are not presented here as each 3D stack is structurally and visually very similar to that of dataset 1.

C. Comparison of the Proposed Method with the State-of-the-Art

We have compared our proposed tracking method with the 'local-graph' based cell tracker [13] and also with a baseline tracker. In order to show the improvements in tracking accuracy using contextual information, we designed the baseline tracker on the same local cell shape features as used to compute the node potentials in Sec. III-E and the tracker associates cell slices across images using 'Hungarian algorithm'. Also, if any associated pair of cells have a feature distance larger than a predefined threshold, the track is terminated and re-initialized. Fig. 9A shows the tracking result by using this baseline tracker on four spatially sampled image slices from a 3D image stack. A number of wrong associations are marked by white arrows. Fig. 9B and Fig. 9C shows tracking results on the same images for [13] and the proposed method respectively. The baseline tracker generates many wrong associations because the cell shapes are often very similar even in a close neighborhood. The tracker proposed in [13] performs much better than the baseline tracker and the errors comprise of both false-positives and false-negatives along with a number of switched tracks. The proposed method, however, performs the best as the errors obtained are much fewer in numbers than both [13] and the baseline and therefore validates that the local contextual information indeed aids spatio-temporal cell tracking for tightly packed multilayer tissues.

D. Learning the CRF Parameters

We use manually ground-truthed correspondences in a small subset of the dataset-1 as our training set to learn the best set of values for different parameters and the same set of learnt parameter values are used for all the experimental results shown in this paper. The parameters used in the cell division detection method (see Sec. III-C) are learnt independently from the CRF parameters. For each of the parameters, we first choose a range for the parameter value and then uniformly sample a number of values from within that range. Now, we generate generate combined sets of parameter values using every possible combination. Finally, on the training dataset, the best set of parameters out of all such candidates is selected using a 5-fold cross validation.

E. Discussion on the Limitations of the Proposed Method

The accuracy of the proposed cell tracking method depends on spatio-temporal registration of the image slices in the 4D stack. The parameters of the tracker can be tuned in order to account for some error in registration but for major transformations across image slices, the tracker would not work satisfactorily. As we have shown in our experiments, the tracker can handle moderate deformations of the growing cells. However, if the deformation changes both the nominal shape of individual cells as well as the topology of their local neighbourhood, it becomes more challenging and in some cases leads to failure of the tracker. Likewise, this present tracking algorithm is not designed to handle large displacements or motions of individual cells, but it can still provide good tracking accuracy as long as the local neighbourhood structure around a cell is not jeopardized.

V. Conclusion and Future Work

We have presented a method for automatically tracking individual cells in closely packed developing multilayer tissues.

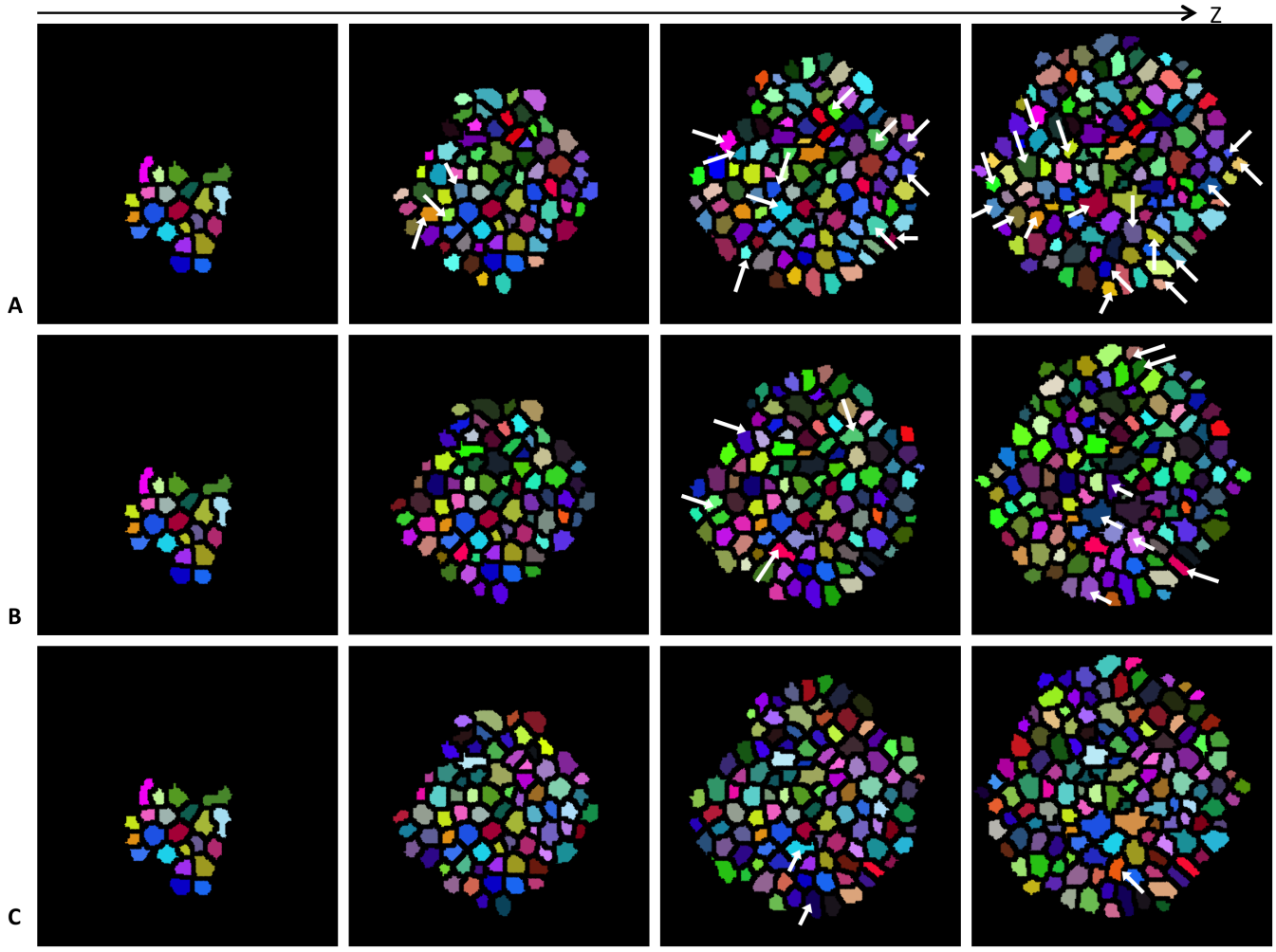

Fig. 9. Comparison of the spatial tracking results as obtained from the proposed method with the results from [13] and a baseline tracker. The results are shown on a set of four spatially sampled image slices from a 3D image stack of Arabidopsis SAM. The tracking results are shown using similar color-coding as in the previous figures and the locations of errors in tracking are marked by white arrows. (A) The results obtained by using the baseline tracker contain many errors as it is designed on local cell shape features and the cell shapes even from a close neighborhood can be very stereotypical. (B) Results obtained by using [13] are much better in accuracy but still contain a number of FP, FN and switched tracks. (C) The proposed method performs the best out of these three with very few errors and no track switching.

We observed that cells in a close cluster in the tissue can have very similar image features and hence we leveraged upon the local spatial geometric structure and topology of the relative positions of the neighbouring cells to robustly track growing cells in the tissue in presence of imaging noise. We have also shown how to detect cell divisions prior to temporal tracking in order to find out the proper terminating point of individual cell lineages. Experiments were conducted on two 4D confocal stacks Arabidopsis SAM having different temporal resolutions and the results indicate the high accuracy obtained through the proposed method for both spatial and temporal tracking.

Future work would include the integration of this spatio-temporal tracking method with our image analysis components such as segmentation [20], registration [21] and the cell resolution 3D reconstruction methods [22], [23] to design a complete 4D image analysis pipeline. This pipeline could be effective for generating cell division and cell growth statistics in a fully automated, high-throughput manner. These statistics can help us model the spatio-temporal interplay between cell growth and cell division in a complex multi-layered tissue.

Acknowledgement

We graciously acknowledge Dr. G.V. Reddy for providing us with the Arabidopsis datasets that have been used throughout the experiments.

References

[1] R. Fernandez, P. Das, V. Mirabet, E. Moscardi, J. Traas, J.-L. Verdeil, G. Malandain, and C. Godin, "Imaging plant growth in 4d: robust tissue reconstruction and lineaging at cell resolution," *Nature Methods*, vol. 7, no. 7, pp. 547–553, 07 2010.

[2] O. Dzyubachyk, W. Van Cappellen, J. Essers, W. Niessen, and E. Meijering, "Advanced level-set-based cell tracking in time-lapse fluorescence microscopy," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 3, pp. 852–867, 2010.

[3] K. Li and T. Kanade, "Cell population tracking and lineage construction using multiple-model dynamics filters and spatiotemporal optimization," in *Proceedings of the 2nd International Workshop on Microscopic Image Analysis with Applications in Biology*, 2007.

[4] K. Li, M. Chen, T. Kanade, E. Miller, L. Weiss, and P. Campbell, "Cell population tracking and lineage construction with spatiotemporal context," *Medical Image Analysis*, vol. 12, no. 5, pp. 546 – 566, 2008.

[5] D. R. Padfield, J. Rittscher, N. Thomas, and B. Roysam, "Spatio-temporal cell cycle phase analysis using level sets and fast marching methods." *Medical Image Analysis*, vol. 13, no. 1, pp. 143–155, 2009.

[6] A. Dufour, V. Shinin, S. Tajbakhsh, N. Guillen-Aghion, J. C. Olivo-Marin, and C. Zimmer, "Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1396–1410, 2005.

[7] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching," in *in CVPR*, 2000, pp. 44–51.

[8] V. Gor, M. Elowitz, T. Bacarian, and E. Mjolsness, "Tracking cell signals in fluorescent images," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 0, p. 142, 2005.

[9] N. N. Kachouie, P. Fieguth, J. Ramunas, and E. Jervis, "Probabilistic model-based cell tracking," *International Journal of Biomedical Imaging*, 2006.

[10] T. Kirubarajan, Y. Bar-Shalom, and K. R. Pattipati, "Multiassignment for tracking a large number of overlapping objects," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 37, no. 1, pp. 2–21, 2001.

[11] R. Bise, Z. Yin, and T. Kanade, "Reliable cell tracking by global data association," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 1004–1010.

[12] H. Grabner, J. Matas, L. J. V. Gool, and P. C. Cattin, "Tracking the invisible: Learning where the object might be," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010*, 2010, pp. 1285–1292.

[13] M. Liu, R. K. Yadav, A. Roy-Chowdhury, and G. V. Reddy, "Automated tracking of stem cell lineages of arabidopsis shoot apex using local graph matching," *Plant journal, Oxford, UK*, vol. 62, pp. 135–147, 2010.

[14] M. Liu, A. Chakraborty, D. Singh, R. K. Yadav, G. Meenakshisundaram, G. V. Reddy, and A. Roy-Chowdhury, "Adaptive cell segmentation and tracking for volumetric confocal microscopy images of a developing plant meristem," *Molecular Plant*, vol. 4, no. 5, pp. 922–31, 2011.

[15] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, pp. 498–519, 1998.

[16] M. Ankerst, G. Kastenmüller, H.-P. Kriegel, and T. Seidl, "3d shape histograms for similarity search and classification in spatial databases," in *Proceedings of the 6th International Symposium on Advances in Spatial Databases*, 1999, pp. 207–226.

[17] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., 1988.

[18] C. Sutton and A. McCallum, "An introduction to conditional random fields," *Foundations and Trends in Machine Learning*, vol. 4, no. 4, pp. 267–373, 2012.

[19] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *In Proceedings of Uncertainty in AI*, 1999, pp. 467–475.

[20] K. Mkrtchyan, D. Singh, M. Liu, G. V. Reddy, A. K. R. Chowdhury, and M. Gopi, "Efficient cell segmentation and tracking of developing plant meristem," in *IEEE International Conference on Image Processing*, 2011, pp. 2165–2168.

[21] K. Mkrtchyan, A. Chakraborty, and A. K. Roy-Chowdhury, "Automated registration of live imaging stacks of arabidopsis," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013, pp. 672–675.

[22] A. Chakraborty, M. M. Perales, G. V. Reddy, and A. K. Roy-Chowdhury, "Adaptive geometric tessellation for 3d reconstruction of anisotropically developing cells in multilayer tissues from sparse volumetric microscopy images," *PLoS ONE*, vol. 8, no. 8, p. e67202, 08 2013.

[23] A. Chakraborty, R. K. Yadav, G. V. Reddy, and A. Roy-Chowdhury, "Cell resolution 3d reconstruction of developing multilayer tissues from sparsely sampled volumetric microscopy images," *IEEE International Conference on Bioinformatics And Biomedicine*, pp. 378–383, 2011.