Model-Based Approach for Tracking Embryogenesis in *Caenorhabditis Elegans* Fluorescence Microscopy Data

Oleh Dzyubachyk,[1*] Rob Jelier,[2] Ben Lehner,[2] Wiro Niessen,[1,3] Erik Meijering[1]

Abstract— The nematode *Caenorhabditis elegans* (*C. elegans*) is a widely used model organism in biological investigations. Due to its well-known and invariant cell lineage tree, it can be used to study the effects of mutations and various disease processes. Effective and efficient analysis of the wealth of time-lapse fluorescence microscopy image data acquired in such studies requires automation of the cell segmentation and tracking tasks involved. This is hampered by many factors, including autofluorescence effects, low and uneven contrast throughout the images, high noise levels, large numbers of possibly simultaneous cell divisions, and touching or clustering cells. In this paper, we present a new algorithm for segmentation and tracking of cells in *C. elegans* embryogenesis image data. It is based on the model evolution framework for image segmentation and uses a novel multi-object tracking scheme based on energy minimization via graph cuts. Preliminary experiments on publicly available test data demonstrate the potential of the algorithm compared to existing approaches.

I. INTRODUCTION

The nematode *C. elegans* is one of the most widely used model organisms in molecular and developmental biology. Its development from a single cell to the full-grown organism built of 959 somatic cells is well studied and completely invariant [1]. This offers excellent possibilities for biological research on cell differentiation and related subjects. The latest advances in microscopy technique greatly facilitate such research by providing imaging data of sufficient temporal and spatial resolution. However, analysis of such data also requires specialized segmentation and tracking tools that can cope with the high complexity of the data.

A. Imaging C. elegans Embryogenesis

Following the development of a *C. elegans* embryo throughout time by means of a microscope poses several challenges for both imaging and subsequent image analysis. Differential interference contrast (DIC) and fluorescence microscopy are two main imaging modalities that have been used in this context. Existing nuclei tracking algorithms in 4D Nomarski DIC image sequences [2] are based on distinguishing difference in texture between nuclei and cytoplasm. Besides being sensitive to noise, this method experiences difficulties in handling cell division events [1] where the daughter cells appear as very small objects, limiting the tracking to the 24-cell stage. Segmentation and tracking

[1]Biomedical Imaging Group Rotterdam, Departments of Medical Informatics and Radiology, Erasmus MC–University Medical Center Rotterdam, The Netherlands, [2] EMBL-CRG Systems Biology Unit, Centre for Genomic Regulation, Barcelona, Spain, [3]Imaging Science & Technology, Faculty of Applied Sciences, Delft University of Technology, The Netherlands.
*Email: o.dzyubachyk@erasmusmc.nl

algorithms have recently been applied more successfully in fluorescence microscopy [1]. Despite the several disadvantages (discussed below), it provides good visualization of a developing *C. elegans* embryo during at least a 6-hour period with a time sampling rate of 1 image per minute, which is sufficient to properly capture all relevant motions.

B. Previous Work on Segmentation and Tracking

Segmentation of *C. elegans* nuclei from fluorescence microscopy images is a highly complicated problem. First, since a living organism is being imaged, autofluorescence is prominent, resulting in the presence of disturbing background structures. Second, lowering the exposure times to allow tracking the embryogenesis over longer time periods, results in low contrast and high noise levels. Third, the limited depth penetration causes a large difference in contrast between nuclei situated close to the surface and those deeper in the embryo. Fourth, though most of the time a nucleus appears in the image as a homogeneously stained sphere, the nuclei shapes and intensity patterns may sometimes be rather irregular, especially during cell division. Fifth, at later stages of *C. elegans* development, nuclei tend to cluster together, and the boundary between touching cells is not always clearly visible in the images. Sixth, this effect is made worse by the fact that the depth resolution is often much smaller than the in-plane resolution, which hampers accurate detection of two touching cells as two separate objects, and may lead to undersegmentation. In many cases, the use of temporal information is the only possibility to segment clustered cells correctly. Seventh, a large number of cell divisions may occur nearly simultaneously, making it difficult to properly connect mother and daughter cells.

So far, few attempts have been made to perform automated segmentation and tracking of fluorescence microscopy *C. elegans* data. In [3], a 3D watershed segmentation algorithm is presented. Using a watershed-based framework requires a substantial postprocessing step for combining oversegmented regions together and splitting undersegmented clusters. The framework for segmenting three-dimensional *C. elegans* data via gradient flow [4] is able to produce good quality segmentation of highly clustered nuclei. Finally, in [1] an algorithm for automated cell lineage tracing is presented. This algorithm uses a data preprocessing step, followed by nuclei detection and linking, and is capable of tracking cell nuclei of a developing *C. elegans* embryo until the 350-cell stage with reasonable quality. However, after the 194-cell stage, when cells become more clustered, the overall performance of this algorithm significantly drops. A program

(StarryNite [5]) implementing the mentioned algorithm and a tool (AceTree) for analysis and editing of its results are publicly available online [1].

C. Our Algorithm

Here we present a new algorithm for tracking *C. elegans* embryogenesis data obtained by fluorescence microscopy imaging. The algorithm is based on the model evolution framework, the key idea of which is that segmentation and tracking of objects are performed simultaneously. The main advantage of this is that information about the state of an object at a previous time point can be incorporated in the detection and segmentation of the same object in the next time point. This greatly facilitates segmentation and tracking of clustered nuclei. In addition, we introduce a novel multi-object tracking scheme based on energy minimization via graph cuts, which turns out to be very efficient and effective for the considered application. To our knowledge, this algorithm is the first application of model-based approach for tracking embryogenesis and the first method to perform both segmentation and tracking on *C. elegans* fluorescence microscopy image sequences.

II. MODEL-BASED TRACKING

As a starting point we use our cell tracking algorithm presented earlier [6], and the coupled active surfaces framework [7]. The main idea behind tracking objects within this framework is that each object is represented by one active contour (level set in our case). The evolution of each active contour is performed via energy minimization, where the energy consists of both data-based and regularity-based terms. After the optimal solution for the current frame is obtained, the set of the active contours is updated: new contours are created for divided cells or cells that entered the frame, and contours corresponding to the cells that died or left the frame are removed from the set. Tracking on the next frame starts from the final position on the previous one.

To apply the algorithm to *C. elegans* data, we need to introduce several modifications. The main point that makes the algorithm in [6] not practically suitable for tracking this kind of data is that it is rather computationally expensive, with computational load proportional to the total number of objects being tracked. Even though real-time performance is not required, the long time required for the algorithm to process the whole data set strongly limits the possibility to perform efficient research. The second major disadvantage of tracking *C. elegans* data by the level-set based algorithm is that level sets are known to have difficulty splitting closely positioned but already separated cells just after division. As a result, if a level set function corresponding to a mitotic cell encounters such situation, the division event might be detected later than it happened in reality. In order to overcome the mentioned difficulties we suggest using a different energy minimization algorithm: instead of level sets we use the graph-cut technique [8].

A. Graph-Cut Based Energy Minimization

Graph cuts recently became a popular energy minimization framework and were successfully applied to various computer vision problems. However, graph-cut based tracking applications (especially for multi-object tracking) are still rare due to the global nature of this technique [9]. Here we resolve this problem in a similar manner to narrow-banded level-set evolution, where, instead of performing the gradient descent within the narrow band around the current region occupied by the object, we apply within that narrow band graph-cut energy minimization on the following function

$$E = \sum_{p \in N_b} R_p(f_p) + \alpha \sum_{\substack{p,q \in N; \\ f_p \neq f_q}} B_{p,q}(f_p, f_q). \quad (1)$$

Here p and q are voxels in the narrow band N_b, N is the set of neighbor pairings, and α is the scalar that controls the smoothness of the detected boundary of the object. The region energy term R_p is set to the difference between the probabilities of a voxel $\mathbf{x}$ to belong to the foreground or to the background

$$R_p(f) = -\ln p(I|\text{foreground}) + \ln p(I|\text{background}). \quad (2)$$

Here we have chosen to use the non-parametric probability density function $p(\mathbf{x})$, which is calculated from the smoothed histogram of the image intensity $I(\mathbf{x})$ within the corresponding region. The boundary term $B_{p,q}$ we set to be the function of image contrast in the form

$$B_{p,q}(f_p, f_q) = \exp\left(-\frac{\|I_p - I_q\|^2}{2\sigma^2}\right) \frac{1}{\|p - q\|}, \quad (3)$$

where the parameter σ can be calculated from the data [9] or set empirically.

This approach enables multi-object tracking, since the energy minimization is performed only in the vicinity of the current position of the object, which allows to detect and resolve possible collisions between objects. Moreover, since level sets are no longer used, there is no need to perform costly re-initializations. Thus we can represent each object by a binary mask, which greatly reduces the computational load of the algorithm.

B. General Algorithm Flow

Our algorithm begins tracking with processing of the first image stack. It is performed in two steps: first an initial segmentation is obtained via the non-PDE based energy minimization method [10] and after that the segmentation result is refined by using the graph cut energy minimization described in the previous section. Next, the obtained solution is split into connected components representing separate objects. In such a way each of the detected nuclei are represented by two attributes: a binary mask for occupied region and a characteristic label.

The tracking on each time point starts from the solution obtained for the previous time point. The minimization of (1) is performed within a narrow band around the current object position. Next, voxels that appear to be part of the foreground

as a result of the energy minimization are included in the object's region, and those of the background are removed. This process is repeated until the number of voxels that changed sign after one step does not exceed a given value, or until the maximum number of iterations is reached.

As mentioned, such local update of the region representing an object being tracked helps to detect situations when two objects come close to each other. In such case, if proper measures are not taken, the region corresponding to one object might also capture the neighboring one. In our algorithm the non-empty intersection of two different narrow bands is the indicator of such situations. For separation of such objects we apply the Radon-transform based scheme described in detail in [6].

After reaching the final segmentation of the image stack at the current time point, we perform the necessary actions: create new labeled regions in case cell division took place and remove the labels in case the track of some of the nuclei is lost. An additional postprocessing step can be used here if recovery of the lost cells is necessary (in a manner similar to the one described in [6]).

III. RESULTS

We tested our algorithm on the part (covering from 4- to 180-cell stages) of the *C. elegans* embryogenesis image sequence described in [1]. The whole image sequence was downsampled to reduce the size of each image twice in each dimension in order to speed up computations and to remove the line skipping artifacts. Beside that no other image preprocessing was used. The result of the segmentation of the image sequence from [1] is shown in Fig. 1. The figure clearly shows that our method provides good quality segmentation of all 180 nuclei present at that time point even in the noisy images with strong background structures.

Quantitatively the tracking performance of our algorithm can be summarized as in Table I. For convenience of comparison of both methods we used the same evaluation measures as in [1]. Notice, that since the exact values for the method of Bao et al. are not provided in the article, we estimated them approximately from the corresponding columns of the error plots. The results above clearly indicate the superiority of our method for tracking on the described data sequence, especially for the 4–51-cell and 51–102-cell stages. All the objects present on the image at the time points corresponding to those stages were properly identified, resulting in the absence of detection errors. Our algorithm also exhibits very good performance with respect to tracking cell divisions. The polar body, which was detected and tracked during a number of time steps, and then again fell below the object size threshold, was the cause of the only false division that happened. Notice, that the polar body is not included in further analysis because its presence or absence had minor influence on the rest of the results, and consequently it can be easily added to or removed from the tracking results manually. The number of missed cell divisions for both 4–51-cell and 51–102-cell stages was equal to one, and in both cases the daughter cells where properly recovered

by the postprocessing detection step. At later time points (corresponding to the 102–180-cells stage) the quality of results of our algorithm decreases, though still remaining better than the one of the algorithm of Bao et al. at this stage. Such change of performance can be explained by the large increase of the number of nuclei being tracked as well as nuclei clustering.

The algorithm was programmed in the Matlab version R14. Processing times ranged from less than 1 minute for the first image stack to about 20 minutes for the last one due to the increase of the number of objects being tracked.

IV. DISCUSSION AND FUTURE WORKS

Results described in the previous section clearly show the potential of our method and its advantage over the method currently used for tracking *C. elegans* embryogenesis imaged by means of fluorescence microscopy. Our tracking algorithm implements combined segmentation and tracking, which allows to incorporate prior temporal information into the segmentation algorithm. The latter greatly improves detection of separate objects in clustered environments. Another major advantage of our approach in comparison with the state-of-the-art method [1] is that it detects the real boundaries of each nucleus instead of approximating its shape by a sphere. In this way we obtain much more information about the real state of the system. Unlike the scheme in [1], where the probability of cell division increased with time, our scheme for detection of mitotic cells relies only on the shape and intensity information. Thus our algorithm is more suitable for performing tracking on mutated *C. elegans* embryos, where cell divisions do not necessarily have to be bound to happen with certain periodicity. Finally, our method is simple in use: it contains a minimal number of parameters, each of which is easy to choose because it has clear physical meaning. Important is that these parameters are not dependent on time.

Using graph-cut based energy minimization has several advantages over using the active contour framework. First, it is faster because it does not require performing time-costly procedures like re-initialization, calculation of generalized Heaviside function, curvature, etc. Second, it is binary, thus very efficient with respect to memory consumption. Third, energy minimization via graph cuts is known to find the global energy minimum, which is very helpful for proper detection of mitotic events.

In the future our algorithm will be tested on more data, especially including later time points, with larger number of objects being tracked and where nuclei have smaller size and cluster together. In order to increase the robustness of our algorithm on such data we are planning to implement a postprocessing step that includes possible splitting of undersegmented cells. Also it is necessary to incorporate in our algorithm a scheme for automatic distinguishing between a nuclei of a small size and the polar body. Finally, a good method for preprocessing the data (noise and background removal) can significantly further improve the performance and robustness of our algorithm.
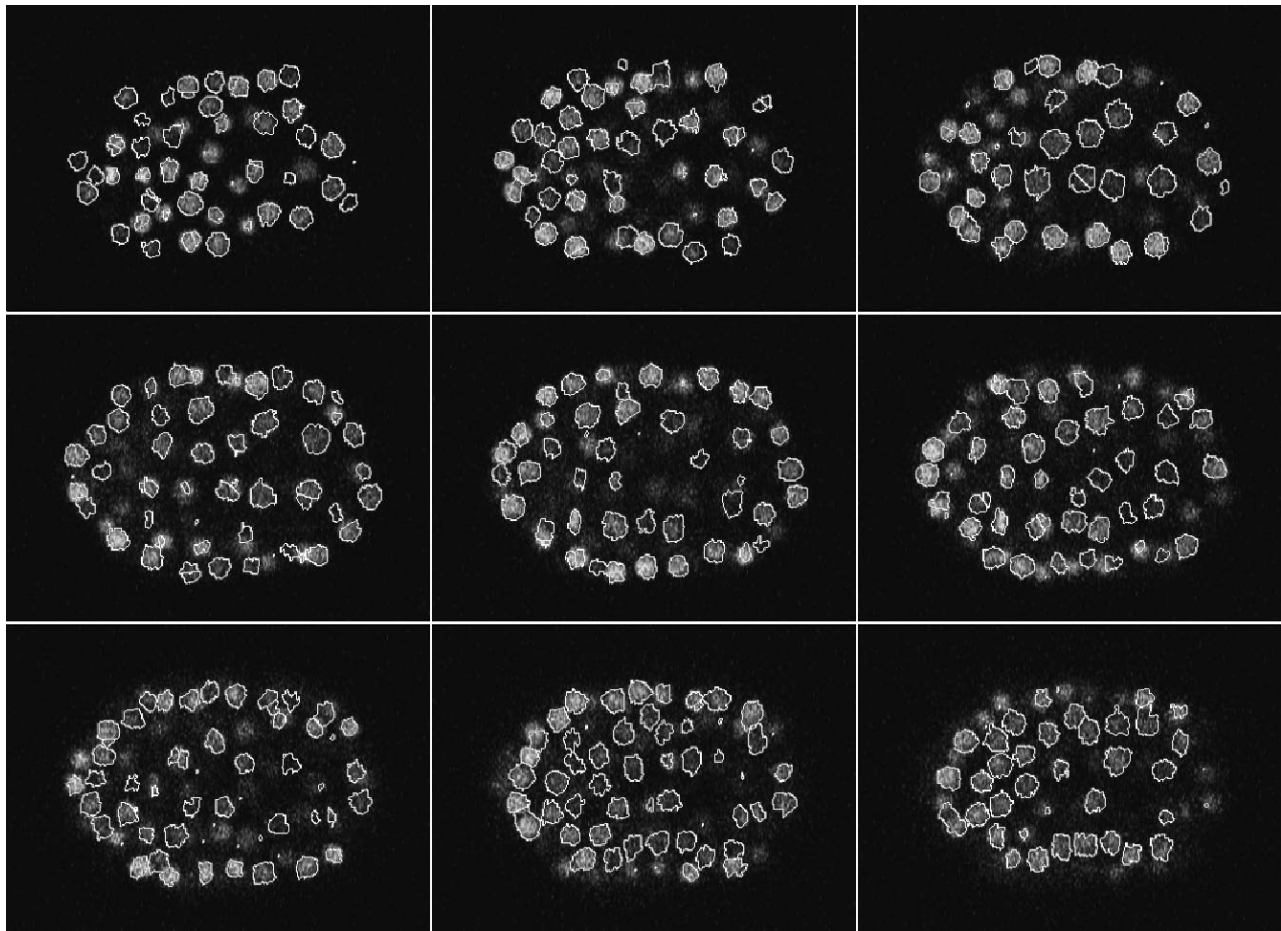
Fig. 1. Example of cell segmentation and tracking in the *C. elegans* embryogenesis image sequence from [1]. The images are 9 slices out of 35 slices for the image stack corresponding to the 180-cell stage. The white contours overlaid on the image show the results of the model-evolution based cell segmentation and tracking algorithm proposed in this paper.

TABLE I

TRACKING PERFORMANCE OF OUR ALGORITHM IN COMPARISON TO THE ALGORITHM OF BAO ET AL. [1].

Feature	False Negative (per 1000 nuclei)			False Positive (per 1000 nuclei)			Error per Division		
Cell stage	4–51	51–102	102–180	4–51	51–102	102–180	4–51	51–102	102–180
Our method	0	0	2.0	0	0	2.3	0.02	0.02	0.06
Bao et al.	$\sim 0-10$	$\sim 0-4$	$\sim 0-8$	$\sim 0-15$	$\sim 0-5$	$\sim 0-7$	$\sim 0-0.14$	$\sim 0-0.15$	$\sim 0-0.08$

REFERENCES

[1] Z. Bao, J. I. Murray, T. Boyle, S. L. Ooi, M. J. Sandel, and R. H. Waterston, "Automated cell lineage tracing in Caenorhabditis elegans," *Proc Natl Acad Sci USA*, vol. 103, no. 8, pp. 2707–2712, 2006.

[2] S. Hamahashi, S. Onami, and H. Kitano, "Detection of nuclei in 4D Nomarski DIC microscope images of early caenorhabditis elegans embryos using local image entropy and object tracking," *BMC bioinformatics*, vol. 6, pp. 125, 2005.

[3] F. Long, H. Peng, and E. W. Myers, "Automatic segmentation of nuclei in 3D microscopy images of C. elegans," in *Proceedings of the 2007 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2007, pp. 536–539.

[4] G. Li, T. Liu, A. Tarokh, J. Nie, L. Guo, A. Mara, S. Holley, and S. Wong, "3D cell nuclei segmentation based on gradient flow tracking," *BMC Cell Biology*, vol. 8, pp. 40, 2007.

[5] J. I. Murray, Z. Bao, T. J. Boyle, and R. H. Waterston, "The lineaging of fluorescently-labeled *Caenorhabditis elegans* embryos with StarryNite and AceTree," *Nature Protocols*, vol. 1, no. 3, pp. 1468–1476, 2006.

[6] O. Dzyubachyk, W. J. Niessen, and E. Meijering, "Advanced level-set based multiple-cell segmentation and tracking in time-lapse fluorescence microscopy images," in *Proceedings of the 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2008, pp. 185–188.

[7] 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.

[8] Y. Boykov, O. Veksler, and R. Zabih, "Markov random fields with efficient approximations," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1998, pp. 648–655.

[9] J. Malcolm, Y. Rathi, and A. Tannenbaum, "Multi-object tracking through clutter using graph cuts," in *Workshop on Non-Rigid Registration and Tracking Through Learning*, 2007, pp. 1–5.

[10] B. Song, *Topics in variational PDE image segmentation, inpainting and denoising*, Ph.D. thesis, Dept. Math., Univ. California, Los Angeles, USA, 2003.