AUTOMATED LINEAGE TREE RECONSTRUCTION FROM CAENORHABDITIS ELEGANS IMAGE DATA USING PARTICLE FILTERING BASED CELL TRACKING

Noemí Carranza, Ihor Smal, Oleh Dzyubachyk, Wiro Niessen, Erik Meijering

Biomedical Imaging Group Rotterdam,
Departments of Medical Informatics and Radiology
Erasmus MC – University Medical Center Rotterdam, The Netherlands
Email: h.meijering@erasmusmc.nl

ABSTRACT

Caenorhabditis elegans is an important model organism for the study of molecular mechanisms of development and disease processes, due to its well-known genome and invariant cell lineage tree. Such studies generally produce vast amounts of image data, and require very robust and efficient algorithms to extract and characterize lineage phenotypes and to determine gene expression patterns. Previously published methods for this purpose show only mediocre performance and often require extensive manual post-processing. The challenge remains to develop more powerful and fully automated methods. In this paper we propose a new algorithm for C. elegans cell tracking and lineage reconstruction, based on a Bayesian estimation framework, implemented by means of particle filtering. The tracking is enhanced with a detection stage, based on the h-dome transform. Preliminary experiments on several image sequences demonstrate that the new tracking algorithm is able to reconstruct the lineage tree, at least until the 350-cell stage, without manual intervention, at low computational cost and with low error rates.

Index Terms— Caenorhabditis elegans, embryogenesis, fluorescence microscopy, particle filtering, cell tracking.

1. INTRODUCTION

The nematode worm Caenorhabditis elegans (C. elegans) is a model organism that is used extensively in biology. Comprising about 1,000 cells, it is the most primitive animal to exhibit characteristics that are important in the study of human biology, the effects of mutations, and various disease processes. Its well-known genome, rapid generation time, and the fact that its cell lineage is invariant, describing the fate of every cell during development to the adult stage [1], make this organism very suitable for the study and characterization of genetic mutants. An important challenge for geneticists is to design screens that will identify mutations that specifically disrupt biological processes of interest. Imaging based screens generate vast amounts of 3D image sequences, and an understanding of the embryo's body plan from the data can be obtained most effectively by reconstructing the lineage tree of the dividing cells within the embryo [2].

In recent years, some tools have been developed for automated cell segmentation and tracking in embryogenesis, based on spherical model fitting [3] or level-set based model-evolution [4]. The experimental results reported in these works demonstrated that such approaches work well in the early stages of development, with relatively small numbers of cells, but that performance quickly deteriorates in the 200-350 cell stage, and completely breaks down in later stages, requiring many hours or even days of manual post-editing to correct for errors. For comparison of lineage trees in a high-throughput fashion, it is of critical importance to develop more robust and accurate methods, which are capable of fully-automated reconstruction with minimal errors until at least the stage before the last division.

In this paper we propose a new scheme for tracking of C. elegans embryogenesis. It is based on probabilistic approaches to object tracking using sequential Monte Carlo methods, in particular particle filtering, which uses an M-component mixture model for tracking multiple targets. In other applications this has been shown to perform superiorly compared to non-probabilistic methods [5]. The scheme is adapted with a new observation model for cell tracking, and extended with a detection stage based on the h-dome transform from grayscale morphology, which is capable of robustly detecting missed cells in very noisy data.

2. METHOD

2.1. Bayesian estimation framework

Using the Bayesian estimation framework we aim to estimate the position and size of the cells. At each time step t, the cell is characterized by the state vector $\mathbf{s}_t = (x_t, y_t, z_t, h_t)$, where (x_t, y_t, z_t) denotes the spatial position and h_t the size (radius) of the cell (Fig. 1). Additionally, we know some premises: it is not possible for new cells to enter the field-of-view other than by mitosis, and this division is only from one cell ("mother") to two cells ("daughters"). Some cells

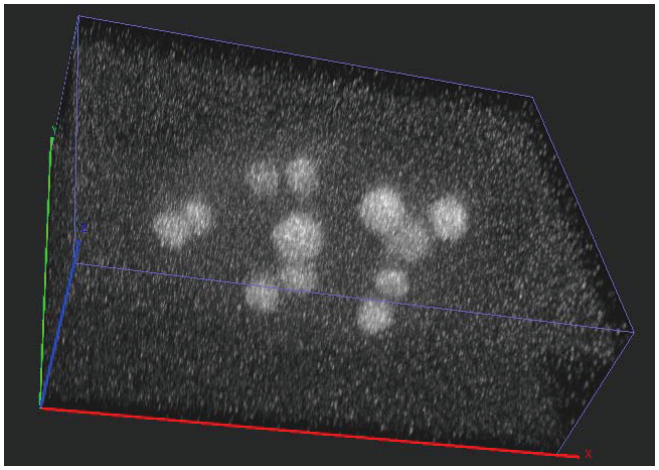

Fig. 1. Sample 3D rendering of one time step (12-cell stage) of a C. elegans fluorescence microscopy image data set.

will die in the last stages. The Bayesian tracking approach is used to recursively estimate a time evolving posterior distribution $p(\mathbf{s}_t|\mathbf{z}_{1:t})$ that describes the cell state given all the observations (images) $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$ up to time t. This probability density function (pdf) $p(\mathbf{s}_t|\mathbf{z}_{1:t})$ can be obtained recursively in two steps, prediction and update. For the prediction step, the state evolution $D(\mathbf{s}_t|\mathbf{s}_{t-1})$ is needed:

$$p(\mathbf{s}_t|\mathbf{z}_{1:t-1}) = \int D(\mathbf{s}_t|\mathbf{s}_{t-1})p(\mathbf{s}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{s}_{t-1}. \quad (1)$$

The update step is as follows:

$$p(\mathbf{s}_t|\mathbf{z}_{1:t}) \propto L(\mathbf{z}_t|\mathbf{s}_t)p(\mathbf{s}_t|\mathbf{z}_{1:t-1}), \quad (2)$$

where $L(\mathbf{z}_t|\mathbf{s}_t)$ is the likelihood that relates the noisy measurements (images) to states $\mathbf{s}_t$. As we are dealing with a multi-modal problem (we have multiple cells in each time step), the posterior distribution can be modeled as an M-component mixture:

$$p(\mathbf{s}_t|\mathbf{z}_{1:t}) = \sum_{m=1}^{M_t} \pi_{m,t} p_m(\mathbf{s}_t|\mathbf{z}_{1:t}), \quad (3)$$

where the number of cells M_t changes in time, $\sum_{m=1}^{M_t} \pi_{m,t} = 1$, and $p_m(\mathbf{s}_t|\mathbf{z}_{1:t})$ is the pdf for cell m.

2.2. Particle filters

Particle filters are sequential Monte-Carlo techniques for estimating the posterior pdf with a set of samples [6, 7]. They approximate the pdf by N_s random measures ("particles") associated to N_s corresponding normalized weights $\{\mathbf{s}_{m,t}^{(i)}, w_{m,t}^{(i)}\}_{i=1}^{N_s}$ as

$$p_m(\mathbf{s}_t|\mathbf{z}_{1:t}) \approx \sum_{i=1}^{N_s} w_{m,t}^{(i)} \delta(\mathbf{s}_t - \mathbf{s}_{m,t}^{(i)}). \quad (4)$$

The weights are updated according to

$$w_{m,t}^{(i)} \propto w_{m,t-1}^{(i)} L(\mathbf{z}_t|\mathbf{s}_{m,t}^{(i)}), \quad (5)$$

and $\mathbf{s}_{m,t}^{(i)}$ are obtained by means of the dynamic model $D(\mathbf{s}_t|\mathbf{s}_{t-1})$. In our case, the motion of the cells is random, so the dynamic model is a random walk (using a normal distribution, where the mean is the position of the cell at time $t-1$ and the variance is the variance of the surviving particles). In each iteration, particles with low weights will be removed, and only N_e with significant weight will survive, where

$$N_e = \frac{1}{\sum_{i=1}^{N_s} (w_t^{(i)})^2}. \quad (6)$$

At each time step t, we define the estimated state of the cell, $\hat{\mathbf{s}}_{m,t}$, by a maximum a posteriori (MAP) estimate, given by the particle $\mathbf{s}_{m,t}^{(i)}$ with the highest weight.

2.2.1. *Observation model*

For each cell, the likelihood $L(\mathbf{z}_t|\mathbf{s}_t)$ of the state $\mathbf{s}_{m,t}$ depends on the size and location of the cell. The model is depicted in Fig. 2. Using the average image intensities in the three different regions (the center μ_c, inside μ_i, and outside μ_o), the corresponding variances in those regions (σ_i, σ_o), and the numbers of pixels that were used to compute the means and variances (n_i, n_o), we define the likelihood as

$$L(\mathbf{z}_t|\mathbf{s}_{m,t}) = \begin{cases} \left(\frac{\mu_i - \mu_o}{\sqrt{\sigma_i^2/n_i + \sigma_o^2/n_o}} \right)^2 \cdot \Phi, & \mu_c \geq \mu_i > \mu_o, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where

$$\Phi = \prod_{m'=1, m' \neq m}^{M_t} \phi(d_{m,m'})$$

penalizes particles close to other cells, with $d_{m,m'}$ the spatial distance between the states $\mathbf{s}_{m,t}$ and $\hat{\mathbf{s}}_{m',t}$, and $\phi(d) \approx 0$ if d is small and $\phi(d) \approx 1$ if d is large.

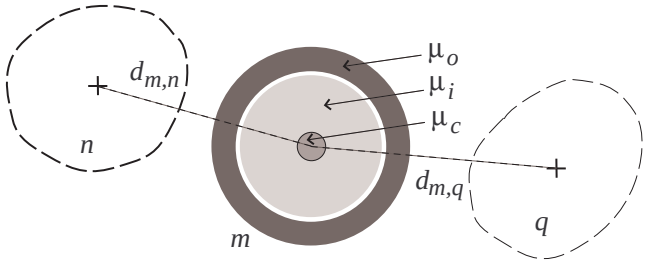

Fig. 2. 2D representation of the real 3D observation model. The estimated positions of the neighboring cells n and q are marked with "+", and the distances from them to the hypothesized location of the cell m are denoted by $d_{m,n}$ and $d_{m,q}$. The gray areas represent the image regions, which are used to compute the average intensities μ_i, μ_o, and μ_c.

2.2.2. *Division and death of cells*

One of the most important aspects of tracking embryogenesis image sequences is how to detect cell divisions. For this purpose we use the mean-shift clustering technique [8]. With this method, we detect if the particles associated to a cell are divided in one or two clusters. When the center of the two clusters is separated more than the typical cell size, we determine that the cell has divided (Fig. 3). Near the 350-cell stage, some cells start to die. In our algorithm, we detect a cell death event when the variance of the particles is too high or the number of particles is too small.

2.3. Detection with h-dome transform

During cell division, tracking may fail if the daughter cells move too far with respect to the original cell position in the previous frame. Furthermore, if cells are missed, mismatches may occur. To avoid these problems, a detection stage is added, based on the h-dome transform, which will detect any possible missed cells. The transform "cuts off" structures of a predetermined height from the top around local intensity maxima, producing "dome"-like structures [9]. The h-dome image $D_h(I)$ of an image I is given by

$$D_h(I) = I - \rho_I(I - h) \tag{8}$$

where the grayscale reconstruction $\rho_I(\psi)$ of image I from an image ψ ($I \geq \psi$) is obtained by iterating grayscale geodesic dilations of ψ "under" I until stability is reached [10]. Geometrically speaking, an h-dome D of image I is a connected component of pixels such that every pixel p, neighboring D, satisfies $I(p) < \min\{I(q)|q \in D\}$ and $\max\{I(q)|q \in D\} - \min\{I(q)|q \in D\} < h$. In this case, as we only want to find a small number of objects (lost cells), h can be easily estimated from the existing cells and is not critical, and we avoid detecting noise that could be interpreted as a cell. If missing cells are detected, they are back-tracked with the particle filtering algorithm and matched with the correct mother cell.

3. RESULTS

3.1. Images

Four image sequences (of size 500 x 700 pixels x 31 slices x 250 time steps and of varying quality) of C. elegans embryogenesis were analyzed with the proposed method. To reduce computation time, the images were downsampled in x and y by a factor of 3. In these preliminary experiments (except in one of the sequences, where annotations were available only until the 180-cell stage), tracking was performed from the 4-cell stage until the 350-cell stage, the latter of which has been shown in previous works [3] to be a critical stage where existing methods break down.

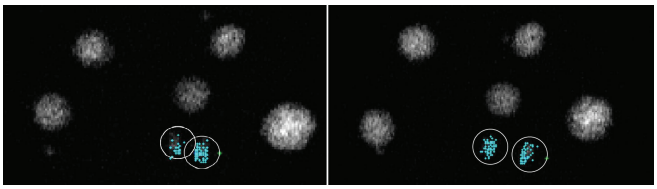

Fig. 3. Example of cell division. In the right image, two clusters of particles are found (for a given cell) that are well separated (distance larger than the typical cell size), whereas in the left image (previous frame) they are not, indicating that a cell division has occurred.

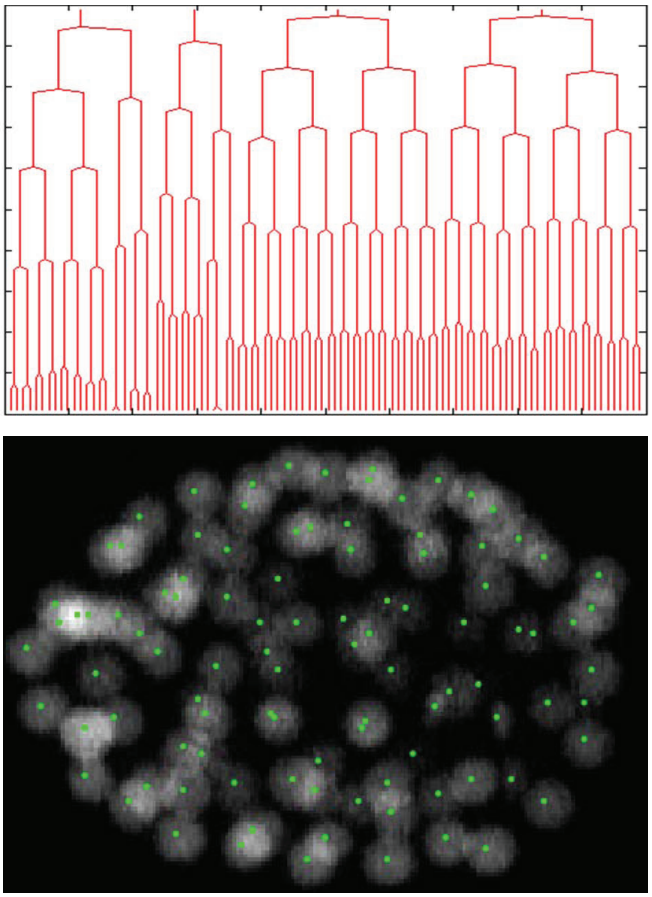

Fig. 4. Lineage tree (top) obtained for one of the sequences until the 100-cell stage and a 2D projection (bottom) of the image data at the latter stage together with the found center positions (indicated in green).

3.2. Validation

The efficiency and accuracy of the proposed method was tested. The output positions of the cells were compared with expert manual annotations. One example of a resulting lineage tree, starting with four cells (EMS, P2, ABa, ABp), is shown in Fig. 4. The tracking results were binned into a limited number of critical stages (ranges of numbers of cells) as in [3]. Errors were classified into false negatives (incorrectly lost cells), false positives (incorrectly detected new cells), and errors of cell division (a correctly tracked cell division is one for which the mother and the two daughter cells are all correctly identified and matched).

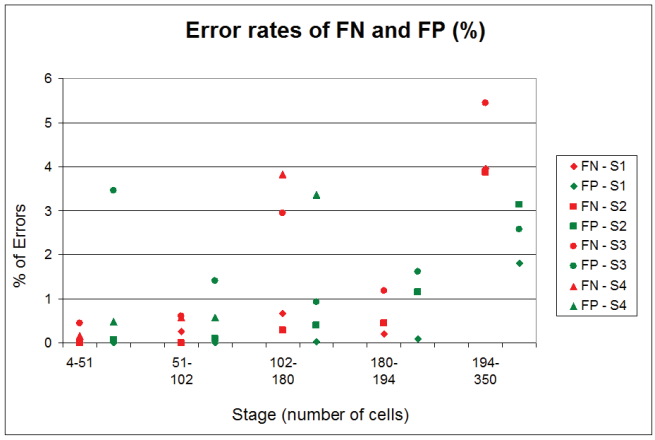

Fig. 5. Error rates for false negatives (FN) and false positives (FP) for each of the four image sequences (Sx).

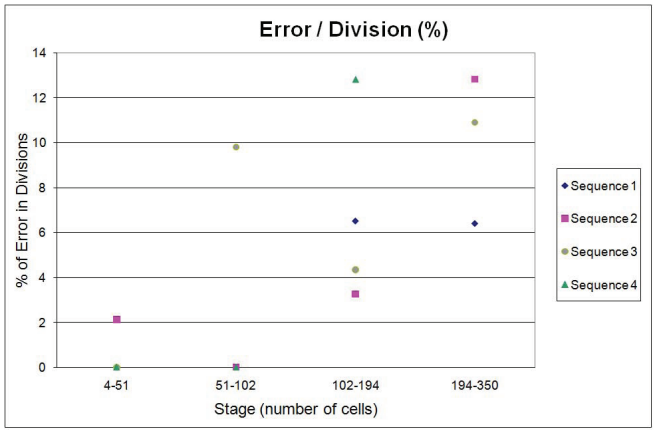

Fig. 6. Error rates for tracking cell divisions in each of the four sequences.

Error rates for false negatives and false positives are shown in Fig. 5, while division errors are shown in Fig. 6. The error rates are relatively low, especially in the early embryonic stages. The difference in error rates between the four sequences is mainly due to the difference in image quality (the sequences are numbered in the order from highest to lowest quality), where sequences with lower quality usually have higher error rates, as expected. The majority of errors found in our experiments are false negatives and false positives, and are comparable to [3], while the errors in tracking cell divisions are slightly lower.

The downsampling step does not affect the FN and FP error rates in the first stages, where cells are still relatively large and well separated. In the last stages, the error rates could be lowered by keeping a higher resolution, but the required extra computation time was too high compared with the achieved improvements. Currently, the computation time of the algorithm is around 1.5 hours for each sequence, with an implementation in Matlab and Java (for the more time consuming parts), running on a standard PC (Intel Xeon Quad-Core 2.26GHz CPU with 12GB RAM). Further optimization of the code is possible.

4. CONCLUSIONS

In this paper, we have proposed a novel particle filtering based method for cell tracking of C. elegans embryogenesis, enhanced with a detection stage implemented with the h-dome transform. The method is built within a Bayesian estimation framework, incorporating prior knowledge about cell behavior and appearance. The results of validation experiments on four sequences until the 350-cell stage by comparison with expert manual annotations show that the method is able to reconstruct lineage trees with low error rates. Our future work will address reaching the later stages of development while maintaining low error rates.

5. ACKNOWLEDGMENTS

This research was financially supported by the European Commission in the Seventh-Framework Programme (FP7, grant agreement 201842, the ENCITE project). The C. elegans test data sets were kindly provided by the Waterston Lab, University of Washington, Seattle, WA, USA.

6. REFERENCES

[1] J. E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson, "The embryonic cell lineage of the nematode Caenorhabditis elegans," *Develop Biol*, vol. 100, no. 1, pp. 64–119, 1983.

[2] E. M. Jorgensen and S. E. Mango, "The art and design of genetic screens: Caenorhabditis elegans," *Nat Rev Genet*, vol. 3, no. 5, pp. 356–369, 2002.

[3] 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.

[4] O. Dzyubachyk, R. Jelier, B. Lehner, W. Niessen, and E. Meijering, "Model-based approach for tracking embryogenesis in Caenorhabditis elegans fluorescence microscopy data," in *Proc IEEE Eng Med Biol Soc*, 2009, pp. 5356–5359.

[5] I. Smal, K. Draegestein, N. Galjart, W. Niessen, and E. Meijering, "Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: Application to microtubule growth analysis," *IEEE Trans Med Imaging*, vol. 27, no. 6, pp. 789–804, 2008.

[6] S. M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans Signal Process*, vol. 50, no. 2, pp. 174–188, 2002.

[7] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, Berlin, 2001.

[8] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans Pattern Anal Mach Intell*, vol. 24, no. 5, pp. 603–619, 2002.

[9] I. Smal, W. Niessen, and E. Meijering, "A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering," in *Proc IEEE Int Symp Biomed Imaging*, 2008, pp. 264–267.

[10] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans Image Process*, vol. 2, no. 2, pp. 176–201, 1993.