DELTR: DIGITAL EMBRYO LINEAGE TREE RECONSTRUCTOR

X. Lou [1], *F. O. Kaster* [1], *M. S. Lindner* [1], *B. X. Kausler* [1], *U. Köthe* [1],
B. Höckendorf [2], *J. Wittbrodt* [2], *H. Jänicke* [1], *F. A. Hamprecht* [1*]

[1] Interdisciplinary Center for Scientific Computing, University of Heidelberg, 69115 Heidelberg, Germany
[2]Institute for Zoology, University of Heidelberg, 69120 Heidelberg, Germany

ABSTRACT

We present DELTR, an automated pipeline for the analysis of time-resolved light sheet fluorescence microscopy images of zebrafish embryogenesis. It comprises 3D nucleus segmentation using shape-regularized graph cuts, parallelized extraction of geometrical features, and cell tracking by means of combinatorial optimization. We also discuss the interactive visualization software used for validating the results, and describe our advances towards reconstructing the entire cell lineage tree of the zebrafish. Our method achieves ca. 96 % accuracy for cell nucleus detection and ca. 90 % accuracy for the association of nuclei across subsequent time steps.

Index Terms— Cell Lineage Reconstruction, Segmentation, Shape Regularization, Cell Tracking, Visualization

1. INTRODUCTION

Digital scanned laser light sheet fluorescence microscopy (DSLM) is a recent time-resolved 3D live-cell imaging technique which provides unprecedented spatio-temporal resolution and signal-to-noise ratio at low energy load [1]. This makes it an excellent tool for elucidating the embryonic development of vertebrates by tracking every cell over time and determining both its lineage and the subsequent fate of its progeny: this information can be summarized into a cell lineage tree. The zebrafish (*Danio rerio*) is a particularly promising model organism for such studies due to the transparency of the larvae and their nearly constant size in the first 24 hours of development.

However, manual data analysis is impracticable due to the large amount of data: typical image volumes have a size of several hundred megavoxels, and several hundreds of these volumes are acquired over time: hence automatical image analysis is called for. In this paper, we present an automated pipeline for cell lineage reconstruction from DSLM data.

2. RELATED WORK

Cell lineage reconstruction has been pioneered in the nematode *C. elegans* [2]. Recent work [3, 4] on cell lineage reconstruction culminates in the publication of the zebrafish lineage tree up to the 1000-cell stadium, based on label-free microscopic imagery [5]. However, the only published method for DSLM data was presented in [1]: There the authors segment cell nuclei by local adaptive thresholding and perform local nucleus tracking by nearest neighbor search. By our segmentation method, we improve upon the nucleus detection efficiency of that paper.

Our segmentation method is most closely related to the one presented in [6]. As in this paper, we use blob filter responses that are coherent across multiple scales as initial seeds for our segmentation and refine them via discrete graph cut optimization. However, our approach differs by the use of more flexible foreground cues based on discriminative random forest classifiers [7] instead of the Poisson mixture model employed in that article, and by explicit shape regularization using a multi-object generalization of the graph cuts algorithm presented in [8]. An alternative approach uses continuous optimization by level set evolution for nucleus segmentation as well as for initial denoising and seed generation [9]: however, an advantage of discrete optimization techniques is that convergence to the global optimum can be guaranteed in many cases.

Our tracking approach based on integer linear programming has the advantages that the tracking is interpretable as the exact optimization of a defined energy functional and that it can natively account for cell divisions; furthermore it is computationally efficient given the segmented cell nuclei. Alternative approaches combine segmentation and tracking in one step by initializing a set of active contours in the first time frame and updating these contours in the subsequent time frames using motion filtering. This requires that cell divisions are handled in a special way, e.g. by stitching together different single-cell trajectories [4].

3. AUTOMATED LINEAGE RECONSTRUCTION

3.1. Workflow Overview

Reconstruction of cell lineage mainly relies on two steps: segmentation of cell nuclei and tracking of their movement and division. Joint approaches [4] for both steps do exist, but we

*Corresponding Author: Fred.Hamprecht@iwr.uni-heidelberg.de.

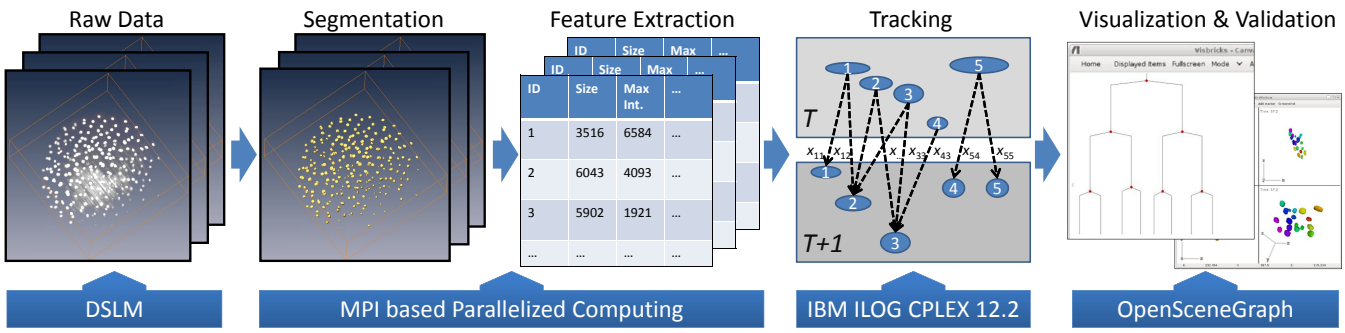

Fig. 1. Workflow of DELTR. Segmentation of the raw DSLM data produces cell nuclei on which features are extracted. These two computation-intensive tasks are parallelized volume-by-volume. These features provide information of inter-frame nuclei association for the tracking step, which is efficiently performed using IBM ILOG CPLEX. The final lineage tree is visualized by a newly developed tool based on OpenSceneGraph, which also allows interactive viewing of the raw data and the segmentation.

choose to address these problems subsequently because of the extremely high computational load (Fig. 1). Already the segmentation of the volumetric data at a single time point (4×10^8 voxels) is pushing the limits of most computing systems that are currently available, and a joint approach that has to propagate temporal information over hundred of time frames would be even more difficult to realize. In the following we discuss the details of each step.

3.2. Segmentation with Shape Regularization

The segmentation of DSLM data poses the following challenges: relatively low SNR compared to 2D imaging techniques, a striped background at early stages, severely cluttered cell nuclei at late stages, and high variability of cell brightness (the intensities of the dimmest and the brightest cells differ by a factor of 50). While various contributions have previously been made to the segmentation of such volumes, the quality of the segmentation results has not been sufficiently evaluated before with respect to the regularity of the shape. Segmented cell nuclei with irregular shape not only hinder visual inspection and evaluation but also jeopardize further tasks such as growth phase classification and tracking.

We model this problem using Markov Random Fields (MRF) and propose a novel energy formulation with four energy terms (Eq. 1): a data term as the probability of voxels being assigned to foreground or background, a smoothing term for spatial regularization, and a shape term and a flux term for shape/length regularization. The maximum a posteriori (MAP) solution of the MRF is obtained using the max-flow/min-cut algorithm [8]. The details of our segmentation method are explained elsewhere [10] and the software will be made publicly available.

$$\begin{aligned}\min_{\boldsymbol{l}} \quad & \Big\{ \lambda_{\text{data}} \sum_{p\in\boldsymbol{I}} E_{\text{data}}(l_p) + \lambda_{\text{smooth}} \sum_{\{p,q\}\in\boldsymbol{N}} E_{\text{smooth}}(l_p, l_q) \\ & + \lambda_{\text{shape}} \sum_{\{p,q\}\in\boldsymbol{N}} E_{\text{shape}}(l_p, l_q) + \lambda_{\text{flux}} \sum_{p\in\boldsymbol{I}} E_{\text{flux}}(l_p) \Big\} \\ \text{s.t.} \quad & \lambda_{\text{data}} + \lambda_{\text{smooth}} + \lambda_{\text{shape}} + \lambda_{\text{flux}} = 1, \\ & \lambda_{\text{data}} > 0, \lambda_{\text{smooth}} > 0, \lambda_{\text{shape}} > 0, \lambda_{\text{flux}} > 0. \end{aligned} \tag{1}$$

3.3. Feature Extraction and Tracking

Connected component labeling is used to transform the binary image generated by the segmentation step into a list of individual nucleus objects. The individual objects are efficiently stored in a dictionary of keys (DOK)-based sparse matrix representation. We characterize them by computing various geometrical and statistical features such as the intensity-weighted mean position (center of mass), volume or central moments and percentiles of the intensity distribution.

In order to efficiently track the large number of nuclei over time, we find the optimal joint association between nuclei for every pair of subsequent time frames. Formally, let i denote a nucleus from frame T and j, k denote nuclei from frame $T+1$, and let $\emptyset$ represent *no association*. We consider the following events: *move*, *division*, *disappearance* and *appearance*. While an (apparent) cell disappearance may be caused by apoptosis, it is more typically due to a cell leaving the field of view, misdetection in the segmentation step or erroneous removal; an (apparent) appearance happens when it is detected again at a later time. As shown in Table 1, all these events have associated costs: the constants c_{div}, c_{dis} and c_{app} are chosen such that appearances and disappearances are heavily penalized compared to divisions and moves. The results are sufficiently robust towards moderate variations of these val-

ues, so that no parameter learning was performed. Also, in order to rule out implausible events, we only consider at most k nearest neighbors of i within a given distance threshold.

Event	Notation	Cost
i *moves* to j	$i \to j$	$d^2_{i\to j}$
i *divides* into j and k	$i \to j+k$	$d^2_{i\to j} + d^2_{i\to k} + c_{\text{div}}$
i *disappears* in frame T	$i \to \emptyset$	c_{dis}
j *appears* in frame $T+1$	$\emptyset \to j$	c_{app}

Table 1. Summary of the events considered in the tracking. Here $d_{i\to j}$ is the Euclidean distance between the center of mass of the nuclei i (from frame T) and j (from frame $T+1$).

Let $\mathcal{M}$ be the set of all possible moves and $\mathcal{D}$ the set of all possible divisions. For each event in $\mathcal{M}$ and $\mathcal{D}$, we define a binary variable x indicating whether this event takes place or not. Thus, finding the optimum joint association is then an integer linear programming (ILP) problem:

$$\begin{aligned}\min_{\boldsymbol{x}} \quad & \Bigg\{ \sum_{(i\to j)\in\mathcal{M}} x_{i\to j}(c_{i\to j} - c_{i\to\emptyset} - c_{\emptyset\to j}) + \\ & \sum_{(i\to j+k)\in\mathcal{D}} x_{i\to j+k}(c_{i\to j+k} - c_{i\to\emptyset} - c_{\emptyset\to j} - c_{\emptyset\to k}) \Bigg\} \\ \text{s.t.} \quad & \sum_{j:(i\to j)\in\mathcal{M}} x_{i\to j} + \sum_{j,k:(i\to j+k)\in\mathcal{D}} x_{i\to j+k} \le 1 \quad \forall\, i, \\ & \sum_{i:(i\to j)\in\mathcal{M}} x_{i\to j} + \sum_{i,k:(i\to j+k)\in\mathcal{D}} x_{i\to j+k} \le 1 \quad \forall\, j. \end{aligned} \tag{2}$$

All cells not accounted for by either a division or a move are assumed to appear or disappear. Typically there are a few ten thousand variables (one for each division or move) and a few thousand constraints (one for every nucleus in each frame). We use a state-of-the-art ILP solver (ILOG CPLEX 12.2[1]) to solve this problem to global optimality within less than a minute per frame pair on a standard desktop machine. Note that several frame pairs may be processed in parallel.

3.4. Visualization

We created a software for interactive visualization of the results, which is based on the OpenSceneGraph 3D graphics API[2]. It offers the following capabilities:

- Visualization of all segmented nuclei in a given subvolume by their center-of-mass positions along with the principal component semiaxes or by volume rendering.
- Validation of individual nuclei by showing the cross-section of a selected nucleus across the plane defined by the leading principal components together with the segmentation isocontour.
- Visualization of the 3D trajectories of individual cells and their progeny over time.
- Synchronized display of the raw image data, nucleus segments and the cell lineage tree topology.

[1] http://www.ilog.com/products/cplex/
[2] http://www.openscenegraph.org

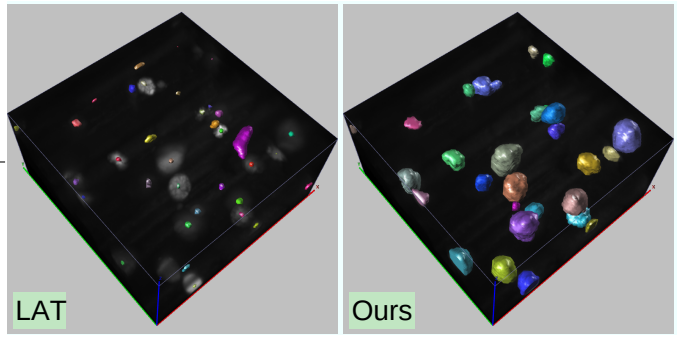

Fig. 2. Comparison between LAT [1] and our segmentation method [10] in terms of shape extraction. The segmentation is overlayed on the raw data.

4. RESULTS ON DIGITAL ZEBRAFISH EMBRYO

We applied our pipeline to reconstruct the cell lineage of the first 80 frames (2 to 4 hours of development; from a 66 nuclei stadium up to ca. 2400) of the animal view of a DSLM dataset of zebrafish embryos. Only a single nucleus channel is available. Both the segmentation and the tracking are evaluated with respect to manually generated ground truth.

4.1. Segmentation: Accuracy and Shape Extraction

Qualitatively, for 6 selected time steps our method achieves a false negative rate of 4 ± 1 % for cell detection, which is a reduction by 40 % over the original local adaptive thresholding (LAT) method used in [1]. Also, as shown in Fig. 2, our method extracts smooth and regularly shaped cell nuclei which is not possible by LAT.

4.2. Tracking: Detection Rates

The cell lineage ground truth for the tracking evaluation was manually generated for the first 25 time steps. To minimize the dependency on the segmentation accuracy, we considered only cell nuclei that were successfully found by the segmentation. The detection rate is 91 % out of 3008 move events and 71 % out of 180 division events; the number of appearance and disappearance events is too low to report significant rates. By visual inspection on the incorrect associations, we found that errors were mainly caused by high intensity speckles that were mistaken for cell nuclei by the segmentation.

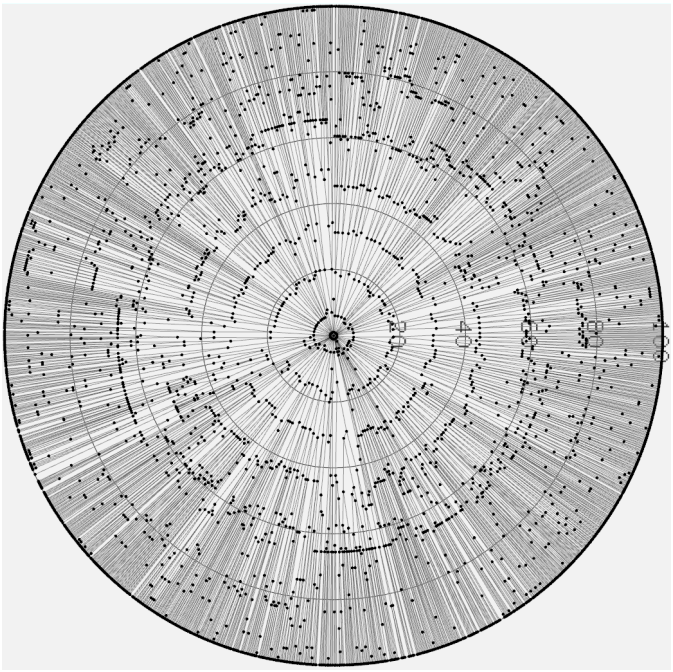

Fig. 3. Visualization of the pruned lineage tree using radial layout. Black dots indicate cell divisions. Five synchronous cell division cycles can be easily observed (up to frame 60).

4.3. Visualization: Interactivity and Efficiency

For our data with several thousand cell nuclei over 80 time steps, initialization of the visualization tool takes less than one minute. The computation of the lineage tree layout is instantaneous. Streaming technology allows to navigate the whole data range in a matter of seconds. Well designed linking and brushing among multiple views helps us to efficiently inspect the cause of errors such as disconnected links or abnormal branches (see Fig. 1, rightmost). After basic pruning that removes abnormally short branches, the entire lineage tree is visualized using radial layout, as shown in Fig. 3.

5. CONCLUSIONS & FUTURE WORK

In this paper we presented DELTR, an automated cell lineage reconstruction pipeline, and reported its performance on cell lineage reconstruction of the first 80 frames of the digital zebrafish embryo dataset. In comparison to manual or semi-automated approaches, an automated pipeline that encodes human knowledge (e.g. nucleus shape, temporal coherence) is crucial. For example, to understand the mechanism of mutation, a large number of such datasets have to be processed and compared against each other.

Built upon state-of-the-art machine learning, computer vision and optimization techniques, DELTR shows promising results. In the future, we plan to improve the pipeline as follows: (I) enhance robustness of segmentation against background structures; (II) improve the tracking energy formulation by incorporating the result of probabilistic mitosis detection; (III) devise a visualization method for efficient exploration of very large tree structures with smoothly synchronized viewing of the raw data and the segmentation.

6. REFERENCES

[1] P. J. Keller, A. D. Schmidt, J. Wittbrodt, et al., "Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy," *Science*, vol. 322, no. 5904, pp. 1065 – 1069, 2008.

[2] J. E. Sulston, E. Schierenberg, J. G. White, et al., "The embryonic cell lineage of the nematode caenorhabditis elegans," *Dev. Biol.*, vol. 100, no. 1, pp. 64–119, 1983.

[3] O. Al-Kofahi, R. J. Radke, S. K. Goderie, et al., "Automated Cell Lineage Construction," *Cell Cycle*, vol. 5, no. 3, pp. 327–335, 2006.

[4] K. Li, E. D. Miller, M. Chen, et al., "Cell population tracking and lineage construction with spatiotemporal context," *Med. Image Anal.*, vol. 12, no. 5, pp. 546–566, 2008.

[5] N. Olivier, M. A. Luengo-Oroz, L. Duloquin, et al., "Cell lineage reconstruction of early zebrafish embryos using label-free microscopy," *Science*, vol. 329, no. 5994, pp. 967 – 971, 2010.

[6] Y. Al-Kofahi, W. Lassoued, W. Lee, et al., "Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images," *IEEE T. Biomed. Eng.*, vol. 57, no. 4, pp. 841 –852, 2010.

[7] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[8] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE T. Patt. Anal. Mach. Int.*, vol. 23, no. 11, pp. 1222–1239, 2001.

[9] C. Zanella, M. Campana, B. Rizzi, et al., "Cells Segmentation From 3-D Confocal Images of Early Zebrafish Embryogenesis," *IEEE T. Image Proc.*, vol. 19, no. 3, pp. 770–781, 2010.

[10] X. Lou, U. Köthe, J. Wittbrodt, et al., "Accurate Reconstruction of Digital Embryo Volume with Multi-object Shape Regularization," *Med. Image Anal.*, 2010, in preparation.