2024 IEEE International Symposium on Biomedical Imaging (ISBI) | 979-8-3503-1333-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISBI56570.2024.10635875

A JOINT SCALE ANALYSIS AND MACHINE LEARNING FRAMEWORK FOR CELL DETECTION AND SEGMENTATION IN TIME LAPSE MICROSCOPY

Nagasoujanya Annasamudram *Sokratis Makrogiannis*

Division of Physics, Eng., Math. and Comp. Science, Delaware State Univ., Dover, DE 19901, USA

ABSTRACT

Cell segmentation is a crucial step for understanding cell mechanisms, and behaviors and for analyzing them with significant applications in disease modeling, personalized medicine, and drug development. In this work, we propose an automated system for cell segmentation in time-lapse microscopy. This work seeks space-time interest points in multiple scales and performs spatio-temporal scale selection in image sequences. Spatio-temporal features drive segmentation and identification of cells by a multilayered neural net. We validated our method on datasets of image sequences of live cells and reference masks from the Cell Tracking Challenge (CTC) consortium. Our methodology produced promising segmentation results over multiple test image sequences. The code is available at `https://github.com/smakrogi/CSTQ_Pub`.

Index Terms— segmentation, time-lapse series, spatio temporal features, scale selection, neural net

1. INTRODUCTION

Study of living organisms by cell identification, quantification and characterization using imaging techniques are emerging research areas in biological and medical studies [1,2]. Recent developments in time-lapse microscopy enable the observation and quantification of cell-cycle progression, cell migration, and growth control.

Cell segmentation and tracking methodologies involve the tasks of preprocessing, cell segmentation and motion tracking [3–6]. One of the challenging aspects of cell analysis is to develop methods that can provide good quality and accuracy on segmentation for different modalities of cell imaging. Previous works include cell nuclei detection [7,8] and hybrid models of deep learning methods and image processing [9].

In this work, we introduce a system for cell detection and segmentation in time-lapse image sequences. We compute motion activity measures, inspired from the computer vision fields of video processing, blob detection, and scale selection [10, 11]. We generate multi-scale feature descriptors for the each of three consecutive frames. We introduce a technique for selecting automatically a scale with high contrast and least background noise, inspired from techniques of spatio-temporal point-of-interest detectors [11]. The computed features drive the cell detection and segmentation. A multi-layered neural net is applied at the superpixel level to separate foreground from background.

We validate our model with datasets from the Cell Tracking Challenge consortium [12]. We evaluated our results on their ground truth masks, which are human-made reference annotations, agreed by experts. Our method achieved satisfactory results for 2-D + t datasets of the Cell Tracking Challenge. Using segmentation and detection measures for performance evaluation, we obtained DSC values of at least 0.80 for all test sequences.

2. METHODOLOGY

2.1. Multi-scale Interest Points, Scale Selection and Superpixels

2.1.1. *Spatio-temporal Anisotropic Diffusion*

In this stage we solve a system of three coupled PDEs applied to each frame and its direct temporal neighbors. The goal of this stage is to smooth the background while preserving the spatio-temporal discontinuities of cells in the frame and to produce multiple scales of visual representation. More specifically, given 3 consecutive frames of the sequence at time points $\tau = \{t-1, t, t+1\}$, we define a system of three coupled PDEs for each frame.

$$\begin{aligned}\frac{\partial I(i,j,\tau,s)}{\partial s} =& g(|\nabla I(i,j,\tau,s)|)\cdot \Delta I(i,j,\tau,s)\\ &+ \nabla g(|\nabla I(i,j,\tau,s)|)\cdot \nabla I(i,j,\tau,s) \quad (1)\end{aligned}$$

We apply *initial condition* $I(i,j,\tau,0) = I_0(i,j,\tau)$ and *boundary condition* $\frac{\partial I}{\partial \vec{n}} = 0$ on $\partial\Omega \times \partial T \times (0,S)$. In equation (1), $g(\cdot)$ is the edge stopping function and s is the scale index.

We iteratively generate multi-scale diffused frame maps. The diffused frame maps are utilized in the subsequent stages of spatio-temporal feature detection, segmentation, and foreground/background separation.

2.1.2. Spatio-temporal Feature Maps and Motion Descriptors

The goal is to detect blob-like structures and other keypoint types in the spatio-temporal domain. We expect that the use of non-linear diffusion models will enable the detection of non-spherical features that may correspond to non-circular cell shapes. In addition to this step, we also estimate motion displacement field vectors of the moving regions using Combined Local Global Optical Flow Estimation (CLGOF) [13].

Spatio-temporal Moments These are the spatio-temporal second moments [14] computed as an extension of Harris corner detector. This approach computes the structure tensor

$$M = w(\cdot, \sigma_i^2, \tau_i^2) * \begin{pmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_y I_x & I_y^2 & I_y I_z \\ I_z I_x & I_z I_y & I_z^2 \end{pmatrix}, \quad (2)$$

where $w(\cdot, \sigma_i^2, \tau_i^2)$ denotes a smoothing kernel with integrating factors σ_i^2, τ_i^2. Then, it uses the Harris corner detection criterion S to identify the points of interest

$$S = det(M) - \kappa \cdot trace^3(M). \quad (3)$$

Spatio-temporal Hessian This approach computes the Hessian matrix H in the space-time domain

$$H(\cdot, \sigma^2, \tau^2) = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix}. \quad (4)$$

It measures the strength the interest points using the determinant magnitude, $S = |det(H)|$.

Spatial Hessian on Temporal Derivative This approach computes the spatial Hessian matrix on first-order temporal derivatives I_t that we denoted by H_t. It measures the strength the interest points using the determinant $S = |det(H_t)|$.

Automated Scale Selection

The goal of the automatic scale selection stage is to identify the diffused map that best represents the feature descriptors of the foreground regions and limits the noise level. In this stage, we form multi-scale feature descriptors in spatio-temporal domain defined by the regional maxima across all generated scales. These feature vectors are then normalized so that the algorithm can automatically choose a scale of the diffused maps.

We utilize the second order derivative of the spatio-temporal features over the generated scales as the criterion for scale selection similarly to [11]. The diffused frame and the features from the selected scale, drive the subsequent stages of cell detection and segmentation.

Watershed and Key-point Based Segmentation

In this step we fuse the regional maxima of the feature maps with regional minima of the edge map to form markers that drive watershed segmentation. We first invert the stochastic map produced by Parzen density estimation to form regions separated by spatio-temporal discontinuities. We identify maxima of the keypoints and use them as markers along with the edge map minima.

2.2. Machine Learning-based Foreground/Background Separation

We developed a fully connected multi-layered neural network to classify the watershed superpixels into cell (foreground) or non-cell (background) classes. The layer structure of the neural net contains $9 \times 12 \times 6 \times 2$ nodes, where each fully connected layer is followed by a relu activation module. We used the cross-entropy loss function and LBFGS optimization to train the network. The training/test vectors consist of the regional areas, and the regional averages and standard deviations of the diffused map intensities, spatio-temporal descriptors, motion displacement vector magnitudes, and the spatio-temporal gradient magnitudes.

Because the class of non-cells typically has many more samples than the cell class, we developed an unsupervised learning-based stratification approach. We employed DBSCAN [15, 16] to cluster samples in the original feature space. We then randomly under-sampled the most populous cluster to equalize its size with that of the second most populous cluster. This step reduces class imbalance in training data while preserving the structure of the data as estimated by nonparametric clustering. The network learns the relationships among the regional features and separates foreground and background regions without user intervention in the test sequences.

2.3. Local Intensity Clustering Level Sets with Bias Field Estimation (LSE-BF)

We employ region based energy minimizing level set models to refine the delineation of the cells that were detected in the previous stage of foreground/background separation. The property of intensity inhomogeneity is used to get a local clustering criterion in neighborhood of each point in the image [17]. This local criterion is then used to obtain a global criterion in the neighborhood center of the image. This criterion is used to compute the energy minimized by level set functions that partition the image domain and to estimate the bias field.

2.4. Cell Cluster Separation

Adherent cells form clusters that sometimes are detected as a single cell. In this stage we identify and separate the cell clusters. We identify cell cluster candidates using morphological characteristics based on the solidity of the detected regions. We then compute the signed distance map of the binary cell map and use the distance map boundaries to divide the cell clusters.

3. EXPERIMENTS

3.1. Dataset Description

We evaluated our method on Cell Tracking Challenge datasets of 2D time-lapse live cell sequences. We employed 3 datasets of fluorescence (Fluo) microscopy and 1 phase-contrast (PhC) microscopy for training and testing. Each of the datasets includes two sequences labeled as 01 and 02, along with reference masks. We tested our algorithm on Fluo-C2DL-MSC (MSC), which has low resolution cytoplasm of rat mesenchymal stem cells, Fluo-N2DL-HeLa (HeLa), which contains low resolution nuclei of cervical cancer cells, Fluo-N2DH-GOWT1 (GOWT1) that consists of high resolution nuclei of mouse embryonic stem cells, and PhC-C2DH-U373 (U373) which has high resolution cytoplasm of glioblastoma-astrocytoma U373 cells. The datasets differ in noise level, cell density, number of cells leaving and entering the field of view, resolution, and mitotic events.

3.2. Evaluation Methods

We trained our classifier on the first sequence of the dataset and evaluated its performance on the second sequence. Our methodology was evaluated for accuracy of segmentation against CTC reference data.

The CTC datasets include two sets of reference segmentation masks for evaluation, named gold standard corpus or gold-truth (GT), and silver standard corpus or silver-truth (ST). GT reference sequences were manually segmented by experts with background in biology at three different institutions. On the other hand, the ST reference sequences are computer-generated annotations from the top results submitted by former participants.

The CTC training dataset serves as a validation set for measuring segmentation accuracy by the Dice Similarity Coefficient (DSC), Jaccard index, SEG measure (SEG), and DET measure (DET). We denote by R_s the set of all binary cell regions delineated by our cell segmentation method, while R_{Ref} is the set of cell pixels from the reference region.

SEG measure – It measures the Jaccard index between test and reference data, for each cell that has a reference mask. Then the method sets to zero the indices of cells for which $|R_s \cap R_{Ref}| \leq 0.5 \cdot |R_{Ref}|$. It finally computes the average of all the individual indices to yield SEG.

DET measure – This is a cell detection accuracy measure. DET is evaluated by comparing the nodes of the acyclic graphs generated by the tested algorithm with nodes of the acyclic graphs of the reference masks. It is defined as

$$DET = 1 - min(AOGM - D, AOGM - D0)/AOGM - D0. \tag{5}$$

Here *AOGM* denotes Acyclic Oriented Graph Measure for detection, and *AOGM-D* is the cost of transforming a set of nodes provided by the cell segmentation method into the set of reference nodes. *AOGM-D0* is the cost of creating the set of reference nodes.

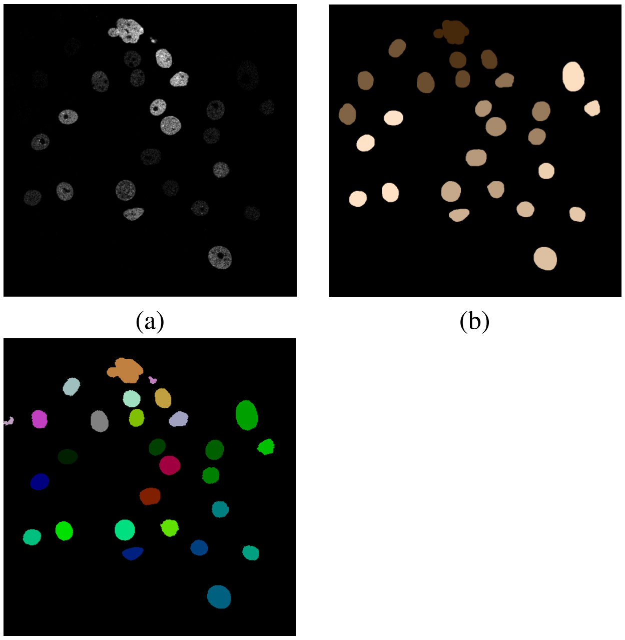

Fig. 1. Segmentation comparisons for Fluo-N2DH-GOWT1-02. (a) Intensity frame, (b) GT reference segmentation, (c) segmentation by the proposed method.

4. RESULTS

We report DSC and Jaccard segmentation accuracy results in Table 1, and SEG and DET measures in Table 2. We display examples of segmentation results and reference masks in Figures 1 and 2. We utilized the gold standard corpus GT and silver standard corpus ST reference masks in our experiments. The accuracy scores computed are for the second sequence in the dataset that was not used for training the neural net.

Table 1. DSC and $Jaccard$ results of $CSTQ$ $v3.0$ versus ST

Dataset	DSC	Jaccard
Fluo-C2DL-MSC-02	0.837	0.729
Fluo-N2DL-HeLa-02	0.903	0.823
Fluo-N2DH-GOWT1-02	0.947	0.900
PhC-C2DH-U373-02	0.799	0.672

We have compared the results of our current method denoted by $CSTQ$ $v3.0$ to previous versions $CSTQ$ $v2.0$ and $v2.9$. $CSTQ$ $v2.0$ implements a probability density model for foreground and background detector, no scale selection, and no motion descriptors for segmentation [6]. We also report scores of $CSTQ$ $v2.9$, that employs a non parametric

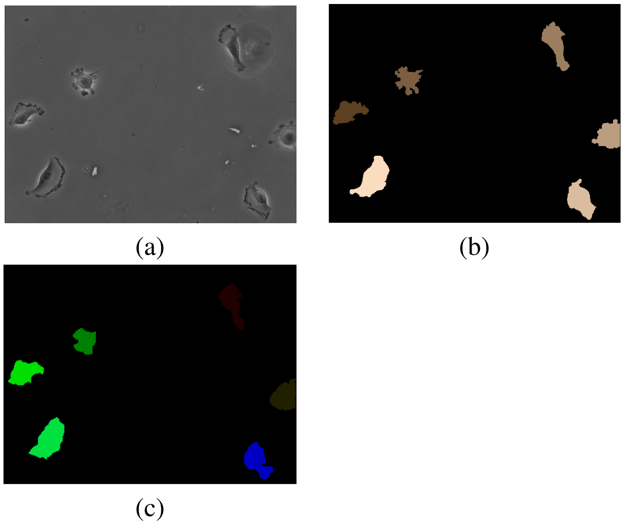

Fig. 2. Segmentation comparisons for PhC-C2DH-U373-02. (a) Intensity frame, (b) GT reference segmentation, (c) segmentation by the proposed method.

Table 2. SEG measure results versus ST, SEG measure and DET measure results versus GT

Version	Dataset	SEG ST	SEG GT	DET GT
$CSTQ\ v2.0$	MSC-02	0.647	0.594	0.734
$CSTQ\ v2.9$	MSC-02	0.782	0.691	0.838
$CSTQ\ v3.0$	MSC-02	0.770	0.694	0.755
$CSTQ\ v2.0$	HeLa-02	0.735	0.646	0.830
$CSTQ\ v2.9$	HeLa-02	0.853	0.802	0.964
$CSTQ\ v3.0$	HeLa-02	0.855	0.801	0.968
$CSTQ\ v2.0$	GOWT1-02	0.861	0.853	0.926
$CSTQ\ v2.9$	GOWT1-02	0.903	0.902	0.876
$CSTQ\ v3.0$	GOWT1-02	0.906	0.907	0.876
$CSTQ\ v2.0$	U373-02	0.576	0.609	0.607
$CSTQ\ v2.9$	U373-02	0.681	0.684	0.844
$CSTQ\ v3.0$	U373-02	0.677	0.687	0.832

likelihood density model for foreground/background separation and no motion descriptors for segmentation. Our method was developed in the Matlab framework and built as a single command line binary file. The $\mu \pm \sigma$ estimate of computation time per frame over the test datasets is 89.9 ± 74.51 seconds.

5. DISCUSSION AND CONCLUSION

The experiments show that our method produces promising results and performs well in the presence of intensity homogeneity in the sequences.

The algorithm was validated on 4 real datasets. We calculated the DSC and Jaccard scores in Table 1 using the ST reference masks, because ST reference sequences have dense annotations, i.e., annotations for all cells in all frames for a given dataset. This is in contrast to the sparse GT reference masks, where only certain cells in selected frames were annotated. We computed the SEGMeasure for our results using both GT reference masks and ST reference masks. We report SEGMeasure and DETMeasure results versus GT in Table 2, and SEG results versus ST and GT in Figure 3.

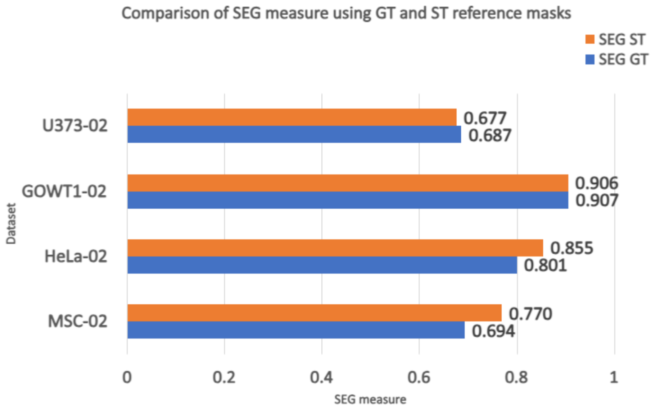

Fig. 3. Performance comparison versus GT and ST reference masks.

When comparing the SEG and DET measures, we observe that the detection rates are higher than the segmentation rates. This is expected, because in segmentation the goal is to delineate the boundary cells with accuracy.

We observed that incorporation of spatial and temporal feature detectors has improved the results for watershed segmentation. Furthermore, the neural net classifier improves the separation of foreground regions from background regions. The learning of relationships among the region descriptors has enabled the classifier to identify foreground regions with increased resilience to noise. On the other hand, detection of low contrast cells for GOWT1-02 and HeLa-02 datasets is an area of improvement for our method. Similarly, delineating thin structures for MSC-02 may be challenging, due to variations in pixel intensities, presence of noise, and resolution limitations.

In conclusion, our research work integrates machine learning methods with automated scale selection of the differential spatio-temporal features along with optical flow estimation for the motion of the cells. This is a promising research area that seems to be applicable to the problem of cell segmentation and detection. Future goals are to closely investigate deep learning spatio-temporal feature detection techniques in order to improve the robustness and generalizability of cell segmentation and to evaluate the algorithm on additional cell microscopy techniques.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted using data publicly available at http://celltrackingchallenge.net/. Ethical approval was not required as confirmed by the license attached with the open-access data.

7. ACKNOWLEDGMENTS

This research was supported by the Army Research Office under grant #W911NF2010095, National Institute of General Medical Sciences of the National Institutes of Health (NIH) award #SC3GM113754, #1U54MD015959-01A1 and the National Science Foundation under award #2234871. Authors have no conflict of interest.

8. REFERENCES

[1] Roland Eils and Chaitanya Athale, "Computational imaging in cell biology.," *J Cell Biol*, vol. 161, no. 3, pp. 477–481, May 2003.

[2] Erik Meijering, Oleh Dzyubachyk, and Ihor Smal, "Methods for cell and particle tracking.," *Methods Enzymol*, vol. 504, pp. 183–200, 2012.

[3] Fatima Boukari and Sokratis Makrogiannis, "Spatio-temporal diffusion-based dynamic cell segmentation," in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*, 2015, pp. 317–324.

[4] Klas E G. Magnusson, Joakim Jalden, Penney M. Gilbert, and Helen M. Blau, "Global linking of cell tracks using the viterbi algorithm.," *IEEE Trans Med Imaging*, vol. 34, no. 4, pp. 911–929, Apr 2015.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015, pp. 234–241.

[6] Sokratis Makrogiannis, Nagasoujanya Annasamudram, Yibing Wang, Hector Miranda, and Keni Zheng, "A system for spatio-temporal cell detection and segmentation in time-lapse microscopy," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 2266–2273.

[7] Tomáš Sixta, Jiahui Cao, Jochen Seebach, Hans Schnittler, and Boris Flach, "Coupling cell detection and tracking by temporal feedback," *Machine Vision Applications*, vol. 31, no. 4, pp. 1–18, 2020.

[8] Hongming Xu, Cheng Lu, Richard Berendt, Naresh Jha, and Mrinal Mandal, "Automatic nuclei detection based on generalized laplacian of gaussian filters," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 826–837, 2016.

[9] Filip Lux and Petr Matula, "DIC image segmentation of dense cell populations by combining deep learning and watershed," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 236–239.

[10] Yanshan Li, Rongjie Xia, Qinghua Huang, Weixin Xie, and Xuelong Li, "Survey of spatio-temporal interest point detection algorithms in video," *IEEE Access*, vol. 5, pp. 10323–10331, 2017.

[11] Tony Lindeberg, "Spatio-temporal scale selection in video data," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 4, pp. 525–562, 2018.

[12] Martin Maska, Vladimir Ulman, Pablo Delgado-Rodriguez, and et al, "The cell tracking challenge 10 years of objective benchmarking," *Nature Methods*, vol. 20, pp. 1010–1020, 2023.

[13] Fatima Boukari and Sokratis Makrogiannis, "Automated cell tracking using motion prediction-based matching and event handling," *IEEE/ACM Transactions on Computational Biology Bioinformatics*, vol. 17, no. 3, pp. 959–971, 2020.

[14] Ivan Laptev, "On space-time interest points," *International journal of computer vision*, vol. 64, no. 2, pp. 107–123, 2005.

[15] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery in Databases and Data Mining*, 1996, pp. 226–231.

[16] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu, "Dbscan revisited, revisited: why and how you should (still) use DBSCAN," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.

[17] Chunming Li, Rui Huang, Zhaohua Ding, J Chris Gatenby, Dimitris N Metaxas, and John C Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri," *IEEE transactions on image processing*, vol. 20, no. 7, pp. 2007–2016, 2011.