Imperial College London

Department of Computing

Automatic Cell Tracking in Noisy Images for Microscopic Image Analysis

by

Pedro Damian Kostelec

September 2014

Supervised by Ben Glocker

Contents

6 Experimental results $\boxed{\text{IN PROGRESS}}$

In this chapter we quantitatively and qualitatively analyse the performance of the automatic cell detector and tracker. Although some evaluation of the performance of the detection method is performed by the original authors in [4] it is useful to see how the method performs on the studied datasets in order to understand how much of the tracking accuracy is lost due to cells missed by detection module. First, in section 6.1 we evaluate the performance and computation time of the cell detector and in section 6.2 those of the cell tracker. Finally, in section 6.3, we explore the limitations of the methods and in section 6.4 summarize the results.

> See the cell population tracking and linear construction with spationtemporal ocntet by Kang et al for a good results section

6.1 Cell detector $\boxed{\text{IN PROGRESS}}$

In this section we evaluated the performance of the automatic cell detection module. First, we introduce the performance metrics used to evaluate the accuracy of the cell detector. Then we present detection accuracy results. To evaluate the accuracy and generalizability of the detection module we perform two sets of experiments. First, we train the cell detector on a number of frames from each individual dataset, and measure the accuracy on the same dataset. Second, we train the detector on combinations of datasets in order to judge the performance degradation due to the learning on the wider types of cells. Because of the varying size of the cells in the datasets, and the varying brightness of the cells, we expect that such a trained detector will detect a larger number of cells than are actually present, sometimes mistakenly detecting small artefacts in the background as cells. Finally, we compute the average detection time per frame for each dataset.

The aim of this research was to develop an automatic cell detection and tracking pipeline that would require as little manual work as possible. This implies that a balance between accuracy and amount of manual work had to be established. There is also an direct relationship between accuracy and computation time. In order to reduce the amount of manual work we decided that the cell detection module should perform well on all the tested datasets without any manual adjustment of parameters. This consequences of this decision are twofold:

1. The features computed on the candidate cell regions are the same for all datasets and have been presented in section 3.5. Although some datasets could be analysed faster or more accurately with a different subset of features, using the same features for all dataset eliminates the complicated feature selection process for the user and makes the system generalizable to a

large number of different cell types.

2. The parameters of the MSER detector should be adequately set to perform well on all datasets. This means that the MSER detector should be able to detect cells of varying size and contrast in the different datasets. The consequence of this limitation for datasets with large cells and some background noise is that a potentially much larger number of candidate regions will be detected than necessary. Since each candidate region has to be evaluated this results in an increased computation time.

We were able to identify features that compute in an acceptable time for all these datasets (see section 3.5). However, it should be noted that in the case of testing the detector on very large datasets with thousands of frames, some adjustments of the parameters could result in a significant reduction in computation time and increased accuracy.

6.1.1 Performance metrics NEW

We measure the performance of the cell detector in terms of precision and recall. The metrics are defined in terms of:

True Positive instances (TP) are candidate cell regions that are manually annotated as cells and the detector successfully classified as cells.

False Positive instances (FP) are candidate cell regions that are not manually annotated as cells, but the detector incorrectly classified them as cells.

False Negative instances (FN) are candidate cell regions that are manually annotated as cells, but the detector incorrectly classified.

We then define precision as:

$$\mathrm{PRE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}},$$

and recall (also known as sensitivity) as:

$$\mathrm{REC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \ .$$

6.1.2 Detection accuracy NEW

As mentioned previously, we performed two different experiments to measure the performance of the cell detection module. The first experiment consisted of training and testing of the algorithm on the same dataset (the training set was 70%, the testing set 30% of the entire dataset). In this

way we measured for each dataset the maximum precision and recall values we could expect from the algorithm. In the second experiment the training was performed on combined datasets. The goal of this experiment was to observe how well the algorithm is able to generalize, and still return acceptable results.

Training and testing on single dataset

The training was performed on 70% of the manually dot-annotated images from each dataset, and tested on the remaining 30%. Table 6.1 displays the computed precision and recall values, together with the total number of cells that were annotated in each dataset.

> Repeat the measurements with the best feature selection

Dataset	Precision [%]	Recall [%]
A	58.3	45.4
B	89.8	93.6
C	82.5	81.6
D	81.9	90.5
E	82.3	94.9

Table 6.1: Precision and recall values for the cell detector trained on each dataset individually.

> Discuss: which better, why(better images? less noise?), what could be improved...

Training on composite dataset

In this second experiment we were instered in measuring how well the algorithm is able to generalize when it is trained on a larger, combined dataset. For this purpose, several of the datasets were grouped, and the detector was trained to recognize cells of all types. Then, the results were tested individually on the 30% of the annotated images from each dataset, that were not used for training. Table 6.2 shows the results.

> Repeat the measurements with the best feature selection

> Try different dataset combinations based on the observations... similar datasets

Combined dataset	Precision [%]	Recall [%]
A+B+C+D+E	A: 60, B: 60, C: 60, D: 60, E: 60	A: 60, B: 60, C: 60, D: 60, E: 60

Table 6.2: Precision and recall values for the cell detector trained on combined datasets.

> Add discussion: generally how much precision/recall we loose, which dataset seems to suffer less/more, why do they suffer less/more (have fewer cells to be trained on, like like noise in another dataset)

> Here show an example of detection in a dataset so the use has a better idea of the difficulty of connecting the dots, like this image below...maybe create one for each dataset

Figure 6.1 displays a temporal view of the detected cells. The vertical axis represents the frame of the sequence. The figure clearly shows that "cell tracks" are clearly discernible, even if the number of outliers is significant. For the tracking module it is better to have a higher recall than precision, as outliers can be much more easily discarded than segmented tracks linked.

6.1.3 Computations time NEW

In order to positionally track cells in image sequence the computation time of extracting the cell position from each frame is just as important as the accurate identification of cells. Table 6.3 displays the average detection times per frame for each dataset. The detection time is dependent on the number of candidate cell regions extracted and their dimensions.

> Explain hardware and software, say 1 core... can use more

> It might be useful to show the comparison with the old detector if I have the time

> Measure the speed of detection in images of different sizes, and different number of cells

> Explain what the candidate-annoation ration is, and what it tells. Also the ocmputations timer per candidate vary from 0.0086-0.0226

Dataset	Time per frame [s]	Time per annotated cell [s]	Candidate-Annotation ratio
A	0.9224	0.5854	26:1
B	1.5899	0.2015	16:1
C	1.1214	0.0586	7:1
D	1.5215	0.1454	18:1
E	0.6016	0.0853	9:1

Table 6.3: Average computation times per frame and annotated cell using a single worker.

> Discuss which is faster, try to explain why– less noise – less candidate cells...

6.2 Cell tracker NEW

> Define the different measures of accuracy

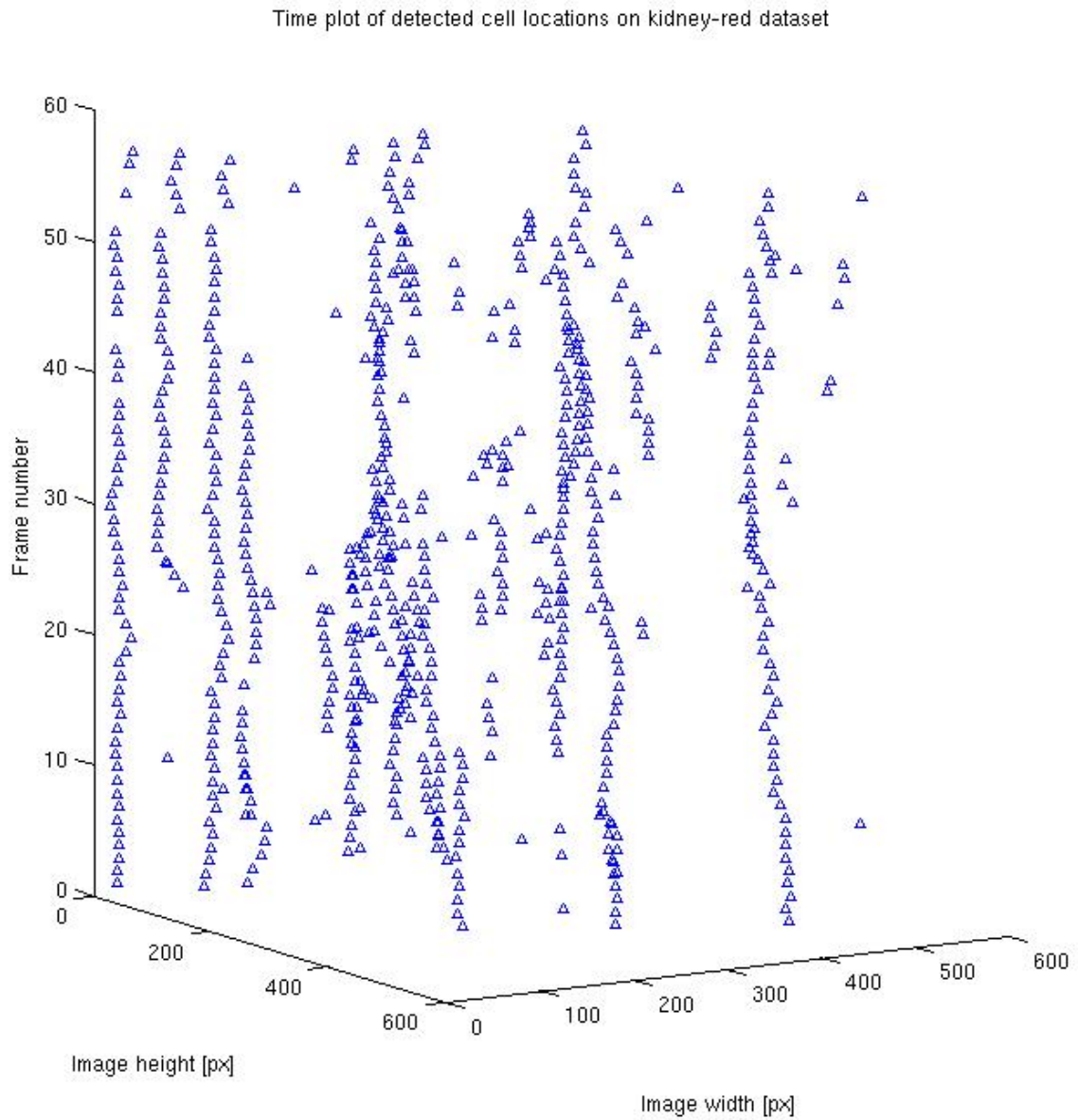Time plot of detected cell locations on kidney-red dataset

Figure 6.1: Cells detected over 60 consecutive frames are visualized as a time series. The vertical axis corresponds to the frames. Even in this raw detection data, it is possible to see the tracks of some of these cells.

6.2.1 Performance metrics NEW

TODO: test detector on full dataset, but train it on ALL (notjust70%) of the annotation TODO: train tracker on 70% of the trajectories, test on rest

great Metrics: Research Article, Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics

- Explain how the testing data was generation. from annotation to mapped deteciton to genreated tracklets

6.2.2 Tracking accuracy NEW

- trained on single dataset

- for each dataset explain how the parameters were setup

- trained on combined dataset

6.2.3 Computation time NEW

Meause the speed of generating tracks, as a measure of per 1, 100, 1000 frames, depending on the number of tracks

6.3 Limitations and areas of improvement NEW

Answer: what, why, how to improve in future

- display examples where the tracker did not perform well, and anlyse why. Suggest possible improvement. - detection training: only first few frames of datasets, not random – expect to detect later frames worse - testing on only long datasets: no data on short datasets. diffucult to train (what to link?), difficult to annotate - speed of detector. Reduce number of hypothesis

6.4 Summary NEW

Brief review of accuracy... whether it is comparable to other methods in literuature review Whether is could be improved in the future... how much

Appendices

Bibliography

[1] P. K. Elzbieta Kolaczkowska, "Neutrophil recruitment and function in health and inflammation," 2013. 7

[2] J. Pillay, I. den Braber, N. Vrisekoop, L. M. Kwast, R. J. de Boer, J. A. M. Borghans, K. Tesselaar, and L. Koenderman, "In vivo labeling with 2h2o reveals a human neutrophil lifespan of 5.4 days," *Blood*, vol. 116, no. 4, pp. 625–627, 2010. 7

[3] P. S. Tofts, T. Chevassut, M. Cutajar, N. G. Dowell, and A. M. Peters, "Doubts concerning the recently reported human neutrophil lifespan of 5.4 days," *Blood*, vol. 117, no. 22, pp. 6050–6052, 2011. 7

[4] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, "Learning to detect cells using non-extremal regions," in *Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I*, MICCAI'12, (Berlin, Heidelberg), pp. 348–356, Springer-Verlag, 2012. 8, 9, 11, 17, 18, 19, 20, 21, 22, 45, 49

[5] Y. Chen, K. Biddell, A. Sun, P. Relue, and J. Johnson, "An automatic cell counting method for optical images," in *[Engineering in Medicine and Biology, 1999. 21st Annual Conference and the 1999 Annual Fall Meetring of the Biomedical Engineering Society] BMES/EMBS Conference, 1999. Proceedings of the First Joint*, vol. 2, pp. 819 vol.2–, Oct 1999. 10

[6] X. Chen, X. Zhou, and S.-C. Wong, "Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy," *Biomedical Engineering, IEEE Transactions on*, vol. 53, pp. 762–766, April 2006. 10, 13

[7] L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *Image Processing, IEEE Transactions on*, vol. 2, pp. 176–201, Apr 1993. 10

[8] J. Serra, *Image Analysis and Mathematical Morphology*. Orlando, FL, USA: Academic Press, Inc., 1983. 10

[9] D. Mukherjee, N. Ray, and S. Acton, "Level set analysis for leukocyte detection and tracking," *Image Processing, IEEE Transactions on*, vol. 13, pp. 562–572, April 2004. 11, 13

[10] C. Tang, Y. Wang, and Y. Cui, "Tracking of active cells based on kalman filter in time lapse of image sequences of neuron stem cells." 11, 14

[11] D. Xu and L. Ma., "Segmentation of image sequences of neuron stem cells based on level-set

algorithm combined with local gray threshold.," Master's thesis, Harbin Engineering University, 2010. 11

[12] C. Arteta, V. S. Lempitsky, J. A. Noble, and A. Zisserman, "Learning to detect partially overlapping instances.," in *CVPR*, pp. 3230–3237, IEEE, 2013. 11, 12, 19

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proceedings of the British Machine Vision Conference*, pp. 36.1–36.10, BMVA Press, 2002. doi:10.5244/C.16.36. 11

[14] T. Joachims, T. Finley, and C.-N. J. Yu, "Cutting-plane training of structural svms," *Mach. Learn.*, vol. 77, pp. 27–59, Oct. 2009. 11

[15] R. Bise, T. Kanade, Z. Yin, and S. il Huh, "Automatic cell tracking applied to analysis of cell migration in wound healing assay," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pp. 6174–6179, Aug 2011. 12, 25

[16] S. Huh, *Toward an Automated System for the Analysis of Cell Behavior: Cellular Event Detection and Cell Tracking in Time-lapse Live Cell Microscopy*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, March 2013. 12, 13

[17] D. House, M. Walker, Z. Wu, J. Wong, and M. Betke, "Tracking of cell populations to understand their spatio-temporal behavior in response to physical stimuli," in *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pp. 186–193, June 2009. 13

[18] B. Xu, M. Lu, P. Zhu, Q. Chen, and X. Wang, "Multiple cell tracking using ant estimator," in *Control, Automation and Information Sciences (ICCAIS), 2012 International Conference on*, pp. 13–17, Nov 2012. 14

[19] 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 (MIAAB)*, September 2007. 14

[20] A. Massoudi, D. Semenovich, and A. Sowmya, "Cell tracking and mitosis detection using splitting flow networks in phase-contrast imaging," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 5310–5313, Aug 2012. 14

[21] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, June 2008. 15, 27

[22] C. Huang, B. Wu, and R. Nevatia, "Robust object tracking by hierarchical association of detection responses," in *Computer Vision - ECCV 2008* (D. Forsyth, P. Torr, and A. Zisserman,

eds.), vol. 5303 of *Lecture Notes in Computer Science*, pp. 788–801, Springer Berlin Heidelberg, 2008. 15, 27

[23] R. Bise, Z. Yin, and T. Kanade, "Reliable cell tracking by global data association.," in *ISBI*, pp. 1004–1010, IEEE, 2011. 15, 25, 27, 30, 49

[24] H. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955. 15

[25] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761 – 767, 2004. British Machine Vision Computing 2002. 18

[26] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun, "Support vector machine learning for interdependent and structured output spaces," in *Proceedings of the Twenty-first International Conference on Machine Learning*, ICML '04, (New York, NY, USA), pp. 104–, ACM, 2004. 20

[27] K. Li, E. D. Miller, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell, "Cell population tracking and lineage construction with spatiotemporal context," *Medical Image Analysis*, vol. 12, no. 5, pp. 546 – 566, 2008. Special issue on the 10th international conference on medical imaging and computer assisted intervention - {MICCAI} 2007. 25, 51

[28] M. Looney, E. Thornton, D. Sen, W. Lamm, R. Glenny, and M. Krummel, "Stabilized imaging of immune surveillance in the mouse lung.," *Nature Methods*, vol. 8, no. 5, pp. 91–6, 2011-01-01 00:00:00.0. 36