AN AUTOMATIC OVERLAP-BASED CELL TRACKING SYSTEM

Joe Chalfoun, Antonio Cardone, Alden A. Dima, Michael Halter, Daniel P. Allen

National Institute of Standards and Technology (NIST)

ABSTRACT

In order to facilitate the extraction of quantitative data from live cell image sets, automated image analysis methods are needed. This paper presents an overlap-based cell tracking algorithm that has the ability to track cells across a set of time-lapse images based on the amount of overlap between cellular regions in consecutive frames. It uses the overlap to identify mitotic cells as well. This cell tracker is designed to be highly flexible, requires little user parameterization, and has a fast execution time. We demonstrate the performance of this tracker on NIH-3T3 mouse fibroblast cell line.

Index Terms— *overlap-based cell tracking, time-lapse cell imaging, cell motility, live-cell imaging*

1 INTRODUCTION

Automated microscopy has facilitated the large scale acquisition of live cell image data [1,3,4]. However, obtaining quantitative data related to single cell behavior requires image analysis methods that can accurately segment and track cells.

Many popular cell tracking techniques are based on complex probabilistic models. In [1] Gaussian probability density functions are used to characterize the selected tracking criteria. In [6] cells are tracked by fitting their tracks to a persistent random walk model based on mean square displacement. An important class of tracking techniques consists of level set methods [5,8]. They produce fairly accurate tracking results but are difficult to implement and are computationally expensive. The majority of the available techniques have a large number of parameters to adjust for every track. Therefore new, flexible techniques that can produce accurate tracking with small set of adjustable parameters are needed.

Tools for automated live cell microscopy have made it possible to routinely collect large sets of time-lapse images under a number of experimental conditions and record changes in the cellular response. Our experience shows that when acquiring time-lapse images at intervals ranging from 5 to 15 min, the movement of cultured mammalian cells between two consecutive frames will be relatively small. This means that between consecutive frames a typical cell will occupy nearly the same position. In order to effectively analyze large volumes of data (>10,000 images) an automated process is needed requiring little manual intervention and a simple, meaningful set of parameters.

The overlap-based cell tracking software developed at NIST was designed with this goal in mind. It tracks cells across a set of time-lapse images based on the amount of overlap between cellular regions in consecutive frames. It is designed to be highly flexible and suitable for use in a wide range of applications, requires little user interaction during the tracking process and has a fast execution time. The algorithm is likely to work effectively under conditions where the change in a cell's location from one frame to the next is relatively small. The core tracking algorithm is shown in Figure 1.

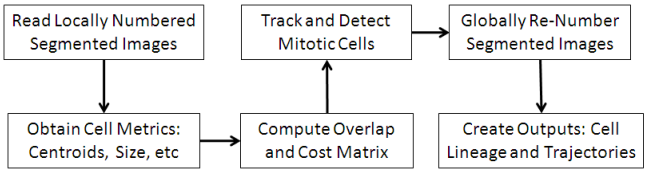

Figure 1- Core algorithm

In this paper, a general formulation of the motion tracking problem will be given, followed by a description of the input data and of the tracking criteria employed. The detection of mitotic cells will be described and some applications of the tracking software will be presented to further illustrate its capabilities. We will conclude with a summary of our results.

2 PROBLEM STATEMENT

Cell tracking techniques are used to obtain motion and life cycle behavior information about cells by following the cells of interest through multiple time sequential images. The cell tracking problem can be defined as: given a cell *A* from a current (source) image, identify the corresponding cell *B*, if any, in the subsequent (target) image. If cell *A* is tracked to *B*, then the two cells are the same cell at successive moments in time. This process involves examining all possible combinatorial mappings of the cells in a source image to the cells in the target image and finding the optimal mapping.

The input to the cell tracker is a series of segmented images derived from the raw microscopy data. In our case we used masks derived from phase contrast images of NIH-3T3 fibroblast shown in Figure 2. Many segmentation techniques exist in the literature; some are general purpose

and others are specific to a cell line and/or image acquisition parameters. The specifics of the segmentation algorithm used in this project will not be addressed here and in general the NIST tracking algorithm can be used with any segmentation algorithm. It is important to note, however, that the reliability of the tracking outcome is highly dependent on the accuracy of the segmentation.

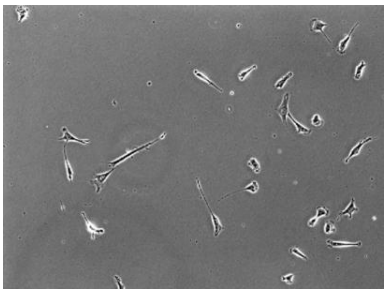

Figure 2- NIH-3T3 fibroblast phase contrast microscopy image

In what follows I_k refers to the k^{th} segmented image or mask with $k = 1,\ 2, \ldots, N$, where N is the total number of images. The segmentation process sets the background pixels in the mask to zero and pixels in a cellular region to a positive integer representing the cell number (Figure 3). The cell numbers are assigned to each segmented region starting at 1 and continuing incrementally until all segmented regions have been labeled. c_i^k is used to identify cell number i from the k^{th} image, with $i = 1, 2, \ldots, M_k$. M_k represents the total number of cells that are present in the k^{th} image.

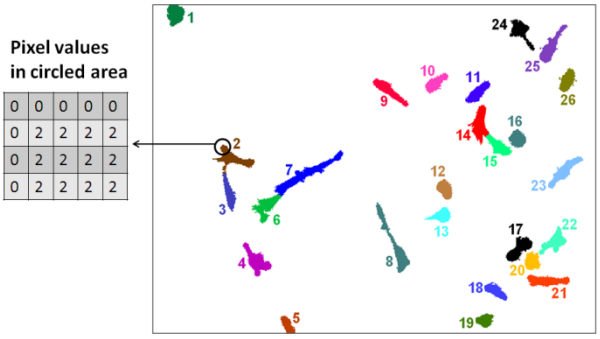

Figure 3- Segmented image of Figure 2

3 THE OVERLAP BASED TRACKING CONCEPT

The cell tracker computes a cost for each possible cell-to-cell mapping based on some simple tracking criteria. The cost value gives a measure of the probability that cell c_i^k from image I_k should be tracked to cell c_j^{k+1} in the subsequent image. The cost function has been defined so that the higher the cost value, the lower the probability that two cells should be identified as the same cell across frames.

Before describing the tracking criteria, consider two consecutive segmented phase-contrast images superimposed on top of each other as shown in Figure 4. Note that individual cells do not significantly change position between consecutive frames. This suggests that the number of common pixels (the overlap) between a pair of cells can be used as the principal measure of cost. If two cells share a large number of overlapping pixels, then they are most likely the same cell in different images. If more than two cells overlap, we employ additional criteria to further refine the cost. For this technique to work reliably, the images must be acquired at a sufficient rate to minimize cell movement between successive frames. If the interval between images is too long, the cells might migrate great distances and will exhibit little or no overlap.

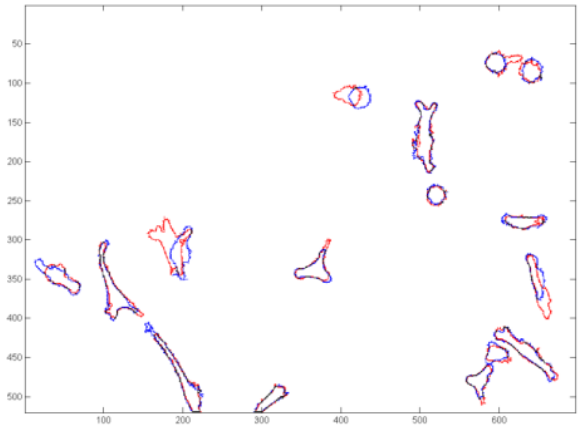

Figure 4- Two consecutive images superimposed

The cost function is defined as a sum of individual metrics. A mathematical statement of the cost function is:

$$d(c_i^k, c_j^{k+1}) = w_o \times O(c_i^k, c_j^{k+1}) + w_c \times \delta_c(c_i^k, c_j^{k+1}) + w_s \times \delta_s(c_i^k, c_j^{k+1})$$

where: w_o is the weight of the overlap term, O is an overlap metric, w_c is the weight of the centroid offset term, $\delta_c$is a centroid offset metric, w_s is the weight of the cell size term and δ_s is a cell size metric

The mathematical representation of the cost function carries desirable properties such as differentiability and the ease of including additional tracking criteria via new terms. The metrics are defined in a way that lower values indicate a higher probability that the source and target cells are the same cell. Each metric is normalized between 0 and 1. A value of zero denotes a perfect match between a pair of cells: all pixels overlap, the centroids are in the same location and cells have the same size. The weights provide flexibility and allow tailoring for different cell lines and image acquisition conditions. For example if the motion of cells between frames is low and cells overlap greatly between consecutive frames then w_o should be set to a high value. If the cell size changes little between two consecutive frames then a larger weight can be given for the size term.

3.1 The overlap metric

The overlap metric for a source/target pair is a measure of the number of pixels two cells have in common between two consecutive frames. It is computed using the formula:

$$O(c_i^k, c_j^{k+1}) = 1 - \left[\frac{n_o(c_i^k, c_j^{k+1})}{2}\left(\frac{1}{s_i^k} + \frac{1}{s_j^{k+1}}\right)\right]$$

where:
s_i^k = the size in pixels of the source cell
s_j^{k+1} = the size in pixels of the target cell
$n_o(c_i^k, c_j^{k+1})$ = number of common pixels to the two cells

3.2 The centroid metric

The centroid metric is a measure of the Euclidean distance between the centroids of the source and target cells between two consecutive frames. Let the width and height (in pixels) of a frame be represented by the symbols I_{width} and I_{height} and denote the centroid coordinates (in pixels) of cell i in frame k by the symbols (X_i^k, Y_i^k). The centroid metric for a source/target pair is computed as:

$$\delta_c(c_i^k, c_j^{k+1}) = \frac{\sqrt{\left(X_i^k - X_j^{k+1}\right)^2 + \left(Y_i^k - Y_j^{k+1}\right)^2}}{\sqrt{I_{height}^2 + I_{width}^2}}$$

3.3 The size metric

The size metric measures the relative difference between source and target cell size in two consecutive frames. It is computed as:

$$\delta_s(c_i^k, c_j^{k+1}) = \frac{\left|s_i^k - s_j^{k+1}\right|}{\max(s_i^k, s_j^{k+1})}$$

3.4 The cost matrix

For each pair of consecutive frames the cost function is computed for all possible mappings of source and target cells. The resulting costs form a cost matrix as shown below. The cost matrix is used to choose the optimal tracking solution for the corresponding pair of frames.

$$Cost_{l_k}^{l_{k+1}} = \begin{bmatrix} d(c_1^k, c_1^{k+1}) & \cdots & d(c_1^k, c_{M_{k+1}}^{k+1}) \\ \vdots & \ddots & \vdots \\ d(c_{M_k}^k, c_1^{k+1}) & \cdots & d(c_{M_k}^k, c_{M_{k+1}}^{k+1}) \end{bmatrix}$$

4 MITOSIS

A cell divides into two daughter cells by the process of mitosis. For adherent cells, the process typically lasts approximately 80min during which the cell rounds then undergoes mitosis and cytokinesis [2]. Figure 5 illustrates an example of a mother cell that goes into mitosis in the first frame and divides in two daughter cells in the next one. The two images superimposed (Figure 6) reveal that the mother cell has an overlap with both daughter cells.

The overlap-based cell tracker uses this information to detect mitotic cells between two consecutive images. In general, when mitosis happens, one mother cell c_m^k from image I_k, is tracked to the two daughter cells c_i^{k+1} and c_j^{k+1} in image I_{k+1}. The cell tracker detects mitosis as follows: 1) The cost matrix is minimized column-wise forming a one row tracking vector. The cell tracker uses this vector to capture mitotic cells by looking for pairs of target cells that are tracked to the same source cell. 2) The overlap of each daughter cell is compared against a user-defined mitosis threshold. The cell tracker will record the mitosis event only if both overlaps meet the threshold. 3) The cell roundness of all potential mother cells is checked before mitosis. If a mother cell does not meet a user-defined roundness threshold, the cell tracker ignores the division.

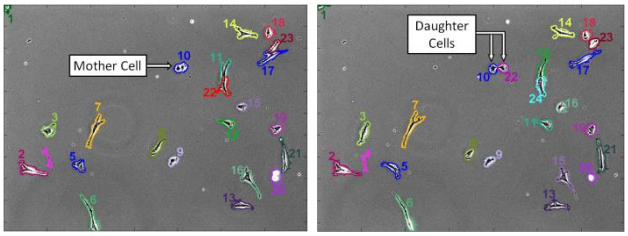

Figure 5- Example of a mitotic cell in two consecutive frames

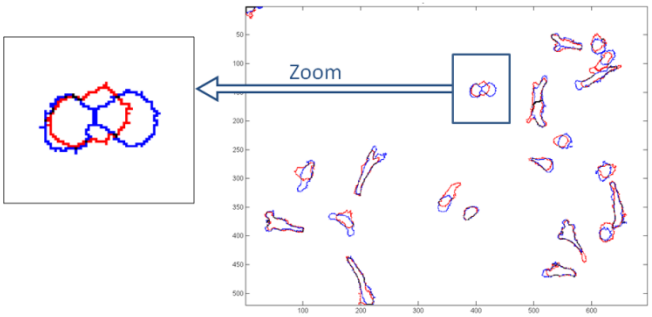

Figure 6- Superimposing image 1 (red) and image 2 (blue)

5 TRACKING SOLUTION AND OUTPUTS

After dealing with mitosis, the cell tracker will assign a track, when possible, between the remaining source cells and the remaining target cells. Tracks are assigned so that a source cell *A* can share a track with only one target cell *B* and vice versa. The unassigned source cells are considered dead (i.e., cells leaving the frame or mitotic mother cells) and the unassigned target cells are considered newborn cells (i.e., cells entering the frame from the borders or cells originating from mitosis). In order to achieve such a solution, the Hungarian algorithm [7] is applied on the cost matrix. By using this algorithm we are able to find an optimal solution that minimizes the sum of the above-defined tracking costs for the remaining cells after dealing with mitosis cases.

Once the individual cell mappings between consecutive frames have been computed, the frame-to-frame mappings are combined to produce a complete life cycle track of all the cells in the set of images. The sequential cell numbers

that were assigned by the segmentation process for each frame are replaced by unique track numbers that identify the movement of each cell over time across the entire set of images. Therefore a unique track number t_n will be associated to each uniquely identified cell, $n = 1, 2, \ldots, T$ where T represents the total number of unique cells found in the image set. The pixels in the images are relabeled to reflect the new track numbers such that when a given cell is assigned with a tracking number, the pixels from all images that belong to this cell will all have the same value. This is formally stated as follows.

$$if\ c_i^k \xleftrightarrow{t_n} c_j^{k+1} \Longrightarrow c_i^k = t_n = c_j^{k+1}$$
$$\Longrightarrow \forall x, y\ /\ p(x,y) \in (c_i^k\ \vee\ c_j^{k+1}), p(x,y) = t_n$$

Figure 7 shows the tracking results for two consecutive frames based on the following weight combination found by trial and error to be: $w_o = 2, w_c = 1$ and $w_s = 0.5$. The cells have been renumbered according to their global unique numbering.

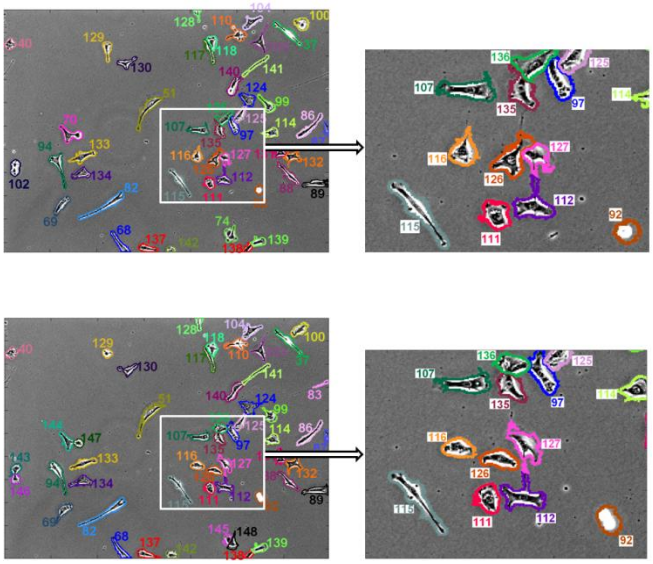

Figure 7- Tracking cells across frames 149-150

Figure 8 shows one output of the cell tracker, the cell lineage. Every line represents the life of a cell. The length of a cell line is proportional to the length of its lifetime (15 min interval between consecutive frames). The cell lineage supports statistical analysis of the cells e.g. the number of descendant generations of a cell, the length of the cell cycle. Figure 8 also shows the centroid trajectories in 3D. This helps determine the cell motility.

The average computation time for tracking 500 cells in a phase contrast NIH-3T3 set of 252 images (520x696 pixels) of a MATLAB implementation on a single core Pentium 3.4 GHz 3 GB RAM is 47 s. This translates to an average speed of 5.36 frames/s. For more details about the cell tracker and its performance, consult the following URL: http://www.nist.gov/itl/idus/compbio.

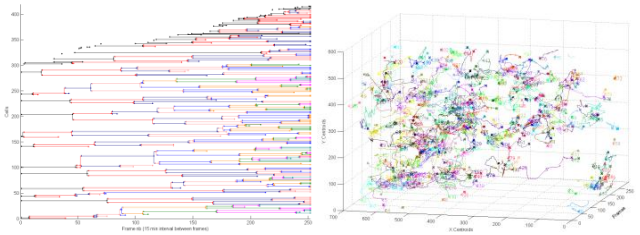

Figure 8- Cell lineage and 3D centroid trajectories

6 CONCLUSION

A highly flexible overlap-based cell tracking algorithm that requires little user parameterization and has a fast execution time was introduced. This cell tracker has the ability to track cells across a set of time-lapse images acquired at sufficiently high rates such that there is significant overlap between cellular regions in consecutive frames. This cell tracker is not dependent on any particular segmentation technique used to obtain the input data.

The overlap-based cell tracker's performance has been evaluated on a large set of NIH-3T3 fibroblast phase contrast microscopy images. Although the cell tracker has exceeded all our expectations, in the future we will formally compare manually segmented and tracked cells with those from the cell tracker as a validation of the underlying algorithm.

7 REFERENCES

[1] A. Bahnson, C. Athanassiou, D. Koebler, L. Qian, T. Shun, D. Shields, H. Yu, H. Wang, J. Goff, T. Cheng, R. Houck, and L. Cowsert, "Automated measurement of cell motility and proliferation," *BMC Cell Biology*, 6(1): 19, 2005.

[2] A. Bruce, B. Dennis, L. Julian, R. Martin, R. Keith, and D. W. James, "*Molecular Biology of the Cell 1*," Garland Publishing, New York, 19832.

[3] Davis PJ, Kosmacek EA, Sun Y, Ianzini F, Mackey MA, "The large-scale digital cell analysis system: an open system for nonperturbing live cell imaging," *J Microsc*; 228 (Pt 3):296-308, 2007.

[4] Gordon A, Colman-Lerner A, Chin TE, Benjamin KR, Yu RC and Brent R, "Single-cell quantification of molecules and rates using open-source microscope-based cytometry," *Nat Methods*; 4 (2):175-181, 2007.

[5] A. Mansouri, "Region tracking via level set PDEs without motion computation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7): 947-961, 2002.

[6] L. Martens, G. Monsieur, C. Ampe, K. Gevaert, and J. Vandekerckhove, "Cell_motility: a cross-platform, open source application for the study of cell motion paths," *BMC Bioinformatics*, 7: 289, 2006.

[7] J. Munkres,"Algorithms for the assignment and transportation problems," *Journal of the Society of Industrial and Applied Mathematics*, 5(1):32-38, March 1957.

[8] Y. Shi and W. C. Karl, "Real-time tracking using level sets," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2: 34-41, 2005