

Automated Dynamic Cellular Analysis in High Throughput Drug Screens

Xiaowei Chen^{1,2} and Stephen T.C. Wong^{1,2,*}

¹HCNR Center for Bioinformatics, Harvard Medical School, Boston, MA 02115

²Department of Radiology, Brigham & Women's Hospital, Boston, MA 02115

*Corresponding author: stephen_wong@hms.harvard.edu

Abstract—To better understand drug effects on cancer cells, it is important to analyze cell nuclei dynamics from time-lapse fluorescence microscopy data. Existing methods, however, are rather limited in dealing with such time-lapse datasets while manual analysis is unreasonably time-consuming. We have therefore developed an automated system that can segment and track thousands of nuclei concurrently in time-lapse fluorescence microscopy data. Numerical nuclei features can be extracted based on our segmentation and tracking results. These features can be used for quantitative analysis of nuclei for high throughput drug screens.

1. INTRODUCTION

To better understand antimitotic drug effects, it would be ideal to be able to analyze quantitatively behavior of a large number of cell nuclei as a function of time. Automated time-lapse fluorescence microscopy imaging provides an important method to observe and study nuclei in a dynamic fashion [5, 6]. However, the increasing quantity and complexity of image data from dynamic microscopy renders manual analysis unreasonably time-consuming. Therefore, automatic techniques to analyze quantitatively nuclei behavior are of considerable interest in realizing the full potential of time-lapse microscopy in large scale quantitative biological research or high throughput drug screens.

In this paper, an automated system is presented for dynamic cellular analysis, which includes two modules nuclei segmentation and nuclei tracking. We used a global threshold technical to segment nuclei from their background and then employ watershed technique [1, 8] to separate touching nuclei. To deal with the over-segmentation problem, a method based on nucleus shape and size is proposed to merge the over-segmented nucleus fragments. Nuclei track in time-lapse microscopy is not a simple task for the dramatic shape and position changes of nuclei during cell mitosis. Figure 1 shows nucleus migration and shape changes during cell mitosis. Active contour trackers [3, 7, 9] will fail to track because daughter cell nuclei may not overlap at all with parent nuclei. Correlation tracker [4, 11] will also fail to track because the dramatic shape changes.

In this paper, a location and size based tracking technique is proposed to deal with these challenges.

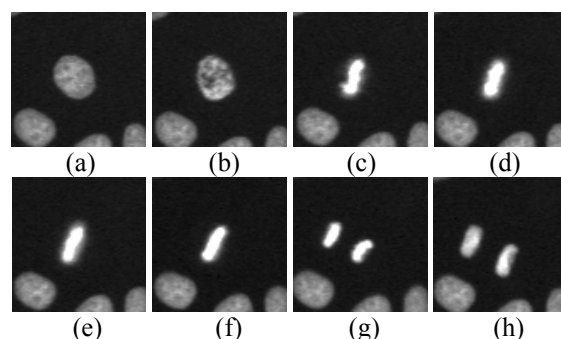


Figure 1: Nucleus migration and shape changes during cell mitosis (a)-(h) Consecutive image sub frames from a time-lapse sequence showing nucleus migration and shape change during cell mitosis.

2. DYNAMIC CELLULAR ANALYSIS

2.1 Nuclei Segmentation and Fragments Merging

We used the ISODATA algorithm to perform image thresholding [10]. This algorithm correctly segments most isolated nuclei, but it is unable to segment touching nuclei. To handle the case of touching nuclei, a watershed algorithm was used [1]. To handle the case of over-segmentation caused by watershed algorithm, a merging process is needed. We calculate the compactness of each segmented object. Compactness is defined as the ratio of the square of the perimeter to the area of the nucleus:

$$Compactness\ s = \frac{Perimeter^2}{4\pi \times Area} \quad (1)$$

Compactness is 1 when the shape of a nucleus is circle. If a round nucleus is divided into several fragments, the compactness of each fragment will be larger than the compactness of the entire nucleus. Based on this observation, a merging technique can be used. This technique identifies over-segmented nucleus fragments,

based on their size and shape, and then merges them into single nucleus units.

The procedure itself can be described as follows. Let N be the total number of segmented objects found by the Watershed segmentation algorithm. Let T_{size} be the minimum size of a nucleus in the image. All touching objects are evaluated using a checking procedure. Two objects are considered touching if they belong to the same object in the binary image before the Watershed algorithm was applied. The merging process finds the smallest touching objects in each iteration cycle and then uses this checking process to update the segmentation until no more touching objects can be merged. The checking process is implemented as follows:

- If the size of a touching object is less than T_{size} it is merged with its smallest touching neighbor.
- If the size of a touching object is greater than T_{size} , three compactness values are calculated; namely, for the object, the object's touching neighbor, and the two objects as a whole. If the compactness calculated decreases after the merging, these two objects are merged.



Figure 2: Nucleus fragments merging examples (a) Small fragments merging. (b) Big fragments merging.

Figure 2 shows two merging examples. In Figure 2 (a) the two small over-segmented fragments are merged based on their size. In Figure 2 (b) two large fragments are merged based on their compactness.

2.2 Nuclei Tracking

After nucleus segmentation, a matching process is used to set up the correspondence between nuclei at time t and nuclei at time $t+1$ by computing the distances between them. In this section, an association matrix is used to measure the distance, which is defined as follows:

$$distance = \begin{cases} 1 - \frac{C \cap \Omega}{C \cup \Omega} & C \cap \Omega \neq 0 \\ 1 + \frac{D(C, \Omega)}{\max(C_{size}, \Omega_{size})} & C \cap \Omega = 0 \end{cases} \quad (2)$$

where C stands for a nucleus at t , Ω stands for one of the nuclei appearing in its nearby location at $t+1$, $D(C, \Omega)$ is the Euclidean distance from the center of gravity of C to the center of gravity of Ω . The association matrix finds possible matches for nuclei at time t . A match is found if the distance is below a certain threshold. When nuclei divide, daughter cell nuclei may not overlap with their parents. In this case,

the value of the associate matrix will be large. In the matching process, four cases can occur:

1. Only one nucleus at time $t+1$ matches a nucleus at time t .
2. No nucleus at $t+1$ matches a nucleus at t .
3. A nucleus at t matches more than one nucleus at $t+1$.
4. More than one nucleus at t match a nucleus at $t+1$.

For Case 1, a successful match is found. Case 2 occurs when the nucleus either moves out of the field of view or becomes too dim to be detected. Only the border nuclei can move out of view. Thus, this situation can be detected by checking the nucleus position. In rare occasions, a nucleus becomes too dim to be detected. This will generate a tracking error. Case 3 indicates a nucleus split. The fourth case indicates unsuccessful segmentation where two or more nuclei touch together or overlap each other and cannot be separated. Case 3 and Case 4 cause ambiguous correspondences between the nuclei at t and $t+1$. We will focus on solve these ambiguous correspondences in the next subsection.

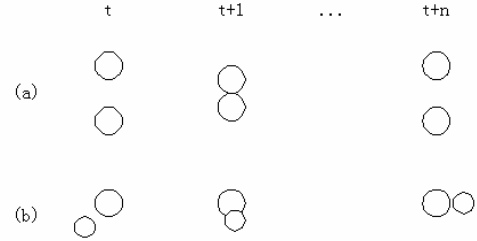


Figure 3: Examples of ambiguous correspondence caused by under-segmentation, (a) Nucleus touching, and (b) Nucleus overlapping

2.3 Solving Ambiguous Correspondence

We use the following strategy to solve ambiguous correspondence. When nuclei cannot be separated from each other and ambiguous correspondence happens, some information of these nuclei is recorded. Later, when they move away from each other, the tracking algorithm is able to solve the ambiguous correspondences by comparing their currently status with previously stored information. In our method, the sizes of nuclei and their relative locations were recorded. Figure 3(a) illustrates the way that the tracking method solves ambiguous correspondence caused by nucleus touching. When two touching nuclei move away from each other, their relative location does not change comparing with the time they move together. Thus, the relative location information can be used to solve such an ambiguous correspondence. Figure 3(b) illustrates the way that the tracking method solves ambiguous correspondence caused by nucleus overlapping. When a small nucleus moves away from its underneath big nucleus, correct

correspondences can be set up by associating a nucleus to the one of the same size.

One kind of split is caused by over-segmentation in which a single nucleus is divided to multiple pieces. This kind of ambiguity can be identified by comparing the change of nucleus size and the relative location of each nucleus at time $t+1$. If the nuclei at $t+1$ are separated by one pixel with the sum of their sizes is approximately the same as the nucleus at time t , the fragments are merged and viewed as a single unit.

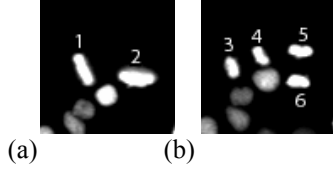


Figure 4: Example of ambiguous correspondence caused by nucleus division: (a) Nuclei before division. (b) Nuclei after division.

Nucleus division can be considered as a special case of nucleus splitting where one nucleus divides into two or more daughter cell nuclei. If all daughter cell nuclei only correspond with one nucleus at t , these daughter cell nuclei are decided as the daughters of that nucleus. If multiple nuclei divide simultaneously together, ambiguous correspondences can happen. Figure 4 shows an example of two nuclei dividing. In this example, the matching processing finds that daughter nucleus 4 corresponds with both nucleus 1 and 2, daughter cell nucleus 5 also corresponds with both nucleus 1 and 2. To solve this ambiguous correspondence the center of gravity for every two daughter cell nuclei are calculated. For nucleus 4, we first calculate the center of gravity of it with nucleus 3, 5, and 6 separately. Then we calculated the distance from these three centers of gravity to each center of gravity for nucleus 1 and nucleus 2. Finally, we found the center of gravity of nucleus 4 and 3 to the center of gravity of nucleus of 1 is the smallest of them all, and thus nucleus 4 is determined to be the daughter cell nucleus of nucleus 1. Similarly, nucleus 5 is determined to the daughter cell nucleus of nucleus 2.

When touching nuclei move together with another touching nuclei, a second ambiguity happens. In this case, the size and relative location of each touching entities are recorded. This information will be used together with previous recorded information to solve the ambiguous correspondence when the nuclei cluster divides.

3. SYSTEM IMPLEMENTATION AND EXPERIMENTAL RESULTS

Four nucleus sequences were used to test the efficiency of the proposed system. Each sequence consists of ninety-six frames over a period of 24 hours. The number of nuclei in

each sequence ranges from 78 to 204. The sequences were recorded at a spatial resolution of 672×512 , and a temporal resolution of one image per 15 minutes with a time-lapse fluorescence microscopy. The cancer cell line used is HeLa cells in our experiments. Two types of sequences were used to denote drug treated and untreated cells. The system was implemented with a window-based C/C++ application. For an image with approximately 300 nuclei, the average computation time was 1.4 seconds on a Pentium IV 2.4GHz computer.

3.1 Segmentation Results

To test the segmentation module, four images were chosen from each cell sequence. This generates a test set which contained 16 images and total 3,071 nuclei. Table 1 shows the segmentation results, which are compared with the result from manual analysis. The segmentation module correctly segmented 97.8% of the 3,071 nuclei.

TABLE 1: NUCLEI SEGMENTATION RESULTS

Correct Segment	Oversegment	Undersegment
3002 (97.8%)	29 (0.9%)	40 (1.3%)

3.2 Tracking Results

We use all of the four sequences to test the tracking module. Table 2 shows the tracking result. The tracking module achieved an average 94.3% tracking accurate rate.

TABLE 2: NUCLEI TRACKING RESULT

Sequence	Nuclei Number	Tracked	Missed
A	204	188	16
B	90	82	8
C	133	130	3
D	78	76	2
Total	505	476(94.3%)	29 (5.7%)

3.3 Application

The system presented is designed to allow large scale quantitative analysis of dynamic cellular behavior captured by time-lapse fluorescence microscopy data. We have applied this system in a biological research project [2, 12] to study cell cycle progress of HeLa cancer cells with and without perturbation of drug treatment, such as Taxol. Based on our segmentation and tracking results, we derived seven numerical nucleus features for identifying different cell cycle phases with good results [2]. Currently, we are applying this system for quantitative dynamic nucleus phenotyping of images acquired in high throughput anti-mitotic drug screens. The availability of such automated tools would improve the productivity and efficiency in decoding drug effects on cancer cells.

4. CONCLUSION

Time-lapse fluorescence microscopy is becoming an important method to study dynamic cellular processes over a large population of cells, with significant potential in achieving new, high throughput ways of conducting drug discovery and quantitative cellular studies. However, currently there is no existing tool available to segment and track large volumes of dynamic cellular image data automatically. Image analysis thus becomes the rate-limiting factor in time-lapse cellular imaging studies. The proposed cellular imaging software system can help to realize the potential of time-lapse fluorescence microscopy from an investigation tool in academic labs for small scale experimentation to high throughput drug screens using image-based assays in pharmaceutical and biotech companies.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the excellent collaboration with their biology collaborators in this research effort, and, in particular, Dr. Randy King and Dr. Susan Hyman, of the Department of Cell Biology and ICCB (Institute of Chemistry and Cell Biology), Harvard Medical School. They would like to thank other members of the Life Science Imaging Group of the Center for Bioinformatics, Harvard Center for Neurodegeneration and Repair (HCNR) and Brigham and Women's Hospital, HMS for their technical comments.

REFERENCE

1. Bleau A. and Leon J.L., "Watershed-based segmentation and region merging," *Computer Vision and Image Understanding*, 77, 3 (2000) 317-370.
2. Chen X., Zhou. X., and Wong S.T.C., "An automated method for cell phase identification in high throughput time-lapse screens," *Proceedings IEEE workshop on LSDM*, Cambridge, United Kingdom, (2004)
3. Christophe Z., Elisabeth L., Vannary M.Y., Nancy G., and Jean-Christophe O.M., "Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: A tool for cell-based drug testing," *IEEE Transactions on Medical Imaging*, 21, 10 (2002) 1212-1221.
4. Goobic A.P., Welser M.E., Acton S.T., and Ley K., "Biomedical application of target tracking in clutter," *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, 1 (2001) 88-92.
5. Hiraoka Y and Haraguchi T, "Fluorescence imaging of mammalian living cells," *Chromosome Res* 4, (1996) 173-176.
6. Kanda T., Sullivan K. F. and Wahl G. M., "Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells," *Current Biology* 8, (1998) 377-85.
7. Kass M., Witkin A., and Terzopoulos D., "Snakes: active contours models," *Internal journal of computer vision*, 1 (1988) 321-331.
8. Malpica N., Solorzano C.O., Vaquero J.J., Santos A., Vallcorba I. Garcia-Sagredo J.M., and Pozo F., Applying watershed algorithms to the segmentation of clustered nuclei, *Cytometry* 28 (1997) 289-297.
9. Nilanjan R., Acton S.T., and Klaus L., "Tracking leukocytes *in vivo* with shape and size constrained active contours," *IEEE Transactions on Medical Imaging*, 21, 10, (2002) 1222-1235.
10. Otsu N. "A threshold selection method from gray level histogram," *IEEE Transactions on System man Cybernetics* 8 (1978) 62-66.
11. Wang Z., Hu B.G., Liang L.C. and Ji Q., "Cell detection and tracking for micromanipulation vision system of cell-operation robot," *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 3, (2000) 1592-1597.
12. Zhou. X, Chen X., and Wong S.T.C., "Data Mining Issues in Time-Lapse Cell Cycle Data Analysis By Gaussian Mixture Model," *Proceedings IEEE workshop on LSDM*, Cambridge, United Kingdom, (2004)