Validation Methods for Cell Cycle Analysis Algorithms in Confocal Fluorescence Images

Dirk Padfield[1,3], Jens Rittscher[1], Nick Thomas[2], and Badrinath Roysam[3]
[1]GE Global Research, One Research Circle, Niskayuna, NY, 12309.
[2]GE Healthcare, Maynard Centre, Cardiff CF147YT, UK.
[3]Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180.

Abstract— **Automated analysis of live cells over extended time periods requires both novel assays and automated image analysis algorithms. Among other applications, this is necessary for studying the effect of inhibitor compounds that are designed to block the replication of cancerous cells. Due to their toxicity, fluorescent dyes that bind to the nuclear DNA cannot be used to mark nuclei, and traditional non-toxic nuclear markers do not yield information about the cell cycle phases. Instead, a non-toxic cell cycle phase marker can be used.**

We previously described a set of image analysis methods designed to automatically segment nuclei in such 2D time-lapse images. While the methods show promise, it is necessary to provide a validation framework for these methods. This paper introduces methods for validating the various stages of the algorithm in order to demonstrate their viability for automatic cell cycle analysis.

I. INTRODUCTION

When eukaryotic cells replicate, they pass through a tightly regulated series of events known as the cell cycle. The cell cycle comprises four phases, G1, S, G2, and M, each of which is marked by distinctive characteristics as DNA is replicated and the cell splits into two daughter cells. Blocking a cell from passing through any of these phases will stop the cell from replicating. This is especially important in the study and treatment of cancer since it is desirable to find ways to block the replication of cancer cells without killing normal cells. To study the effectiveness of such compounds and the mechanisms by which they operate, it is necessary to analyze the behavior of live cells over time, often several days. High-throughput methods are also needed in order to study the cell response to varying concentrations of different types of inhibitors.

Typical nuclear staining that relies on binding to DNA presents a problem for long-term cellular imaging because if cells are stained with a nuclear dye and cultured for extended periods, they die. Another approach is the staining of proteins in the nucleus, but this can potentially manipulate the genetics and does not provide cell cycle information. Instead, a dynamic cell cycle phase marker (CCPM) that does not manipulate the cell genetics can be used to determine cell cycle progression [1]. The analysis of this type of data can contribute significantly to the study of cell cycle inhibitors.

The automatic analysis of this type of data requires a sophisticated image processing approach for several reasons. The nuclei undergo rapid intensity and shape changes through the cell cycle due to the change in fluorescence distribution. The cells split into daughter cells throughout the sequence, and they move rapidly and in and out of the image plane. Depending on the imaging characteristics, there are also often a large number of cells that need to be segmented for each well.

We have developed methods for automatic analysis of these types of datasets. Although these methods show promise for the task of automatic cell cycle analysis, it is necessary to develop a validation framework for the results. Such validation can give both quantitative measures of the results and also enable comparison of various possible algorithms for different parts of the processing.

In Section II, we summarize the automatic cell cycle analysis algorithms. In Section III, we outline our proposed framework for validating the algorithms. Section IV discusses preliminary results and future work.

II. AUTOMATIC ANALYSIS METHODS

We approach the tracking problem by segmenting the 3D spatio-temporal volume of 2D images taken over time. This processing is broken into tractable steps by first using a level set segmentation approach with automatic seed placement to segment the G2 nuclei. This segmentation step is followed by a linking step to connect the nuclei in the various phases.

The segmentation of the nuclei in the G2 phase, which show up as dark structures surrounded by bright cytoplasm, is done using a 3D implementation of level sets [2], [3] constrained with high curvature and strong attraction towards edges. Seeding the level set is achieved through 2D normalized cross-correlation of the image with a ring-like mask. This correlation is done at multiple resolutions, and the results are combined to enable detection of cells of various sizes. However, this method also seeds points in the background that are close to cells, so a point classification method using edge curvature measures is employed to refine the seed detection. This classification method is based on the observation that edges around nuclei have many high-magnitude positive-curvature edges, and those around background points have fewer and weaker positive-curvature edges.

After segmenting the nuclei in the G2 phase, the next step combines and bridges the phases using linking methods. We use Euclidean distance measures to associate two daughters

with their corresponding parent. For more details on these methods, see [4].

These algorithms result in fully connected 3D nuclear "trees" that describe the evolution of the cells through their cell cycles. These tree structures can potentially contain a wealth of biologically relevant information such as whether particular cell cycle phases were extended or shortened, whether this effect was the same or different across the cell population, and whether mitosis occurred correctly and on time or not. However, for these results to prove useful to biologists, they need to be validated for correctness.

III. Validation Methods

The proposed validation framework is composed of two elements: validating the algorithms by comparison to known ground truth, and comparison of results on large-scale toxicological studies to biological expectation. The former involves validation of the seed placement algorithms and the final track and phase information and gives detailed information about individual cells. The latter involves the large-scale comparison of the results of the algorithms on untreated wells to wells treated with cell cycle inhibitors.

A. Algorithmic Validation

The seed placement step occupies an important role in the processing because seeds should be placed in the nuclei and not in the background to enable the correct segmentation of the nuclei. To validate these algorithms, a dataset is used that includes multiple stains. This dataset includes the CCPM along with a nuclear stain. This nuclear stain cannot be used for the time-lapse experiments because it is toxic and kills the cells over time. But it can be used to yield a ground truth map for the location of nuclei at one time frame.

Our approach is to segment the nuclei in the nuclear channel and use this as a mask for determining whether the seeds were placed correctly on the CCPM channel. This segmentation consists of thresholding the image followed by a constrained watershed segmentation to separate the touching nuclei. The seed placement algorithms are then run on the CCPM channel as usual, and each of the seeds that were placed is compared to the nuclear mask to determine whether the seed was placed correctly.

To validate the tracks and phases, a system was developed for facilitating manual generation of ground truth data. This system loads a series of 2D images taken over time and allows the user to track a particular cell through mitosis by marking the centroid of the cell on each image by its phase (G1, S, G2, or M). Each cell cycle is sequentially numbered beginning at the G1 stage after mitosis, and the parent and child relationships are stored for each tracked cell cycle. This ground truth generation process is supervised by an expert biologist to ensure correctness.

This ground truth dataset can be used to determine whether the daughters of each parent were correctly identified and whether the phases were correctly labeled in the automatically processed dataset. Thus, the system can automatically validate the results given the ground truth data.

B. Large-Scale Toxicological Validation

To demonstrate the biological relevance of the results, several datasets were generated with different types and concentrations of cell cycle inhibitors. 97 time-lapse images of a plate with 22 wells were acquired on an IN Cell Analyzer 3000 laser line scanning confocal imager. These wells consist of 6 control wells, and 16 wells treated with varying concentrations of cell cycle inhibitors: 8 with roscovitine and 8 with nocodazole, which arrest cells in the G1 and G2 phases, respectively. The biological expectation for the treated wells is that, for larger concentrations, the cell cycle will be arrested earlier and the cells will either die or remain in the final phase until the end of the experiment. Thus, the final phase length of G1 (roscovitine) or G2 (nocodazole) will be longer, whereas the number of mitosis events will be smaller. The automatically generated measures of cell cycle characteristics can thus be compared to the biologically expected results.

IV. Results and Future Work

To validate the seed placement and seed classification steps, images from 96 wells with nuclear stains were used. For the initial seeds and the classified seeds, the accuracy score was calculated as the ratio of seeds placed in nuclei to all seeds, and the classification step demonstrated an increase in accuracy from 20% to 60%. This validation approach can thus be used to improve the seed placement since new algorithms can be easily tested and compared against previous methods.

The ground truth system was used to generate tracks and phases for over 20 cell cycles corresponding to the 3 cells that split as shown in the results of [4]. The accuracy of this ground truth data was confirmed by an expert biologist. The automatic linking results for those three cells were automatically compared with the manual results, and all three cells demonstrated correct phase linking. The next step is to generate ground truth for more cells to validate more of the automatic results. Also, ground truth will be generated for the toxicological studies, and this will be compared to the automatically generated results to demonstrate the biological relevance of the algorithms.

The performance of the algorithms depends on the appropriate adjustment of certain parameters. Parallel to the validation itself a parameter optimization similar to that in [5] can be used to achieve optimal system performance.

References

[1] Nick Thomas, "Lighting the circle of life: fluorescent sensors for covert surveillance of the cell cycle," *Cell Cycle*, vol. 2, no. 6, pp. 545–9, Nov-Dec 2003.

[2] James A. Sethian, *Level Set Methods and Fast Marching Methods.*, Cambridge University Press, 1996.

[3] Stanley J. Osher and Ronald P. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer, 2002.

[4] Dirk Padfield, Jens Rittscher, Thomas Sebastian, Nick Thomas, and Badrinath Roysam, "Spatio-temporal cell cycle analysis using 3D level set segmentation of unstained nuclei in line scan confocal fluorescence images," in *IEEE ISBI (in press)*, 2006.

[5] N. Krahnstoever, T. Kelliher, and J. Rittscher, "Obtaining pareto optimal performance of visual surveillance algorithms," in *PETS*, 2005.