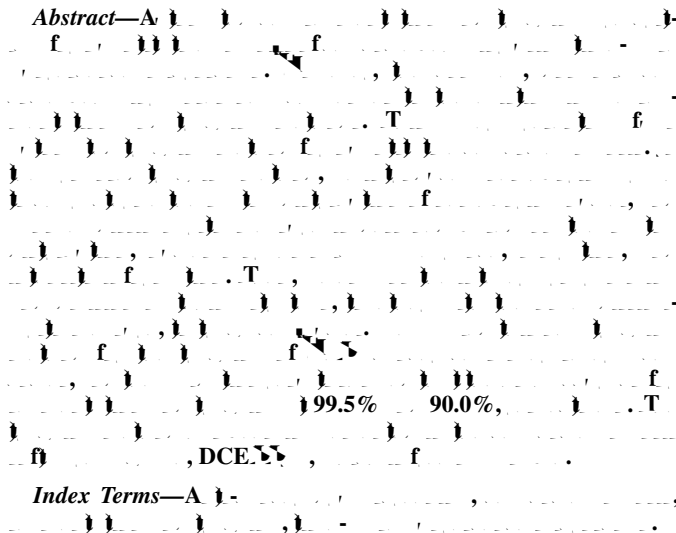


Multiple Nuclei Tracking Using Integer Programming for Quantitative Cancer Cell Cycle Analysis

Fuhai Li, Xiaobo Zhou, Jinwen Ma, and Stephen T. C. Wong*



I. INTRODUCTION

TAXANES are a group of drugs used in cancer treatment, which includes paclitaxel and docetaxel. This group of small molecule drugs prevents the growth of cancer cells by affecting microtubules. In normal cell growth, microtubules are formed when a cell starts dividing, and are broken down when the cell ceases dividing. Taxanes could stop the division of cancer cells by preventing the microtubules from breaking down [1], [2]. However, the subsequent molecular events that lead to apoptosis remain unknown. It is thus essential to dissect cellular processes

Manuscript received July 31, 2008; revised June 17, 2009; accepted July 02, 2009. First published July 28, 2009; current version published January 04, 2010. This work was supported by the National Institutes of Health under Grant NIH R01 LM008696. The work of S. T. C. Wong was supported by the Center for Bioinformatics Program Grant of Harvard Center of Neurodegeneration and Repair (now Harvard Neurodiscovery Center), Harvard Medical School. \blacktriangle

F. Li was with the Department of Information Science, School of Mathematical Sciences, and LMAM, Peking University, Beijing 100871, China. He is now with the Bioinformatics and Biomedical Engineering Programmatic Core, and Research Division, Department of Radiology, The Methodist Hospital Research Institute, Weill Medical College, Cornell University, Houston, TX 77030 USA (e-mail: robert.fh.li@gmail.com).

X. Zhou is with the Bioinformatics and Biomedical Engineering Programmatic Core, and Research Division, Department of Radiology, The Methodist Hospital Research Institute, Weill Medical College, Cornell University, Houston, TX 77030 USA (e-mail: xzhou@tmhs.org).

J. Ma is with the Department of Information Science, School of Mathematical Sciences, and LMAM, Peking University, Beijing 100871, China (e-mail: jwma@math.pku.edu.cn).

*S. T. C. Wong is with the Bioinformatics and Biomedical Engineering Programmatic Core, and Research Division, Department of Radiology, The Methodist Hospital Research Institute, Weill Medical College, Cornell University, Houston, TX 77030 USA (e-mail: xzhou@tmhs.org; stwong@tmhs.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2009.2027813

using small molecule drugs to study the effectors and proteins downstream in the mitotic pathways effectively [1], [2].

However, identifying effective small molecules or anti-mitotic drugs such as Taxanes from a large number of compounds available today is a daunting task [3], [4]. High content screening (HCS) using automated fluorescence microscopy and multiplate bioassays can be used to observe and quantitatively analyze the cell cycle behavior of individual cells treated by small molecules, and has become an important tool to help researchers and scientists better understand the complex cellular processes in disease pathogenesis, drug target validation, and drug lead identification [3], [5]. High content time-lapse cellular images capture abundant spatial and temporal morphological information of a population of cells, which enables the investigation of cell cycle behavior with strong statistical power. However, a large number of cellular images make the conventional manual analysis impractical. Automated quantitative analysis of time-lapse cellular images is critical to the success of such dynamic cell cycle studies [3].

Nuclei segmentation and tracking are the essential parts to quantify the cell cycle behavior. Although a number of segmentation and tracking methods have been reported in the literature, there remain many open problems, mostly due to complex cell cycle behavior including cell migration, morphological changes, cell division, and death. Broadly speaking, current cell tracking approaches can be classified into two categories: 1) the detection and segmentation based tracking and 2) the evolving model based tracking [6], [7].

In the first category, given an image sequence, $S = (F_1, F_2, \dots, F_N)$, the tracking problem is divided into $N - 1$ matching problems: $T = (M_{1,2}, M_{2,3}, \dots, M_{N-1,N})$, where T is a tracking solution of S , and $M_{i,i+1}$ denotes the matching solution of (F_i, F_{i+1}) [6], [8]. The tracking accuracy closely depends on the accurate detection and segmentation, dissimilarity measurement, and sophisticated matching strategies. In the second category, the boundaries or positions of cell nuclei are initialized in the first frame, and then their boundaries and positions evolve frame by frame [7]. Mean-shift [9], [10], parametric active contour [11]–[13], and level set [14], [15] are the widely used tracking approaches from this category. However, both mean-shift and parametric active contour cannot cope with cell division, and nuclei clusters may cause matching errors and inaccurate boundaries when nuclei move fast. While the level set method enables the topological changing for cell division, it also permits the fusion of overlapping cells. Extending these methods in the second category to cope with these challenges, e.g., the cell clustering, division, and fast migration, is nontrivial and will rapidly increase computation time [11], [16].

In this paper, we present a fully automated tracking algorithm using methods of the first tracking category for quantitatively

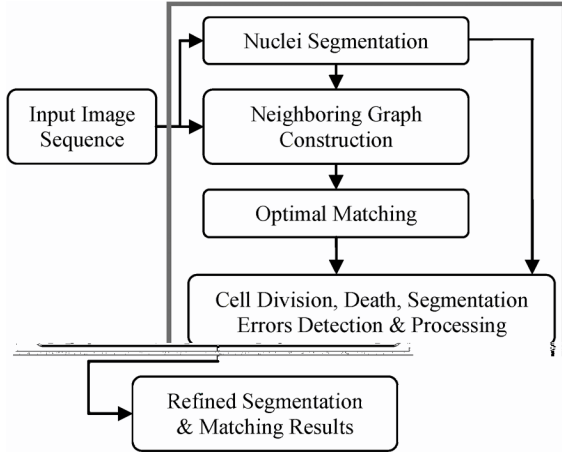


Fig. 1. Overview flowchart of the proposed system.

analyzing the cell cycle progress of a population of cancer cells treated by small molecule drugs such as Taxanes. There are two reasons why we choose methods in the first category. First, we developed a new nuclei segmentation method that can achieve higher accuracy. Secondly, the change in morphological appearance, long distance migration, and overlapping nuclei make the existing methods in the second category prone to errors.

In the proposed method, we defined an accurate dissimilarity measure and designing a sophisticated matching strategy. Fig. 1 provides a flowchart of the proposed system. The system consists of four major components: 1) nuclei segmentation, 2) neighboring graph construction, 3) nuclei division and integer programming based optimal matching, and 4) cell division, death, and segmentation error after postprocessing. In the segmentation component, nuclei are detected and segmented accurately. Then, a neighboring graph is generated to connect all segmented nuclei. In the matching component nuclei morphological appearance, migration and neighboring relationships are integrated into an accurate dissimilarity measure. In this sense, we use the cell phase information and then nuclei are associated by an optimal matching strategy. Finally, cell division, death, and segmentation errors are identified and corrected.

II. MATERIALS AND METHODOLOGY

A.

HeLa H2B-GFP cells were thawed six days before plating for each experiment and cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS). All cells were plated in eight well #1 German borosilicate sterile bottomed plates (Nalge Nunc International) and incubated at 37 °C in 5% CO₂ for 18 h before imaging at 25 000 cells per well (50 000 cells per mL). Untreated cells were received medium while treated cells were received 300 nm nocodazole. Images were acquired on an automated epifluorescence TE2000-E Eclipse microscope (Nikon Instruments, Inc.) with a motorized XYZ-plane stage. Light was generated from a mercury arc lamp with two neutral density filters. Auto-focusing was performed on the first pass and for every subsequent 10 passes to compensate for motor drift. Images were acquired using a 0.2 s exposure time, every 15 min for 50 h, yielding a total of 200 images for each position.

In this study, the nuclei detection and segmentation method proposed in [17] is employed. Essentially, this method consists of three components: binarization, nuclei center detection, and nuclei boundary delineating. We implemented the binarization process using an adaptive thresholding method [18]–[22]. Nuclei center detection is important because it determines the segmentation results by providing seed points for the seeded-watershed algorithm [17], [23]. To detect the centers of nuclei, we first utilize both intensity and shape information by combining the original intensity with the distance image as: $\mathbf{I}_1 = \mathbf{I}_0 + \alpha \mathbf{I}_{Dis}$, where \mathbf{I}_1 is the new image, \mathbf{I}_0 is the original image, and \mathbf{I}_{Dis} is the distance image. The parameter α affects the nuclei center detection results by changing the relative weights of the intensity information (intensity image) and shape information (distance image). A high α value increases the influence of the distance image and a low α value decreases its influence. We empirically set $\alpha = 0.4$ after testing a number of values on a training data set. The validation of how sensitive the parameter α affects the performance of the detection algorithm is provided in Section III-A. We detect the nuclei center using the gradient vector field (GVF) followed by Gaussian filtering [17], [24]. The pseudo-code for nuclei detection with GVF is as follows.

A **(, , D ,)** **F**

M_{nuc} ; /*Nuclei pixel matrix. The background and nuclei pixels are set as 0 and 1 respectively*/

M_{gvf} ; /*GVF matrix storing the gradient vectors of each pixel.*/

$M_{par} \leftarrow zeros(m, n)$; /*Particle convergence matrix, whose values are the number of virtual particles stopping at each pixel.*/

$(ExistNonzeroElement(M_{nuc}))$

$(x_0, y_0) \leftarrow GetOneNonzeroElement(M_{nuc})$; /*
Get coordinates of one nonzero element (nuclei pixel)
in M_{nuc} */

$(x_1, y_1) \leftarrow (0, 0)$;

$((x_1, y_1) \neq (x_0, y_0))$

$(x_1, y_1) \leftarrow$

$MoveVirtualPartical((x_0, y_0), M_{gvf}(x_0, y_0))$;
/*Move the virtual particle from (x_0, y_0) along
the gradient vector to the point (x_1, y_1) . If pixel
 (x_0, y_0) is a local maximum, the gradient vector
will point to itself, then the (x_0, y_0) and
 (x_1, y_1) are the same point.*/

$M_{par}(x_1, y_1) \leftarrow M_{par}(x_1, y_1) + 1$; /*Number of
virtual particles at (x_1, y_1) increases one.*/

$M_{nuc}(x_0, y_0) \leftarrow 0$; /*Set the visited point as 0.*/

$SeedsImage \leftarrow Thresholding(M_{par})$;

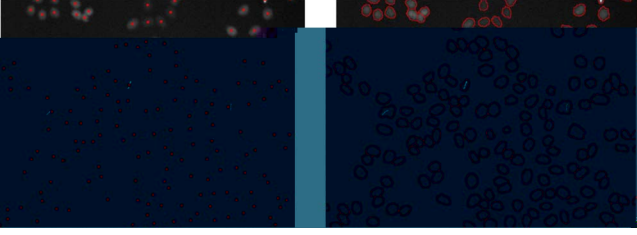


Fig. 2. Representative detection and segmentation results of the proposed segmentation method.

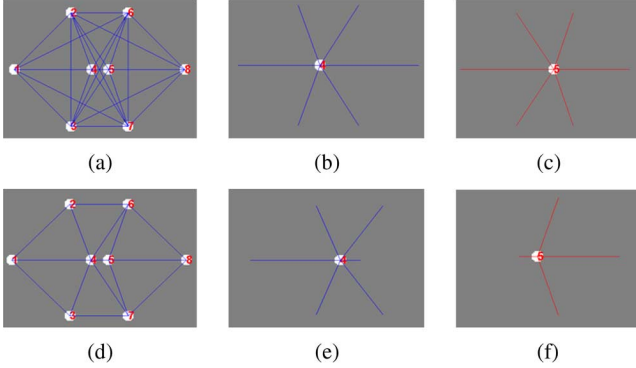


Fig. 3. Example neighboring graphs. (a) Neighboring graph built by connecting any two nuclei; (b) and (c) are the edges connected to the fourth and fifth nuclei in neighboring graph (a). (d) Neighboring graph constructed by using the Delaunay triangulation method; (e) and (f) are the edges connected to the 4th and 5th nuclei in neighboring graph (d).

After detecting the nuclei centers, we delineate the nuclei boundaries using the seeded watershed algorithm. Fig. 2 provides representative detection and segmentation results.

The similar morphology of nuclei makes it difficult to accurately distinguish adjacent nuclei using only their morphological features. Therefore, we make use of the information of neighboring nuclei to better distinguish the adjacent nuclei. A way to describe the neighboring nuclei information is to connect the segmented nuclei into a neighboring graph. In the neighboring graph, the edges that connect nuclei with their neighbors can be used to characterize the spatial neighboring distribution, as seen in Fig. 3. However, this poses a critical question: how can we generate a neighboring graph that has strong ability to distinguish the adjacent nuclei using the neighboring relationships (spatial distribution of neighboring nuclei)? One possible solution is to connect all the neighbors of one nucleus, as shown in Fig. 3(a). However, it is difficult for a human observer to understand the neighboring relationships due to the complexity of this neighboring graph. In addition, more edges in the neighboring graph increase the computation time. Furthermore, the adjacent nuclei may have similar neighboring relationships and thus, it adds to the problem of distinguishing them from each other, as seen in Fig. 3(b) and (c). A different solution was proposed by Delaunay in 1934 Delaunay triangulation [25]. The representative neighboring graphs generated by using the Delaunay triangulation are provided in Fig. 3(d) and Fig. 4. As compared to the above neighboring graph, Delaunay neighboring graphs have less edges and are therefore easier to be understood, as

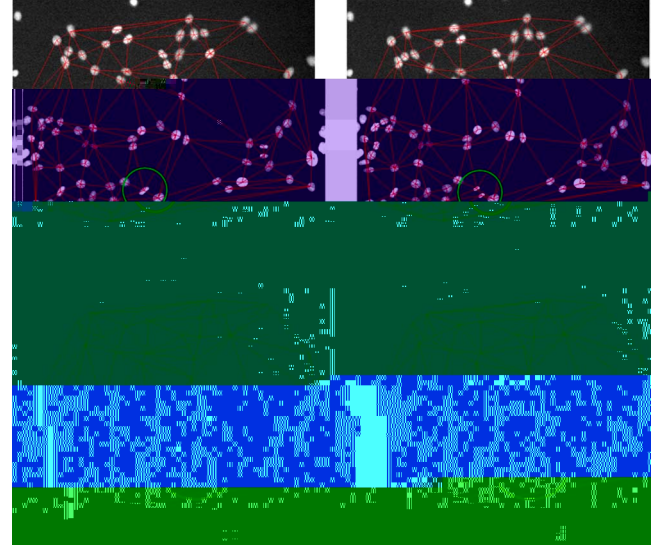


Fig. 4. Illustration of the local variation of the neighboring graph caused by the cell division. (a) Neighboring graph in frame t ; (b) neighboring graph in frame $t + 1$ in which there is a cell division in the green box; (c) neighboring graph after removing the dividing cell in frame $t + 1$; and (d) neighboring graph after removing the divided cells in frame $t + 1$.

seen in Fig. 3(d). Moreover, in such a graph the adjacent nuclei often have very different spatial distribution of neighboring nuclei that can be used to distinguish them easily, as seen in Figs. 3(e) and (f). These properties indicate that Delaunay triangulation technology is a good choice to generate the neighboring graph.

To match the segmented nuclei in two consecutive frames, we need to: 1) define a dissimilarity measure between two nuclei in two consecutive frames, and 2) choose an appropriate matching strategy.

The dissimilarity measure indicates how different two nuclei are and is critical for accurate nuclei matching. To define an appropriate dissimilarity measure, we first characterize a vertex (nucleus) in the k th neighboring graph (frame), \mathbf{v}_i^k , using an attribute vector: $\mathbf{v}_i^k = (\mathbf{c}_i^k, \mathbf{s}_i^k, \mathbf{h}_i^k, \theta_i^k)$, where $\mathbf{c}_i^k = (x_i^k, y_i^k)$ denotes the centroid; \mathbf{s}_i^k is a binary matrix describing the shape of a nucleus, in which the nucleus and background are represented by 1 and 0, as seen in Fig. 5; \mathbf{h}_i^k is the intensity distribution of the nucleus, and we use the intensity histogram to represent the intensity distribution, as seen in Fig. 6; and $\theta_i^k = \{(l_{i,p}^k, \theta_{i,p}^k) | p = 1, 2, \dots, M\}$ is a $2 \times M$ matrix that describes the spatial distribution (length and angle) of M edges connected to the nucleus in the neighboring graph, where θ is defined in radians, as shown in Fig. 7.

Based on the attribute vector, we define the dissimilarity measure between the i th nucleus in frame k and the j th nucleus in frame $k + 1$ as follows:

$$d(\mathbf{v}_i^k, \mathbf{v}_j^{k+1}) = \alpha_1 \times d_n(\mathbf{c}_i^k, \mathbf{c}_j^{k+1}) + \alpha_2 \times q_1 \times d_n(\mathbf{s}_i^k, \mathbf{s}_j^{k+1}) + \alpha_3 \times q_2 \times d_n(\mathbf{h}_i^k, \mathbf{h}_j^{k+1}) + \alpha_4 \times d_n(\theta_i^k, \theta_j^{k+1}) \quad (1)$$

where α_i , $i = 1, 2, 3, 4$ are the weighting parameters which satisfy $\alpha_i \geq 0$, $\sum_{i=1}^4 \alpha_i = 1$; q_i , $i = 1, 2$ are two phase control parameters, and we will describe them in detail after defining

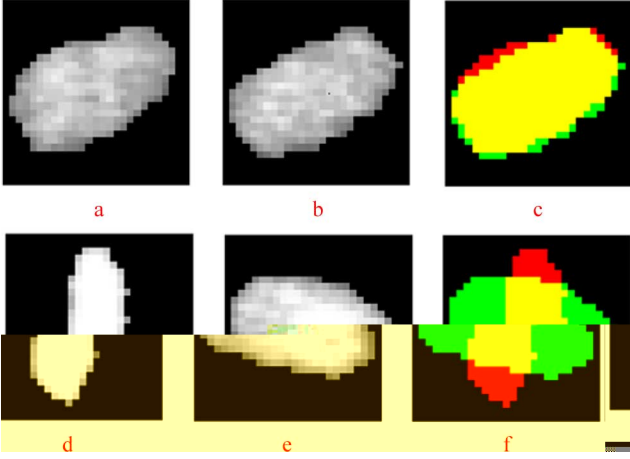


Fig. 5. Illustration of measuring shape variability. (a), (b) Cropped and padded cellular images (matrixes) of the same nucleus in two consecutive frames. The corresponding binary matrixes have the same number of rows and columns as the cellular images. (c) Shows the calculation of shape variability; the yellow color indicates the intersection region; the red color indicates the complementary region of nucleus in (a); the green color indicates the complementary region of the nucleus in (b). (d), (e), (f) show the case of two different nuclei in two consecutive frames.

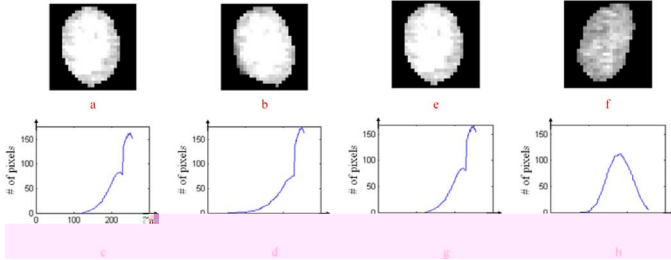


Fig. 6. Illustration of measuring variation of intensity distribution. (a), (b) Same nuclei in two consecutive frames. (c), (d) Are their intensity histograms. (e), (f), (g), (h) Show the case of two different nuclei in two consecutive frames.

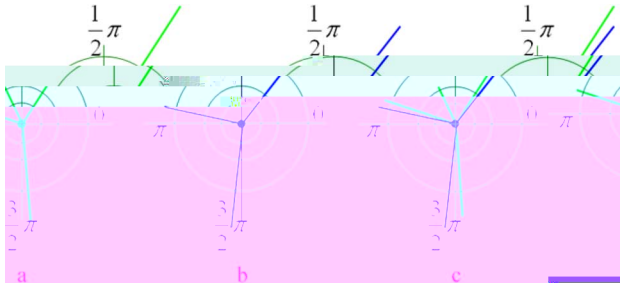


Fig. 7. Illustration of measuring the variability of spatial distribution of neighboring nuclei. The green edges in (a) indicate the spatial distribution of neighboring nuclei of one nucleus in one frame, and the blue edges in (b) indicate the distribution of neighboring nuclei of the other nucleus in the next frame. (c) shows the difference of these two spatial distributions.

the four normalized dissimilarity measures: $d_n(\mathbf{c}_i^k, \mathbf{c}_j^{k+1})$, $d_n(\mathbf{s}_i^k, \mathbf{s}_j^{k+1})$, $d_n(\mathbf{h}_i^k, \mathbf{h}_j^{k+1})$, and $d_n(\mathbf{h}_i^k, \mathbf{h}_j^{k+1})$.

The first term, $d_n(\mathbf{c}_i^k, \mathbf{c}_j^{k+1})$, measures the Euclidean distance between the centroids of two nuclei, as show in (2) at bottom of the page where D denotes the maximum migration distance per frame, and we set $D = 30$ pixels empirically.

The second term, $d_n(\mathbf{s}_i^k, \mathbf{s}_j^{k+1})$ measures the nuclei shape variation which is defined as

$$d_n(\mathbf{s}_i^k, \mathbf{s}_j^{k+1}) = 1 - \frac{|\mathbf{s}_i^k \cap \mathbf{s}_j^{k+1}|}{|\mathbf{s}_i^k \cup \mathbf{s}_j^{k+1}|}. \quad (3)$$

Fig. 5 illustrates the calculation of nuclei the shape variation. The procedure of measuring the nuclei shape variation is as follows. After the initial nuclei segmentation we obtain a label matrix. In such a matrix, pixels within a nucleus region are labeled with a unique integer number which is used to distinguish it from other nuclei regions. We can obtain the binary matrix of a nucleus by cropping a small square region that contains the nucleus region tightly. Before comparing two binary matrices, we pad them symmetrically with zeros (background pixels) such that they have the same number of rows and columns. Then we can calculate their intersect region and union region based on (3).

The third term, $d_n(\mathbf{h}_i^k, \mathbf{h}_j^{k+1})$, describes the variation of the intensity distribution

$$d_n(\mathbf{h}_i^k, \mathbf{h}_j^{k+1}) = \frac{\sum_{t=\min(I)}^{\max(I)} |\mathbf{h}_i^k(t) - \mathbf{h}_j^{k+1}(t)|}{\sum_{t=\min(I)}^{\max(I)} \max(\mathbf{h}_i^k(t), \mathbf{h}_j^{k+1}(t))} \quad (4)$$

where $\max(I)$ and $\min(I)$ denote the maximum and minimum intensity values of the images. The numerator, $|\mathbf{h}_i^k(t)|$

In (5.1), $\{e_j^{(k+1)}\}$ is a collection of m edges selected from $\{e_j^{k+1}\}$, and it enables the comparison between the spatial distributions with different edges. In (5.2), (m/n) measures the difference of edge number; $|l_{i,t}^k - l_{j,p_t}^{k+1}| / \max(l_{i,t}^k, l_{j,p_t}^{k+1})$ compares the difference of the t th edge length; $(\theta_{i,t}^k - \theta_{j,p_t}^{k+1})$ measures the variation of the t th edge angle; and the $\min_{e_j^{(k+1)}}$ operation in (5.2) is to

find the smallest difference between two spatial distributions of neighboring nuclei. Fig. 7 illustrates the variation of the spatial distributions of neighboring nuclei.

The determination of the optimal weighting parameters α_i , $i = 1, 2, 3, 4$, is nontrivial as the validation of tracking results is time-consuming. Thus, we empirically tested a few values for α_i , $i = 1, 2, 3, 4$, and selected the values with the highest tracking accuracy. In this study, we set them as: 0.31, 0.15, 0.23, and 0.31, with the ratio of them is roughly as 4 : 2 : 3 : 4. In Section III-B1, we provide a sensitivity analysis of α_i , $i = 1, 2, 3, 4$. The validation results show that the tracking approach is robust to the parameters.

However, not all the four terms in the dissimilarity measure can be used, when the nuclei change from one phase to another regarding the following facts: 1) when the nuclei enter the prophase from the inter-phase their intensity increases dramatically, 2) when the nuclei enter the metaphase following the prophase, their shape changes dramatically, and 3) when cell divide the new generated cell nuclei will change the local structure of the neighboring graph around the dividing cells, as seen in the green circles of Fig. 4(a) and (b). To address problems 1) and 2), we add two phase control parameters: q_1 and q_2 , to the shape and intensity terms. The two control parameters are defined as

$$q_1 = f_1(p(\mathbf{v}_i^k), p(\mathbf{v}_j^{k+1})) = \begin{cases} 0, & \{p(\mathbf{v}_i^k) = \text{prophase}, p(\mathbf{v}_j^{k+1}) = \text{metahase}\} \\ 1, & \text{otherwise} \end{cases} \quad (6.1)$$

$$q_2 = f_2(P(\mathbf{v}_i^k), p(\mathbf{v}_j^{k+1})) = \begin{cases} 0, & \{p(\mathbf{v}_i^k) = \text{interphase}, p(\mathbf{v}_j^{k+1}) = \text{prophase}\} \\ 1, & \text{otherwise} \end{cases} \quad (6.2)$$

where $p(\cdot)$ is a cell phase identification function. The details of the phase identification function can be found in [26]. The solution to problem (3) is described in the following section.

The matching strategy associates (match) the nuclei in two consecutive frames based on the defined dissimilarity measure. In this study, we propose a division matching and 0-1 integer programming based optimal matching strategy.

To avoid that nuclei of dividing cells affect the neighboring graph, we first match the nuclei of dividing cells with the division matching strategy first and ignore them while we build the neighboring graph, illustrated with green circles of Fig. 4(c) and (d). Two observable facts of cell division are: 1) parent nucleus (Metaphase) divides along the direction perpendicular to its major axis [see Fig. 8(a) and (b)] and 2) two daughter nuclei (Anaphase) have similar appearances; see

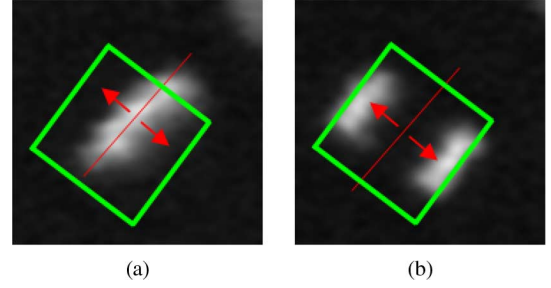


Fig. 8. Illustration of the nuclei division.

Fig. 8(b). Therefore, we need to find the parent nucleus of two daughter nuclei in a local region, as shown in the green rectangular boxes in Fig. 8(a) and (b). Specifically, this can be achieved by 1) finding all the nuclei which are in anaphase in frame $(t+1)$ and placing them in a list L_d and 2) retrieving the first element in L_d , and checking if there is a sibling nucleus that satisfies $d_n(\mathbf{s}_i^k, \mathbf{s}_j^{k+1}) + d_n(\mathbf{h}_i^k, \mathbf{h}_j^{k+1}) < d_T$, where d_T is a dissimilarity threshold. If there is no sibling nucleus, the program checks the next element in L_d ; otherwise, the program enters the next step; 3) checking if there exists a Metaphase nucleus in the desired local region in frame t . If the required nucleus exists, we match it with the two daughter nuclei. The steps 2 and 3 are repeated until all elements in L_d are checked.

After matching the dividing cells' nuclei, the remaining nuclei are matched by the 0-1 integer programming based optimal matching strategy. The optimal matching problem is formulated as follows: let $\mathbf{F}^k = \{\mathbf{v}_i^k | i = 1, 2, \dots, m\}$ and $\mathbf{F}^{k+1} = \{\mathbf{v}_j^{k+1} | j = 1, 2, \dots, n\}$ denote two consecutive frames. For each nucleus, \mathbf{v}_i^k , in frame k , there are several matching candidates: $\mathbf{M}_i = \{(\mathbf{v}_i^k, \mathbf{v}_j^{k+1}) | t = 1, 2, \dots, N_i\}$. Therefore, a total of $N = \sum_{i=1}^m N_i$ possible matches: $\mathbf{M} = \cup_{i=1}^m \mathbf{M}_i$ exist. The optimal matching strategy, under the constraint that each nucleus has at most one match is to find a solution $\mathbf{x}_0 = \{0, 1\}^N$ such that

$$\mathbf{x}_0 = \arg \max_{\mathbf{x} \in \{0, 1\}^N} (f(\mathbf{x})) \quad (7.1)$$

$$f(\mathbf{x}) = \sum_{t=1}^N (\mathbf{x}(t) \times d(M_t)) = \sum_{t=1}^N \left(\mathbf{x}(t) \times e^{-d(\mathbf{v}_{i_t}^k, \mathbf{v}_{j_t}^{k+1})} \right). \quad (7.2)$$

In (7.2), $\mathbf{x}(t)$ denotes the t th element of vector \mathbf{x} , and the term $e^{-d(\mathbf{v}_{i_t}^k, \mathbf{v}_{j_t}^{k+1})}$ is the dissimilarity of two nuclei. The matching constraints can be formulated as

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} \quad (7.3)$$

where \mathbf{A} is a $(m+n) \times N$ dimension "involving" matrix and \mathbf{b} is a $(m+n)$ elements vector of ones. In matrix \mathbf{A} , the t th column corresponds to the t th possible match $(\mathbf{v}_{i_t}^k, \mathbf{v}_{j_t}^{k+1})$ in \mathbf{M} , $t = 1, 2, \dots, N$, indicating which nuclei correspond to that particular match

$$A(q, t) = \begin{cases} 1, & q = i_t \text{ or } j_t; \\ 0, & \text{otherwise,} \end{cases}; \quad q = 1, 2, \dots, (m+n). \quad (7.4)$$

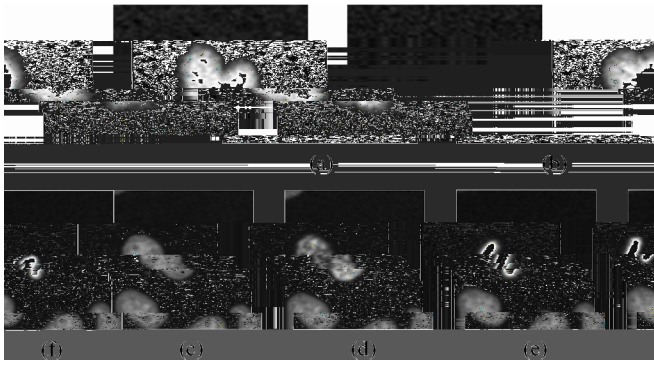


Fig. 9. Bi-nuclei and four cell cycle phases. (a) Bi-nuclei; (b) separated bi-nuclei; (c) inter-phase; (d) prophase; (e) metaphase; and (f) anaphase.

Obviously, this is a 0-1 integer programming problem. The linear programming based branch-and-bound (LPBB) algorithm is widely used to solve the 0-1 integer programming problem [27], [28]. The optimization process of LPBB is to build a searching tree by repeatedly discretizing (0 or 1) the variables (branching) and pruning the tree branches based on the optimal value of the node (bounding), computed by linear programming. In general, in a given search tree node the optimal value of the LP-relaxation problem at this node is greater than the existing best solution. The node and its branches will not be searched and a new node will be searched. If a new feasible 0-1 solution has a lower value, as compared with the existing best solution, the algorithm will update the existing best solution using the new feasible solution. If the value of the LP-relaxation problem at this node is less than that of the existing best solution, but it is not the 0-1 solution, two new searching tree nodes branching from this node are added by discretizing another variable (0 or 1), and the searching begins at a new node. The optimal solution is obtained by repeating the above process.

Since the above optimal matching strategy is under the constraint that one nucleus matches at most once, there exist some nuclei with no match at all due to cell division, cell death (apoptosis), and segmentation errors (over- and under-segmentation). Although we have employed division matching to match all dividing cells' nuclei, some nuclei are still missing due to the bias of the phase identification. We, however, can correct the segmentation errors and identify cell division and cell death by checking the remaining nuclei.

A bi-nucleus is two inter-phase nuclei lying closely together such that they cannot be distinguished by the detection resolution of the microscopy, and thus appear as one nucleus in a few contiguous frames but eventually separate in the later frames, as seen in Fig. 9(a) and (b). When a bi-nucleus separates, a nucleus without match will be generated. Therefore, all the nuclei in frame $t+1$ which have no match are candidates of newly divided nuclei (in anaphase), newly separated bi-nuclei, or over-segmented nuclei, as seen in Fig. 9. Due to the distinct morphological appearances of the anaphase, bi-nuclei, and over-segmented nuclei, it is easy to separate them using a support vector machine (SVM) classifier based on the following six manually selected features: size,

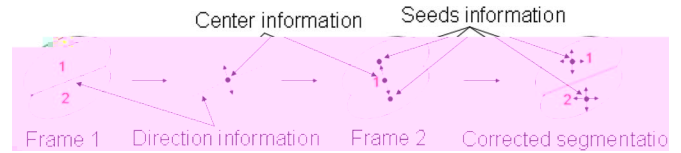


Fig. 10. Illustration of the proposed nuclei splitting method.

average intensity, standard deviation of intensity, compactness, long axis, and short axis. For the new divided nuclei and separated bi-nuclei candidates, we match them by finding the nuclei in frame t that are closest to them. For the over-segmented nuclei candidates, we merge them by finding the ones in frame t that overlap with them and have the smallest size.

All the nuclei with no matching in frame t are either the candidates for apoptotic cells or due to the under-segmentation of some nuclei in frame $t+1$. Therefore, we need to classify the candidates into the two groups: apoptosis and under-segmentation. Given a candidate v_{i1}^k , we find all its possible matching nuclei in frame $t+1$ and then label the candidate as an apoptosis nucleus if there is no matching nucleus, v_j^{k+1} which satisfies (under-segmentation)

$$\begin{cases} S(v_j^{k+1}) > \alpha S(v_{i1}^k) \\ S(v_j^{k+1}) > \alpha S(v_{i0}^k) \end{cases} \quad (8)$$

where $S(\cdot)$ is the operator of calculating the area of a nucleus, and we empirically set $\alpha = 1.8$. Otherwise, we label the nucleus v_{j1}^{k+1} , which satisfies (8) and has the largest size, as the under-segmented nucleus, and split it into two nuclei: $v_{j1,1}^{k+1}$ and $v_{j1,2}^{k+1}$. We propose a splitting method to separate under-segmented nuclei, as shown in Fig. 10. After splitting, we can easily match the two split nuclei $v_{j1,1}^{k+1}$ and $v_{j1,2}^{k+1}$ with v_{i0}^k and v_{i1}^k .

III. VALIDATION RESULTS

A.

Unlike cell cytoplasmic images with complex shapes and significant intensity variations, cell nuclei have regular shape and uniform intensity. If the cell nuclei centroids are detected correctly, the cell boundaries will be well delineated, as shown in Fig. 2. Thus, it is reasonable to validate the segmentation by just counting the over-segmentation errors (two or more detected nuclei centroids are presented inside one nucleus region) and under-segmentation errors (two or more nuclei regions share only one detected centroid).

To validate how the sensitivity of the parameter α affects the detection performance, we uniformly selected 21 nuclei images from each of four nuclei image sequences (frames 1, 10, 20, ..., 200 are selected). These 84 (21×4) nuclei images contain about 8300 nuclei. We used two error measures to evaluate the performance of the detection approach: over-detection (the same as the over-segmentation) and under-detection errors (the same as the under-segmentation). We varied the value of α from 0 to 1. The detailed results of the detection algorithm are provided in Table I. As we can see, the over-detection error is more sensitive to the variation of the value of α

TABLE I
SENSITIVITY VALIDATION OF PARAMETER α IN NUCLEI DETECTION

α	0	0.2	0.3	0.4	0.5	0.6	0.8	1
Over-detection	53	41	42	42	45	53	64	83
Under-detection	18	18	15	14	14	10	13	22
Errors Rate (%)	0.86	0.71	0.69	0.67	0.71	0.76	0.93	1.3

TABLE II
VALIDATION OF SEGMENTATION RESULTS

Seq.	# of nuclei	Correctly segmented	Over-segmented	Under-segmented
1	18683	18564 (99.36%)	35 (0.19%)	84 (0.45%)
2	19469	19394 (99.61%)	40 (0.21%)	35 (0.18%)
Total	38152	37963 (99.50%)	75 (0.20%)	119 (0.30%)

TABLE III
COMPARISON OF SEGMENTATION RESULTS USING NUCLEI IMAGE SEQUENCE 1

Methods	Correctly segmented	Over-segmented	Under-segmented
Watershed	16254 (87%)	2242 (12%)	187 (1%)
Watershed/ Hybrid	17690 (94.69%)	747 (4%)	246 (1.32%)
Detection/ Watershed	18564 (99.36%)	35 (0.19%)	84 (0.45%)

as compared to the under-detection error. If we increase α to 1, both over-detection and under-detection increase. From 0.2 to 0.6, the detection algorithm keeps the similar error rate. From 0 to 0.8, the error rate is confined within 1%. In conclusion, the detection approach is robust to the parameter α .

) Two cell sequences were selected to evaluate the proposed segmentation method with the manual detection results were used as the ground truth.

TABLE V
COMPARISON OF NUCLEI TRACKING RESULTS

Seq.	Proposed tracking method			Location and size based tracker		
	# of cell traces	ETR	EMR	# of cell traces	ETR	EMR
1	97	7 (7.22%)	6 (6.18%)	109	36 (33.03%)	47 (43.12%)
2	109	16 (14.68%)	15 (13.76%)	116	51 (43.97%)	56 (48.28%)
3	107	9 (8.41%)	9 (8.41%)	111	48 (43.24%)	51 (45.95%)
4	131	12 (9.16%)	11 (8.40%)	128	61 (47.66%)	56 (43.75%)
Total	444	44 (9.91%)	41 (9.23%)	464	196 (42.24%)	210 (45.26%)

TABLE VI
COMPARISON OF THE DIFFERENCES BETWEEN THE ERROR TRACES AND ERROR MATCHES OCCURRED UNDER THE VARIATIONS OF PARAMETERS $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ AND THAT OCCURRED UNDER THE ORIGINAL SETTING

Seq.	Common Error Traces	Different Error Traces	Common Error Matches	Different Error Matches
1	7	0	6	0
2	16	1	15	0
3	9	1	9	2
4	12	0	11	0
Total	44	2	41	2

In summary, the location and size based tracker first defines a dissimilarity measure based on the overlapping area and the distance of the centroids of two nuclei in two consecutive frames. Then it matches the nuclei one by one based on the defined dissimilarity measure. We provide the detailed comparison results in Table V. On average about 90% of nuclei traces obtained from the proposed method are correct, and both the ETR and EMR are restricted to 10%. Whereas both the ETR and EMR of the location and size based tracker are 30% higher than the proposed tracking method. Through further investigation, we found that fast moving and cell division caused the most nuclei matching errors for the location and size based tracker while the proposed system deals with these challenges well. In conclusion, the proposed tracking method is accurate and reliable.

E T

In Fig. 11, we illustrate the representative 2-D HeLa cell migration maps, in which different cell migration traces are represented by different colors. As we can see, some cells move in a small region (shown inside the red square box), and some cells migrate farther out (shown in the blue square box). In the anti-cancer drug screening study, the analysis of cell cycle time distribution is important. The drugs or commands that can prolong the time of the cell cycle could be new candidates for anti-cancer treatment. In Fig. 12, the division progression of a nucleus is shown. There are three cell divisions at $T = 3.75, 27.5,$ and 28.5 h, respectively, as shown in Fig. 13(b). We also can see that the cell cycle progression of daughter nuclei may

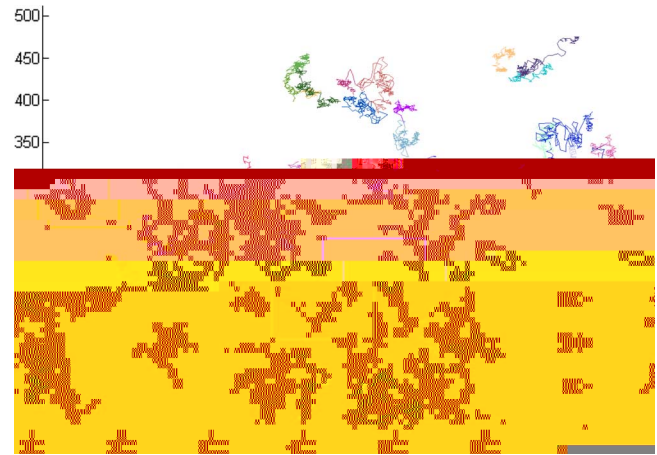


Fig. 11. Two-dimensional cell migration traces.

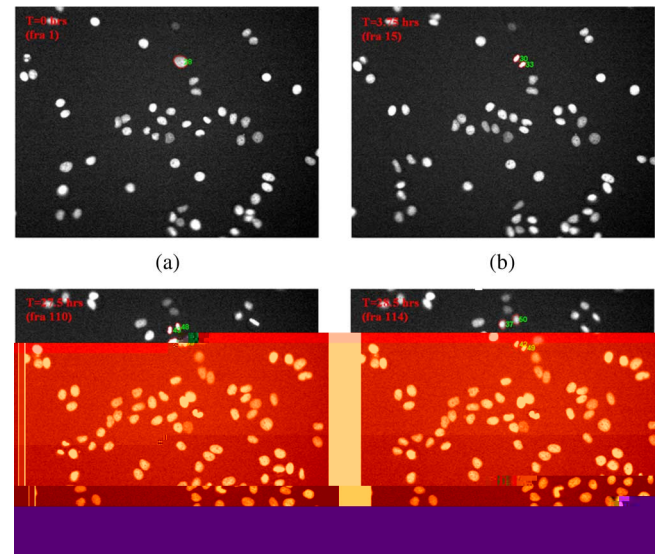


Fig. 12. Illustration of cell division. (a) The #28 nucleus in the first frame; (b) the #28 nucleus divided in frame 15, and the two daughter nuclei are #30 and #33, respectively; and (c) the daughter nuclei #30 and #33 divided in frames 110 and 114, respectively.

have different cell division rates, as seen in Fig. 13(a). In addition, three representative 3-D cell nuclei trajectories extracted automatically are shown in Fig. 14, in which we can clearly see both the cell migration and cell division information.

We incorporated the proposed tracking method into a software package: DCELLIQ (Dynamic Cellular Image Quantifier), which is designed specifically for the analysis of cell cycle progression and dynamics. In DCELLIQ system, cell nuclei segmentation, tracking, and cell cycle phase identification are performed automatically. The software can be downloaded freely at our website.¹

IV. CONCLUSION

Cell cycle behavior analysis of cancer cells is important for understanding the mechanisms of anti-cancer drugs and for the

¹<http://www.cbi-tmhs.org/Dcelliq/index.html>

- [19] U. Adiga, R. Malladi, and R. Fernandez-Gonzalez, "High-throughput analysis of multispectral images of breast cancer tissue," *IEEE T*, vol. 15, no. 8, pp. 2259–2268, Aug. 2006.
- [20] C. Wahlby, J. Lindblad, and M. Vondrus, "Algorithms for cytoplasm segmentation of fluorescence labelled cells," *A*, vol. 24, no. 2–3, pp. 101–111, 2002.
- [21] J. Lindblad, C. Wahlby, and E. Bengtsson, "Image analysis for automatic segmentation of cytoplasm and classification of Rac1 activation," *A*, vol. 57, no. 1, pp. 22–33, Jan. 2004.
- [22] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *E*, vol. 13, no. 1, pp. 146–165, 2004.
- [23] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE T*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [24] T. Lindeberg, "Feature detection with automatic scale selection," *E*, vol. 30, no. 2, pp. 79–116, 1998.
- [25] J. O'Rourke, (*E*). New York: Cambridge Univ. Press, 1998.
- [26] M. Wang, X. Zhou, and R. W. King, "Context based mixture model for cell phase identification in automated fluorescence microscopy," *E*, vol. 8, p. 32, 2007.
- [27] G. L. Nemhauser and L. A. Wolsey, *E*. New York: Wiley, 1988.
- [28] L. A. Wolsey, *E*. New York: Wiley, 1998.
- [29] J. Yan, X. Zhou, and Q. Yang, "An effective system for optical microscopy cell image segmentation, tracking and cell phase identification," in *EEE*, Atlanta, GA, 2006, pp. 1917–1920.