**Abstract**

This paper attempts at generalizing detection and tracking of biological cells in time lapse microscopy. To visualize cells and their behavior, various imaging modalities are used. These modalities introduce anomalies to images such as shading artifacts and non-uniform illumination. Other complications inherent to the domain, e.g. cell residue and cytoplasm are also present. Cell motion is erratic and early detection of cell events are crucial in cases of time sensitive diseases. Manual analysis of cell events is error prone. This paper uses morphological operations along with ellipse fitting and U-net to segment images, considers centroids and distance metric of cell motion features to track and recognize mitosis events.

*Keywords: - Segmentation, Detection, Tracking, Mitosis*

**Introduction**

To gain biological insights from cell behavior, it is often necessary to identify individual cells and follow them over time. To visualize these processes, imaging techniques like phase contrast (PhC), differential interference contrast (DIC), fluorescence microscopy are employed. The image sequences contain a large number of cells over multiple frames which makes manual annotation and event detection tiresome and error prone.

To automate cell segmentation and tracking, computer vision along with machine learning methods are taken advantage of. This entails, image preprocessing, feature extraction, tracking and cell event (e.g. mitosis) recognition. Due to various imaging modalities, the resulting image sequences are prone to anomalies such as low signal to noise ratio, lingering cytoplasm, uneven illumination and shading artifacts.

The 3 datasets considered have significantly distinct features. The cells are all of different types, shapes and sizes. The cells have different boundary structure while the image contrast and intensity fluctuate dataset to dataset. The number of mitosis events also differ hugely in each dataset. Special care has been taken during preprocessing to accommodate as generic an approach as possible.

The 3 databases have some common characteristics as well. The phenomenon of residual cytoplasm or debris is present. Cells entering and leaving in the middle of image sequences is observed. Degree of overlap of cells in all dataset are also similar.

Massive steps have been taken in the direction of tackling these challenges like active-contour, mean-shift, level-set based methods but all of them have been applied to the specific type of images or require extra heuristics to be handled. In this paper, we develop a system to automate cell segmentation and tracking while considering the challenges mentioned above. Firstly, the image is preprocessed using morphological opening followed by otsu thresholding. The resulting contours are fitted with ellipse and bounding boxes for all identified cells are drawn. For tracking, the nuclei of identified cells are computed and compared against previous frame. Euclidean distance metric is used to correspond 2 cell nuclei in consecutive frames. Mitosis events are recognized by finding the distance between cell nuclei. Speed of a user specified cell along with total distance travelled up to that point and total displacement are also calculated.

**Literature Review**

Massive steps have been taken in the direction of tackling these challenges like active-contour, mean-shift, level-set based methods but all of them have been applied to the specific type of images or require extra heuristics to be handled. Cells were detected based on either intensity, texture, or gradient and then linked in two or more frames. The graph-based technique is used to define the relationship between the cells, but this technique is computationally expensive especially when cell density is high. Segmentation errors were also not addressed in post-processing. Hence to overcome all these difficulties another author introduces a method that incorporates different filters for segmentation such as top-hat, h-maxima transformation. Topological features, motion features were to get accurate cell trajectories along with template matching based backward tracking to recover broken paths (EDIT)

From: [https://sci-hub.tw/https://doi.org/10.1016/j.semcdb.2009.07.004](https://sci-hub.tw/https:/doi.org/10.1016/j.semcdb.2009.07.004)    
[1] E. Meijering, O. Dzyubachyk, I. Smal, W. A. van Cappellen. Tracking in cell and developmental biology. Seminars in Cell and Developmental Biology, vol. 20, no. 8, pp. 894-902, October 2009. <https://doi.org/10.1016/j.semcdb.2009.07.004>

Thresholding is commonly used for segmentation due to its ease of use but is error prone. Due to noise in the image caused by various microscopy techniques, more complex methods are required. (template matching, watershed transform, model-based segment merging of the techniques)

Particle detection: First identify loci by using intensity thresholding and then to estimate their positions by computing the centroid. However, not as good as using Gaussian fitting which involves searching for loci with similar intensity profiles of the particle. (Better methods include Laplacian of Gaussian filtering, machine learning)

Linking particles: Due to various cell complications, ‘nearest-neighbour’ may not be optimal. Suggested methods are to use spatiotemporal segmentation approach to search for optimal paths or to use cost matrix from detected particles and their similarity (between 2+ frames) to get the optimal subgraph.

There are many tracking methods utilized for different cases. Methods such as Boosting tracking require training at runtime for positive and negative examples of an object. It needs to learn how an object looks like. However, with cells, their shape is irregular and change over time, even perform mitosis which basically form 2 new objects to track. Thus, the method would prove not be suitable for the task. Centroid tracking focuses on tracking points of an objects and measure discrepancies between the two trajectories. With our task this seemed more suitable thus we decide to adapt it for our purpose with a simpler version due to time constraints.

-top hat with h-maxima transformation for nuclei segmentation

-adaptive thresholding to remove noise from images

-watershed to separate cells with connected boundaries

-morphological opening and closing erosion and dilation to make boundaries visible

-min-max filtering to remove unwanted background

-ellipse fitting for cell localization used to localize nuclei in a cell which has maximum intensity

-rectangle fitting for cell localization fits a rectangular box around the cell

-fast radial symmetry for selecting the regions of interest

-gaussian blurring to remove the noise.

Since the success of AlexNets (Krizhevsky, Sutskever, & Hinton, 2017) that created a paradigm shift towards the use of deep learning in computer vision, convolutional neural network (CNN), a specific structure of deep learning that involves convolutional layers, has quickly found its way into medical imaging analysis and become a methodology of choice (Litjens et. al., 2017). It has not come as a surprising development, as segmentation is an extension to the classification framework using CNNs and is also a common task found in the analysis of both natural and medical images.

A feature commonly found in a CNN infrastructure is (max) pooling, which acts to reduce the spatial size when representing an input image, in order to reduce the number of parameters, computation burdens, and overfitting (Stanford, 2020). However, these benefits do not come without cost: a loss of effective spatial resolution becomes an issue for dense problems like semantic segmentation as high-frequency details are washed out, leading to blurry object boundaries (Marmanis, et al., 2018). One of the most successful methods to prevent this decrease in resolution is a method called ‘Shift-and-stitch’ (Long, Shelhamer, & Darrell, 2015), where a fully CNN (fCNN) is applied to shifted versions of the input image, and the results are stitched together. The idea of fCNN was taken a step further by (Olaf, Philipp, & Thomas, 2015) with the U-net architecture, comprising a ‘regular’ fCNN followed by ‘up’-convolutions that are used to increase the image resolution. While learned upsampling layers have been proposed before (Long, Shelhamer, & Darrell, 2015), the novelty of U-net involves combining opposing convolution and deconvolution layers with so-called skip connections. This allows an entire image to be processed by U-net in one forward pass, which helps consider the full context of the image (Litjens et. al., 2017). U-Net was specifically developed for single-cell analysis, and since its introduction, it has become one of the most widely used CNN architectures in medical image analysis. Other popular deep learning methods that have been successfully applied to a variety of data types include DeepCell and Mask R-CNN, while promising results have been recently obtained by generative and vector embedding approaches (Moen, et al., 2019).

**Methods**

For the cell tracking challenge, the dataset given has three different type of cells that would require specific processing because of the variation in intensities. The difference between background and foreground is indistinguishable in some cases and the presence of cytoplasm and noise is unavoidable. There may be some discrepancy due to the image capturing technique or the duration over which the images were taken. Apart from the noise and unavoidable faults another problem encountered is the motion of cells as there are no laws governing cell motion. To overcome all these challenges, a system that is made of three parts, pre-processing module, detection module and tracking module is introduced.

For the DIC-C2DH-HeLa dataset, in addition to the challenges arising from low signal-to-noise (SNR) and contrast ratio (CR) values and high cell density, it also has the added difficulty for having complex image texture of the cells (Ulman et. al., 2017). We have made several attempts using computer vision methods to detect and segment the cells in this dataset; however, the accuracy obtained from such methods is quite limited. As deep learning methods have proved their exceptional performance for this dataset, we decide to take this route, choosing a successful CNN infrastructure, U-Net, as the backbone network to train on data and make predictions. Our cell segmentation method combines deep learning with watershed segmentation as proposed by (Lux & Matula, 2019). This approach involves training a single U-Net infrastructure first, and outputs produced by this trained network are then processed by watershed transformation to obtain the final image segmentation.

The entire process is consisted of three stages:

1.    Preprocessing:

i)    Normalization of each input image by Contrast limited adaptive histogram equalization (CLAHE).

ii)    Transformation of each input image into markers, cell mask, and a weight map, the last of which is used to define the relative importance of each pixel, resulting in a weighted loss function.

2.    Predicting: Unlike the original U-Net infrastructure proposed by (Olaf, Philipp, & Thomas, 2015), the U-Net model here produces two different predictions: cell markers and pixel-wise binary classification of the input image into foreground versus background.

3.    Postprocessing: a predicted cell marker image obtained from step 2 above is thresholded to form a binary image, upon which a morphological opening operation is applied. Connected components are determined based on this new binary image, and the former are then combined with predicted cell mask (also obtained from step 2) using a marker-controlled watershed transformation to produce the final segmentation.

For task 1 of this project, instead of training this model again from scratch, we decide to use a pretrained model provided by (Lux & Matula, 2019) to obtain the segmentation masks for sequence 3 and 4, since this model was trained using the same dataset (sequence 1 and 2) provided by the Cell Tracking Challenge.

*Pre-processing Module:*

Dataset 2 and 3 have significantly distinct characteristics. Number of cells are greater, as a result mitosis events and overlap scenarios are more.

Traditional methods are best suited to process and segment images from dataset 2 and 3. According to the proposed method, each image from an image sequence in the dataset are read and converted to grayscale. Small cell residue or noise can be observed in the background. Empirically, a 3x3 structuring element is used to erode the grayscale image to cater this challenge. Even this structuring element for erosion resulted in reducing the size of some cells to a single dot. Hence, we dilate the image to increase the size and therefore make cell body distinguishable. The structuring element used was a 5x5 filter. The size of the filter was taken after considering of pixel intensities and helps in brightening the cell interior. The dilated image has variation in cell intensities. Hence, thresholding is done regionally resulting in different threshold values for different regions. Empirically, the region size taken is 3 pixels. The resulting image distinguishes cell boundary. To complete edges and solidify intensity for cell interior, the dilated image is subtracted from the resulting threshold-ed image. Since all cells are darker than the background, Otsu thresholding is employed to consider the binodal nature of image pixel intensities. We aim to increase separation in intensities between background and foreground. If cells were brighter than the background, the resulting image would display foreground in white and background in black. In the proposed method, since the cells observed at the last step were darker than the background, the cells appear black after Otsu thresholding which can be corrected by inverting the image.

To apply connected components and finally label all cells, we erode the final image first to reduce overlap. Empirically, a 2x2 filter with 3 iteration was found to be optimal in terms of information loss and cell boundary overlap reduction. Connectivity chosen for connected components algorithm is 4 – connected objects.

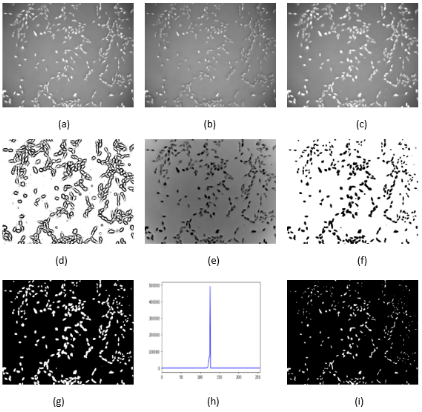


Figure 1: Illustration of illumination correction and noise removal of the input image. (a) Original frame in PhC-C2DL-PSC dataset. (b) Eroded image. (c) Dilated image. (d) Adaptive thresholded image. (e) After background removal. (f) Otsu thresholded image. (g) Inverted image. (h) pixel intensity, number of pixels (x, y) histogram for image (e). (i) Final erosion to remove any noise and overlapping.

Figure 1 illustrates the morphological and thresholding operations performed. Fig 1(a) is a frame in dataset PhC-C2DL-PSC Sequence 2. Figs 1(b) and (c) are erosion and dilation for noise removal and cell brightening. Fig 1(d) is the regionally thresholded image displaying cell boundary. To solidify cell structure, fig1(c) is subtracted from fig1(d). As evident from fig 1(e) a cell has a single intensity. To support this, fig 1(h) displays 2 peaks. The higher background peak and a smaller peak in its left neighborhood denoting the black identified cells. This binodal phenomenon is treated by Otsu thresholding as displayed in fig 1(f). Fig 1(g) inverts the previous image to identify cells as white. To detect overlapping cells, erosion was used the result of which can be seen in fig 1(i).

*Detection Module:*

To accurately detect cells, unique labels have been created for each cell in the pre – processing module. In detection module, due emphasis has been laid on cell structure and mitosis events. An approximation of cell structure as an ellipse rather than fitting the exact cell structure, which can change in any frame, as a technique to make bounding boxes has been employed. Distance between cell radii have been compared to recognise mitosis events.

To define cell structure, the shape of cell over its lifetime has been approximated to ellipses. A unique label for each cell was acquired as a result of preprocessing module. The labels mark each cell with a unique number starting from 1, where 0 is for background. In the proposed method, contours are drawn for each uniquely identified label and are checked for shape. Hence the bounding boxes are created. Centre points of ellipse along with semi major axis are used for mitosis detection and tracking.

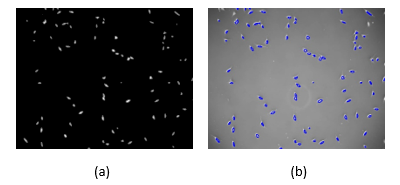


Figure 2: Illustration of cell segmentation and bounding box fitting cell structure in dataset PhC-C2DL-PSC dataset. (a) Bounding boxes on thresholded image. (b) Bounding boxes on original image

Automatic mitosis detection is crucial in the tracking process of the cells. When provided with a large dataset, it is extremely tedious and time-consuming to identify and count each instance of mitosis detection when labelling new cells. The effectiveness of automatic mitosis detection also depends very heavily on the segmentation results as the algorithm works separately and assumes that segmentation is done flawlessly. It can be foreseen that some noise from the images that are not successfully filtered out might cause inaccuracies in the detection process. It is observed from the images that mitotic cells share certain minuscule traits that other individual cells do not. From what is seen, cells that went through mitosis generally stay relatively stationary compared to their counterparts. Obviously, when mitosis takes place, another elliptical oval or circular shape can be seen budding out of the original parent cell. In this way, there will be two nuclei or rather two centroids in this cell structure.

The method that was incorporated to detect mitotic cells is to use the Euclidean distance between the cell centroid. Although the nucleus is not detected, it can be safe to assume that the centroid of the cell has the highest probability of containing the nucleus. The cells are being eroded so that it is corresponding radiuses and contours are as tight to the centroid as possible. Then, the Euclidean distances between each cell centers are tracked and stored. Depending on the type of cells in each dataset, a pre-determined standard distance is being fixed. This is because, in different datasets, the cell sizes are different and so are their proximity within each other. If the Euclidean distance is smaller than the predetermined distance, then it can be assumed that mitosis is likely to have occurred. The mother cell can be identified as being exiting first before the child cells appear and a new cell id is given. In which case, a bounding circle shall be placed around them for easier identification.

*Tracking Module:*

Centres and semi major axis detected to define cell structure come in play in the tracking module. The tracking module tracks cells based on cell centroids, computes and compares Euclidean distance between cell centroids to associate same cells in consecutive frames. The proposed method checks if any cells are being currently tracked, if not, it registers the centroids resulting from pre-processing module and assigns the unique ids to cell centroids. If cells were being tracked in the previous frame and centroids are presented in the current frame, the algorithm determines Euclidean distance between the cell centroids of previous and current frame. Main assumption for tracking made is that the cells will potentially move in successive time frames, but the distance travelled by cell between frames will be smaller than distance between other objects detected. This helps in avoiding tracking of fast-moving residual cytoplasm in image sequences. If cells being tracked in the previous frame are more than the number of centroids provided by the pre-processing module for current frame, algorithm determines the disappearing cells and deregisters them if for n consecutive frames, the cell does not re-appear for which n is a pre-defined number of frames. If number of centroids in current frame are greater than cells being tracked in the previous frame, the algorithm registers the new cells found.

After associating two pairs of centroids, the centroids are used to calculate and store the speed, distances and confinement ratio of the cell. It also uses the centroids to draw the trajectory path of the cell.

**Experimental Setup**

Datasets considered are unique, such as cell size, count, motion etc, and are alike, such as background noise, cell overlap etc, in a lot of aspects.

**Conclusion**

In this project, the first significant task involves detecting and segmenting cells, for which we have made several attempts using both computer vision (for all datasets) and deep learning method (for DIC-C2DH-HeLa).

There are several important lessons that we draw from this experience. First, as traditional computer vision techniques are usually fast and allow practitioners to be in control of each step in the entire process while being able to produce satisfactory results, they should always be a first stop for exploratory experiments. Second, the fact that all the top 3 results for the challenging DIC-C2DH-HeLa dataset are produced by a deep learning framework has reaffirmed the success of deep learning in biomedical applications, particularly for complex datasets. It is also worth noting that the work by (Lux & Matula, 2019), which has produced a top 2 result for this dataset, uses a combination of both computer vision and deep learning in a novel approach. Their work has led to an interesting avenue for future research: rather than being viewed and used as two separate tools, these two methods can be working in tandem and complementing each other to produce exceptional results.

With regards to the mitosis detection and cell tracking, we employ a simple method that relies on a critical assumption of cell behaviour for both problems, using spatial information (proxied by the Euclidean distance) between cells from one frame to the next. While it is able to yield results of a certain accuracy level, we acknowledge the drawback of this approach in situations where cells have rapid movements or live in a dense population (Meijering, Dzyubachyk, Smal, & Cappellen, Tracking in cell and developmental biology, 2009). While it is possible to incorporate several other characteristics of cells such as intensity, volume, orientation, etc. to calculate the ‘nearest distance’, it should be noted that the characteristics of cells are quite dynamic across time and also diverse in different datasets, making it quite challenging to impose certain universal assumptions to solve the mitosis and tracking problem. Since time-lapse microscopy images brings about the added temporal information of cells, their division and movements, several traditional computational methods have taken advantage of this extra information. An alternative framework to make use of the combined spatial-temporal information is deep learning methods, which can learn intrinsic, discriminative features of mitotic cell images from the input. As a result, a CNN model combined with the recurrent neural network (RNN) or graphical model is a potential candidate to further enhance performance (Liu, Lu, Chen, & Su, 2017).