Using Short Term Fourier Transform to Identify Cell Boundary From Microscopic Images

Chaolun Wang

Abstract Cells are known to be the building blocks of every living organisms. Many observations and experimental approaches have been taken to exploring the cellular processes such as mitosis and cell migration. It would be tempting to apply the computational method on the study of cell biology. One of the obstacle for building a computational model of the cell is lacking of data. Since measuring the parameters of a cell is demanding, most frequently, the only data available is the microscopic image. Therefore, it would be necessary to implement data mining techniques on microscopic images of cells. In this study, a method was developed to identify the cell membrane from microscopic images. Short term Fourier transform(STFT), DBSCAN and concave hull algorithm has been used in this project. By identifying the pixels which have a surrounding of high frequency noise, we can separate the background from the region inside the cell. The DBSCAN and concave hull method was used to remove noise and approximate the boundary of cell.

Keywords Cell modeling, Phase-field model, Clustering, Image processing, Short term Fourier transform, DBSCAN, Concave hull

1 Introduction

Every living organisms are composed of cells. Since the cell theory been proposed, many observations and experiments has been taken to elucidating the structure and function of cells. However, as to the computational approaches, one of the greatest obstacle is the lacking of data[1–3]. Because measuring the parameters, such as the salt concentration, mass, and surface area, from a cell is demanding and sometimes even not approachable, most of the computational approaches simulating the cellular processes are using user defined parameters.

The validation of the computational model is usually done by comparing the simulation result with the microscopic images, which, most frequently, is the only data available. It would be necessary to use the data mining techniques on the microscopic images in order to build the computational model more approximate to reality.

Here, a program is developed which can identify and interpolate the boundary of the cell from microscopic image. Although the microscopic images are varies a lot based on the properties of microscope and the method of dying or mounting, a well taken microscopic images of the cell always has a very clear and sharp boundary. Also, since the cells contain complicate substructures, such as nuclei, cytoskeleton and organelles, the inside part of the cell has higher complexity compared to the background. This character has been used in this study to separate the inside from the outside of the cell. The method used for this purpose is the short term Fourier transform. The pixel inside the cell has the neighborhood of high frequency noise compared with the pixel at the back ground. Thus by separating the whole image into small segments and do local Fourier transform separately, we can find out the region with spectrums which are stronger and with higher frequency. Then, the DBSCAN was used to remove noise and separate the cells from each other if there are multiple cells in the image. A concave hull algorithm was used to interpolate the boundary of the cell with segments.

The function which represents the cell boundary can be used to build a energy phase-field model. Figure 1 shows the result of the phase-field method in a previously study about cell migration. An energy phase-field method can be used to study the elastic bending energy of the cell membrane. Given the shape of cell boundary, the phase-field model can provide us the information about the behavior of cytoskeleton inside the cell. Also we can use the model to study the morphological change of the cell through time, and the program developed in this project will help to give a better initial condition for our future numerical simulation. We also saw a strong potential to apply the short term Fourier transform on the detection of other life related signal.

2 Methodologly

The whole project has been separated into three parts: firstly, the short term Fourier transform was used to scale the complexity of different regions of the microscopic image, a weight function was used to stress on the high frequency spectrum; After that, the DBSCAN method was used to remove noise and separate cells from each others if multiple cells exists; After that, a concave hull algorithm was used to interpolate the boundary of the cell.

2.1 STFT method to scale and visualize complexity In order to perform a clustering method, it is necessary to firstly tell the difference between the inside and outside of the cell, which has different level of complexity. Here,

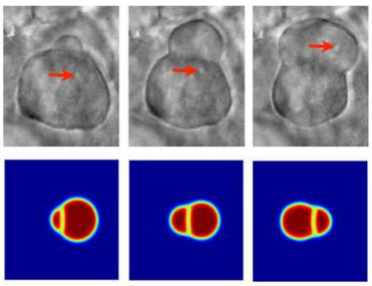

Figure 1: Comparison of our simulations with laboratory observations. First row: the image of cell movement of zebrafish primordial germ cell from [4]. Second row: the pictures from the multi-phase-field simulation at time t=0.1, 0.3, 0.525.

the STFT was used to scale and visualize the complexity[7]. The process is as following: firstly, separate the picture into multiple n by n segments, where n is a user defined variable which was set to be 9 in this study. In order to improve the resolution, there can be some regions of overlapping between the segments. Then the 2D center fast Fourier transform(FFT) was performed to each of the segments in order to get the spectrum. The value of spectrum for each segments have been multiplied by a weight and added together, this value was used to indicate the complexity level of the segment. The complexity level can be visualize using 2D heatmap.

2.2 DBSCAN method for noise removal and cell separation A clustering technique named DBSCAN has been used to remove noise and separate cells from each others. A threshold was set to select the segments with high complexity level(50 percent of the maximum level in this study), Therefor, the region inside or at the boundary of the cell are expected to have high density of high complexity segment. The DBSCAN is a density based clustering method, it is good to separate the cluster with complicate shape, also it shows a high resistant to noise, thus this method is highly suitable for this study. An algorithm was implemented based on the research from Martin et.al[5]. A segment was considered to belong to a cluster if it have more than a certain number of neighbors belong to the same cluster. Also, if for a segment there are not enough neighbors, it will be treated as noise. The radius specifying the neighborhood was set as 5 segment unit and the threshold of number of points inside the neighborhood was set to be 9 in this study.

2.3 Concave hull algorithm to interpolate the boundary of cell After removing the noise and outliers and identifying the cluster which belongs to one cell, the concave hull algorithm[6] was used for the interpolation of the cell boundary. Firstly, the boundary of segments with high complexity was identified by using a classic convex hull algorithm. Then the longest edges from the

convex hull have been deleted iteratively and replaced by shorter edges, which do not cross the preexisted edges and also do not form large angles. The edges will be considered short enough if they are smaller than 10 segment units, and an angle was considered as large angles if they are larger than $\pi/3$. The algorithm will stop when all of the edges inside the concave hull is smaller than 10 segment units. Using this method, we can interpolate the cell boundary piecewisely with segments. If the higher order interpolation is preferred, a B-spline method can be used instead.

3 Results

3.1 Complexity visualization The figures have been divided into 9×9 segments and the STFT has been performed for each of the segments. The complexity level has been calculated based on the spectral patterns. The complexity level of the cell under mitosis and the diatom cells have been shown in Figure 2. From the complexity visualization, we can clearly see that the region with high complexity are exclusively concentrated at the region corresponding to the inside part of the cell. Also, the region near the cell membrane appears to have higher level of complexity compared to other cellular components, this phenomenon is consistent to the fact that many cellular components such as cytoskeleton are highly concentrated near cell membrane, which formed the cortex layer, this kind of structure can contribute to the complexity of this region.

Except from the result in Figure 2, this method has been tested using many other microscopic image which was taken under different types of microscope, with different cell types, cell concentration and ways of mounting and dieing(data not shown). This method has the capability to correctly separate the cell from the background in all of these situations and shows a high level of resistance to noise.

3.2 Clustering and noise removal The DBSCAN has been used for noise removal and cell separation. In this study, the radius specifying the neighborhood was set as 5 segment unit and the threshold of number of points inside the neighborhood was set to be 9. This method has been tested using a wide variety of microscopic images. In most of the case, the method works as expected and can separate the segments into different clusters based on which cell they belongs to, an example is the case in first row of Figure 3. It also shows a strong resistance to background noise, since the noise has an relatively low density of high complexity segments.

This method fail to work properly, however, when the cells are very close to each other(as indicated in second row of Figure 3). In this case, since the

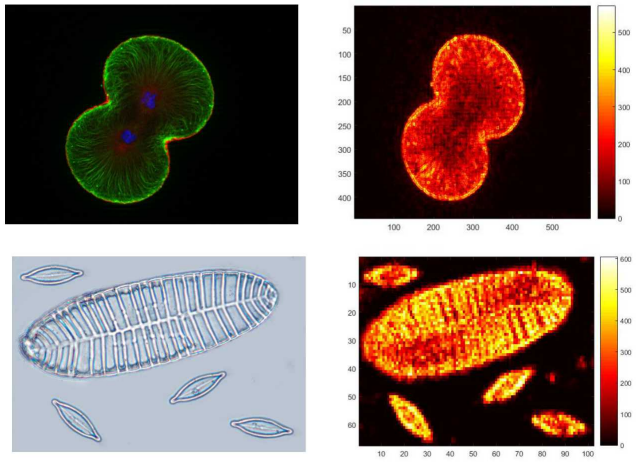

Figure 2: Compare the original microscopic figure with complexity visualization, the region more white in color have higher level of complexity. First row: the image of cell division and its complexity visualization. Second row: the pictures of diatom cells and their complexity visualization

nodes belong to the boundary of the cells are neighboring with each other, the algorithm will treated them as one cluster. This obstacle can be overcomed by adjusting the radius which defines the neighborhood, however, if we keeps reducing the radius, the number of points for each of the cluster will also reduce dramatically. Therefore, this method is more appropriate to be used if the cells in image have certain distance between each other.

3.3 Boundary interpolation The boundary of the cluster has been interpolated using the concave hull algorithm, since some of the cells may have complicate shapes. The result has been shown in Figure 4. We can see that this algorithm works well for the case that the cell boundary is not a convex shape(first row of Figure 4). It correctly interpolates the cell boundary. For the case where there are multiple cells in a figure(second row of Figure 4), this method can also be done by specifying a certain cluster which the user is interested in.

The density of data points in side the cluster is crucial for the performance of the concave hull method. The higher the density is, the boundary of the cluster can also be interpolated with a higher resolution. Since the STFT method tends to have the high complexity region concentrated on the boundary of the cell(first figure of Figure 4), the convex hull is very suitable for the boundary interpolation in this case.

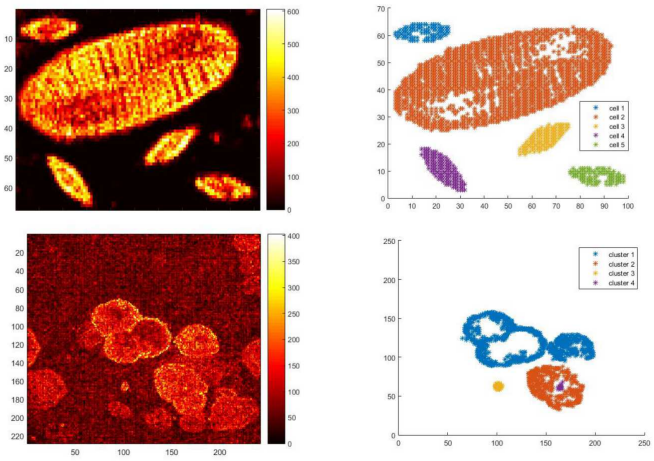

Figure 3: The result of DBSCAN. First row: DBSCAN was used to separate the cluster belongs to each diatom cells, each of the cells has been identified successfully and shown in different colors. Second row: the DBSCAN cannot correctly separate the tumor cells since they are close to each other.

4 Conclusion.

In this project, I successfully implemented a program which can be used to find out the boundary of cell using the microscopic image. By using the approach of STFT, the program can efficiently separate the inside of the cell from the outside of the cell via their difference in complexity. The DBSCAN and concave hull algorithm was then used for cell separation and boundary interpolation. This program is highly stable and can be used to identify the cell boundary from a large variety of microscopic image. As to the efficiency, since the program was implemented in c++, the interpolation of cell boundary can be done within 30 seconds given a figure of jpg format with 1 MB in size.

This project is helpful for my future research about computational cell biology. By interpolating the cell boundary with high resolution, we can build a phase-field model with the initial condition which is more approximate to the observation from laboratory. Also, the work done in this project makes the inverse modeling possible: given the actual shape of a cell, we can use the phase-field model to calculate the surface tension at the cell membrane, it will provides the information about the underlying cytoskeletal structure of cell cortex. This approach will also be helpful to the identification of different cell types, especially in the study of tumor cell migration[8].

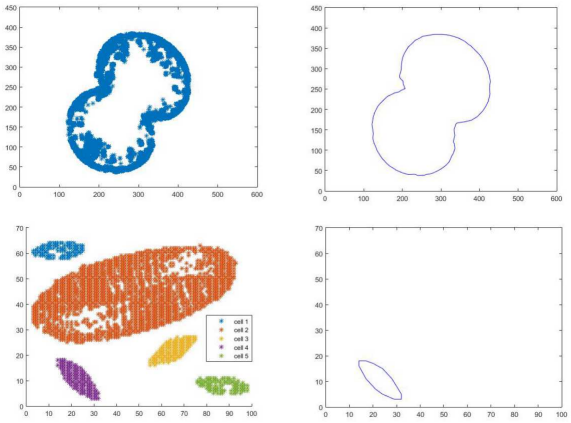

Figure 4: The result of concave hull algorithm. First row: the interpolation of cell boundary of single cell under mitosis. Second row: the interpolation of cell boundary of the diatom cell form figures with multiple cells

References

[1] H. Noguchi and G. Gompper, *Shape transitions of uid vesicles and red blood cells in capillary flows*, Natl.Acad.Sci, 2005.

[2] H. Noguchi and G. Gompper, *Dynamic of fluid vesicles in shear flow: Effect of membrane viscosity and thermal uctuations*, Phys Rev E, 2005.

[3] V. Noireaux, R. A.Goldsteyn, E. Friederich, J. Prost, C. Antony, D. Louvard and C. Sykes, *Growing an actin gel on spherical surfaces.* Biophys. J. 278, 1643C1654, 2000.

[4] H. Blaser, M. Reichman-Fried, I. Castanon and E. Raz, *Migration of zebrafish primordial germ cells: A role for myosin contraction and cytoplasmic flow*, Developmental Cell 11 (2006) 613-627.

[5] E. Martin, K. Hans and X. Xiaowei, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, Knowledge Discovery and Data Mining (1996) KDD-96.

[6] D. Matt, K. Lars and W. Mike, *Efficient generation of simple polygons for characterizing the shape of a set of points in the plane* ,URL:http://www.geosensor.net/papers/duckham08.PR.pdf (2008)

[7] S. Ervin, D. Igor and J. Jiang, *Time-Frequency Feature Representation Using Energy Concentration: An Overview of Recent Advances*, Digital signal processing (2009) 153-183.

[8] P. Friedl and K. Wolf, *Tumour-cell invasion and migration: Diversity and escape mechanisms*, Nature 3 (2003) 362-374.