SINGLE CELL TRACKING OF *SYNECHOCYSTIS* GROWTH IN A MICROFLUIDIC CULTURE DEVICE USING A PROBABILISTIC AUTOMATED IMAGE ANALYSIS TECHNIQUE

F. Yu[1*], K. Song[1], M.A. Horowitz[1], S.R. Quake[1]
[1]Stanford University, U.S.A.

ABSTRACT

Cyanobacterium *Synechocystis sp.* PCC 6803 is a model organism for photosynthesis and play an important role in biofuel research. Single cell analysis of *Synechocystis* can be used to understand how growth characteristics change under different environments or to screen for advantageous phenotypes. This work introduces an automated probabilistic image analysis method for analyzing time lapse images of *Synechocystis*. Combined with a microfluidic cell culture platform previously developed in the group, this method can be used to generate high throughput and robust results of cell growth characteristics.

KEYWORDS: Image Analysis, *Synechocystis*, Single Cell Analysis

INTRODUCTION

Synechocystis sp. is a spherical, fresh water cyanobacterium (Fig. 1abc). It is a model system for studying photosynthetic proteins and complexes such as photosystem I and II, phycobilisome, and chlorophyll. As a prokaryotic system that is amenable to genetic transformations, *Synechocystis* is an attractive candidate for biofuel generation. To characterize bacterial growth at a single cell resolution, we use a microfluidic cell culture device that can perform high throughput culturing experiments with independently addressable environmental parameters [2]. Image sequences of each chamber were captured at high temporal resolutions over several days. In order to extract quantitative data regarding cell growth and motility patterns from these image stacks, we employ a probabilistic image analysis method that uses information from each image as well as temporal correlation between images within a set of time lapsed frames.

Single cell studies reveal individual-to-individual heterogeneity not reflected from population averages. Bacterial single cell lineage tracking is particularly interesting because it can serve as a measurement tool for evaluating environmental or genetic perturbations to cellular metabolism. Even though groups have demonstrated single bacterial lineage tracking algorithms [1], some require finely tuned thresholds and manual post analysis corrections. In this work, we propose an image analysis method from time lapsed images of Cyanobacterium *Synechocystis sp.* PCC 6803 growing in a microfluidic cell culture device using a probabilistic framework. Such a technique decreases the extent of required manual corrections and threshold tuning.

THEORY

There are three main steps in the image analysis pipeline that is described in this work. The first step involves segmentation. Our segmentation uses Minimally Stable Extremal Regions (MSER) that is insensitive to small background variations and deviations from the focal plane [3]. As a result, artifacts such as non-uniform lighting and small variations in z from frame to frame of a video would not significantly affect quality of processed data. Results from segmentation is a black and white image outlining the contours of the cell clusters from which we can extract a certain set of features representing each cluster (Fig. 1d). These features include area, eccentricity, and circularity. Using a set of training images, we identify the distribution of these features for single, double, and many cell clusters (Fig. 2abc).

In the second step, we classify each cluster of white pixels found in the segmented image as single, double, or many-cell clusters. Each probability is calculated as a product of feature probabilities using the distribution obtained from the test set as a reference (Eqn. 1). Instead of making a definitive decision, however, probabilities are generated that describes the confidence of classifying a cluster into a particular group. If probabilities are low, the classification decision is post-phoned until more definitive images are acquired at a later time.

In order to incorporate information from later images, we extract a distance matrix and compute a correlation score related to the exponential of negative distances between pairwise clusters from different images (Eqn. 2). We use this function because it is bounded. Using this score, we link clusters of cells from adjacent images and account for splitting cells using area constraints. We then extract the distribution of distances between centers of two cells forming doublets (Fig. 2d). Mean and standard deviation of the distance are used to determine the precise number of cells in a cluster. Since pinching off of the membrane in *Synechocystis* division occurs very slowly, hard threshold performed per image often yield errors in cell counts. Therefore, we first correlate clusters of cells from adjacent images using previously described distance metric and then use temporal information to determine cell numbers (Fig. 3c).

$$P_i = P_{area(i)} P_{eccentricity(i)} P_{circularity(i)} \tag{1}$$

$$Correlation\ Score = \exp\left(-norm\left(Center_{region\ k, frame\ i} - Center_{region\ k, frame\ i+1}\right)\right) \tag{2}$$

EXPERIMENTAL

Before seeding into the microfluidic cell culture device, cells of Cyanobacterium *Synechocystis sp.* is grown in a 30C incubator. Culturing experiments of *Synechocystis* cells are performed in a microfluidic cell culture chip designed by the Quake group at Stanford. Before each experiment, the microfluidic chip is flushed with water to passivate uncured PDMS (Polydimethylsiloxane) bonds. A pluronic solution is used to treat channels to prevent cells from sticking to the surfaces. Then, chambers are treated with polylysine to promote attachment of cells to surfaces and to facilitate imaging. Cells in exponential phase is extracted and seeded at a concentration of 4 cells per nanoliter. During the experiment, images are acquired every 6-10 minutes from all chambers of the microfluidic chip. A MATLAB based control script coordinates all components of the cell culture system including the microscope, stage, camera, lighting, and determines when images are acquired from each chamber and where they are stored (Fig. 1e). The goal of the image analysis pipeline is to identify cell number and positions in a stack of time lapsed images, which are then used in studying growth and division characteristics.

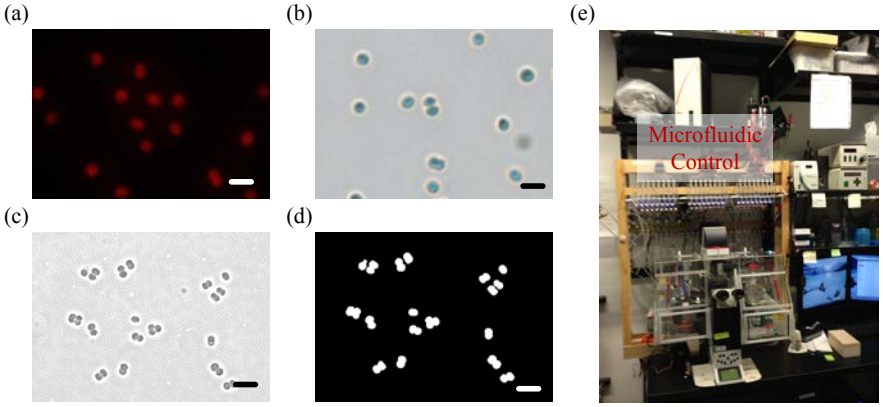

Figure 1: (a) Fluorescent image of Cyanobacterium Synechocystis sp. PCC 6803 auto fluorescence. All scale bars represent 5 µm. (b, c) Bright field images of Synechocystis, showing the green chlorophyll pigments in b. (d) Segmentation result using MSER on grey scale bright field image represented by c. (e) Microfluidic cell culture system used to operate the PDMS microfluidic cell culture chip.

RESULTS AND DISCUSSION

In any image analysis pipeline, segmentation is often required in order to separate the images into regions corresponding to cells of interest. Then, depending on the parameters extracted, cells are often classified into different categories based on extracted properties such as location, size, shape etc. These properties are then used for downstream processing. Difficulties encountered with using typical image analysis software on *Synechocystis* images are often due to two reasons: the fact that *Synechocystis* cells are small (~2 um in diameter) and spherical, and that cells tend to cluster to form 3 dimensional structures. Therefore, most image analysis packages that uses a deterministic pipeline would not be able to correct for mistakes made from a few intermediate images.

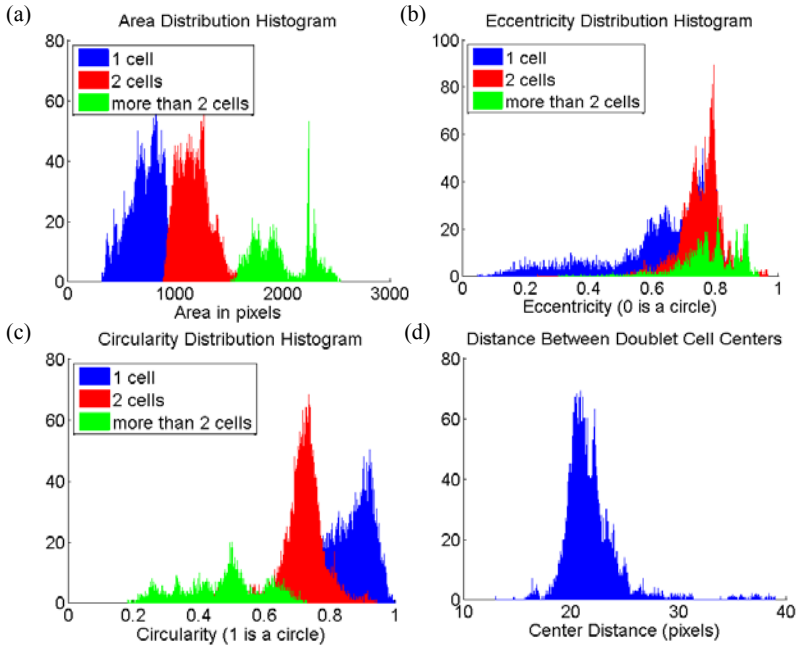

Figure 2: Distributions of features extracted from segmented images of Synechocystis. Distributions are split into single, double, or many-cell clusters. (a) area distribution (b) eccentricity distribution (c) circularity defined as $4\pi A/P^2$ (d) distribution of distances between cell centers from a doublet cell cluster. This distribution is used to determine probability of cell numbers inside a cluster when cell number per cluster is larger than 2.

Using MSER, we are able to extract regions representing clusters of *Synechocystis* cells. The next step tries to classify these regions into single cell, doublet, or multi-cell clusters based on properties of these clusters such as area, eccentricity and circularity (Fig. 2abc). Instead of making a definitive decision on which group a certain cluster belongs to, the proposed method computes probabilities of each cluster belonging to each respective group, thus retaining the possibility of correcting prior classification errors. If the confidence of a cluster belonging to a particular group is high, then the cluster is designated to that group. Otherwise, decision is post-phoned to future frames. Therefore, in addition to using information from each image, this method also incorporates information from later images, thus increasing the amount of information upon which a decision is made.

In order to correctly use temporal data, clusters in time lapse images must be linked. Correlation between clusters from different images of the same set of cells is linked using a distance metric described earlier. Based on this score, new regions that are generated due to doublets splitting can also be accounted for. As shown in Fig. 3b, some regions are newly generated and can be correctly linked to regions in the previous image using area constraints. One such constraint requires areas of linked regions in adjacent images be equal. Finally, using typical distance distributions between cell centers in a doublet (Fig. 2d), the precise number of cells in a cluster can be obtained. Again, due to noise inherent in all image analysis methods, this number could vary from frame to frame. By assigning probabilities to each cluster having certain number of cells, we can again post-phone decisions to future frames when a high confidence result is obtained. Fig. 3c illustrates three sample single cell growth curves obtained. Compared to bulk experiments, single cell measurements provides more insight into heterogeneity of a population and the individual to individual temporal variations in terms of growth rates. Thus, these curves can be used to compare cell fitness under different environmental conditions.

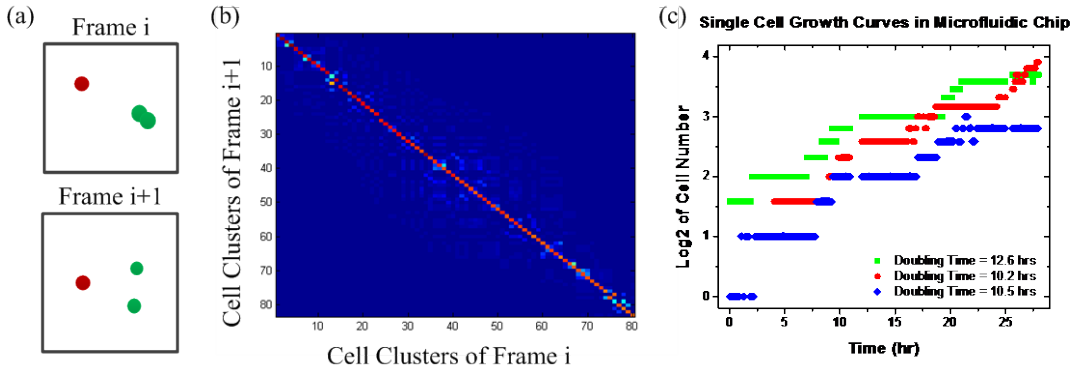

Figure 3: (a) Correlation of cell clusters from frame I to frame i+1 is performed using a distance metric based on the exponential of negative Euclidean distances between cluster centers. Such a metric can also be used to identify new clusters that are generated. (b) Sample heat map showing the correlation matrix between two adjacent frames. Each pixel represents the distance correlation score of two clusters from respective frames. (c) Single cell growth curve of 3 representative lineages of Synechocystis using the proposed image analysis pipeline.

CONCLUSION

We demonstrate an image analysis pipeline that can be used to robustly analyze time lapse images of Cyanobacterium *Synechocystis sp.* PCC 6803. Combined with the PDMS cell culture chip, this system preserves its advantages in terms of high throughput, spatial and temporal resolutions. After segmentation, the proposed method classifies each cluster found from an image using information from current and later images. This way, decisions can be post-phoned if the probability of such a decision is low. In conjunction with the cell culture system, such methods can be used to study Cyanobacterial growth and motility.

REFERENCES

[1] J.C.W. Locke, M.B. Elowitz, "Using movies to analyse gene circuit dynamics in single cells," *Nat. Reviews,* 7, 383-392 (2009)

[2] R. Gomez-Sjoberg, A.A. Leyrat, D.M. Pirone, C.S. Chen, S.R. Quake, "Versatile, fully automated, microfluidic cell culture system," *Anal. Chem.,* 79-22, 8557-8563 (2007)

[3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, "A comparison of affine region detectors," *Int J. Comput Vision,* 65(1/2), 43-72 (2005)

CONTACT

* F. Yu, brianyu@stanford.edu