Detecting cell division of *Pseudomonas aeruginosa* bacteria from bright-field microscopy images with Hidden Conditional Random Fields

Lee-Ling S. Ong[1], Xinghua Zhang[2], Binu Kundukad[1], Justin Dauwels[3], Patrick Doyle[1,4] and H. Harry Asada[1,4]

***Abstract*— An approach to automatically detect bacteria division with temporal models is presented. To understand how bacteria migrate and proliferate to form complex multicellular behaviours such as biofilms, it is desirable to track individual bacteria and detect cell division events. Unlike eukaryotic cells, prokaryotic cells such as bacteria lack distinctive features, causing bacteria division difficult to detect in a single image frame. Furthermore, bacteria may detach, migrate close to other bacteria and may orientate themselves at an angle to the horizontal plane. Our system trains a hidden conditional random field (HCRF) model from tracked and aligned bacteria division sequences. The HCRF model classifies a set of image frames as division or otherwise. The performance of our HCRF model is compared with a Hidden Markov Model (HMM). The results show that a HCRF classifier outperforms a HMM classifier. From $2D$ bright field microscopy data, it is a challenge to separate individual bacteria and associate observations to tracks. Automatic detection of sequences with bacteria division will improve tracking accuracy.**

I. INTRODUCTION

Biofilms are communities of bacteria embedded in extracellular polymatric substances (EPS) [12]. It can form on many surfaces such as teeth, pipes and can contaminate foreign body materials. Free floating bacteria initiate biofilm formation when its adhesion molecules bind to the selected surface. The attached bacteria either divides or recruits other bacteria via cell-cell communication to achieve a critical cell density. As the number of bacteria increase, they produce EPS, forming a 3D biofilm. To understand how individual bacteria organize themselves in the early stages of biofilm formation, we are developing a system that delivers the bacteria trajectories and division events over time.

As the colony expands, bacteria tend to migrate close to each other. More bacteria may migrate into the frame. Existing bacteria may detach [12]. Therefore, tracking each bacteria across multiple time frames becomes a challenge. By labelling division sequences in a track, we can improve the observed bacteria-to-track association. Bacteria divide at a constant rate. Therefore, it is unlikely that a bacteria will undergo two divisions within a short time frame. We can incorporate this information to determine the observed bacteria-to-track association. From a large volume of division labels, we may derive statistics on the affect of the division rate under certain conditions.

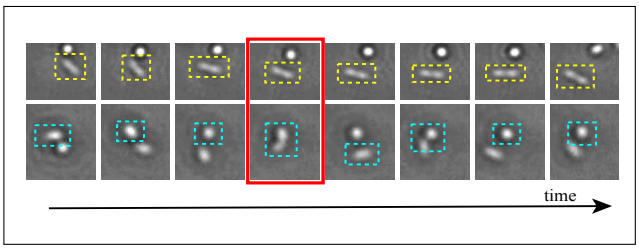


Fig. 1. The red bounding box (over the images in 4th column) shows that two juxtaposed bacteria (cyan box, bottom image) are visually similar to a bacteria undergoing division (yellow box, top image). We can classify the temporal information, which is the tracked bacteria over multiple frames, to detect and validate whether a bacteria is undergoing division or otherwise.

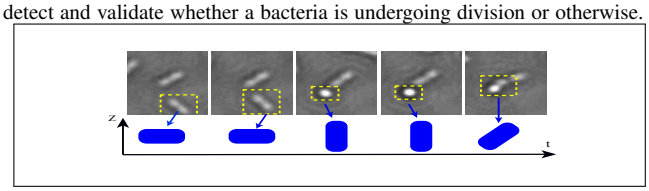


Fig. 2. Dividing bacteria may be migrate over other bacteria and orientate themselves perpendicular to the image plane.

The top images in Figure 1 shows time frames from a sequence of division. Bacteria division begins when its single DNA molecule replicates and then attaches each copy to a different part of its membrane. As division occurs, the cell lengthen, its cytoplasm is cleaved in two and a new bacteria wall is formed [1]. Cell wall formation is indicated by the region of low intensity in the cell. However, two cells which have migrated close to each other appear similar to a bacteria dividing, as shown from the red bounding box in Figure 1, Hence, sequential information is required for bacteria division detection. Furthermore, the formation of the cell wall is undetected if a bacteria undergoing division migrates over other bacteria nor when it is oriented perpendicular to the xy-axis as shown in Figure 2.

In this paper, we apply a two stage approach to detection bacteria division. Firstly, the bacteria boundaries are segmented using level sets and graph cuts. The position, velocity and length of each bacteria were tracked. Next, we train a hidden Conditional Random Field (HCRF) to detect division in tracked candidate sequences. Instead of labelling each

[1]L.L.S.Ong, B. Kundukad, P. Doyle and H.H. Asada are with Singapore-MIT Alliance for Research and Technology, Singapore. sharon.ong, b.kundukad at smart.mit.edu

[2]X. Zhang is with Wuhan University, China. zhxh at whu.edu.cn

[3]J. Dauwels is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. jdauwels at ntu.edu.sg

[4]P. Doyle and H.H. Asada are also with MIT, Cambridge, MA, USA. pdoyle, asada at mit.edu

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

frame independently and smoothing the result, a sequence of frames are jointly labelled.

Related work to detect division events have only been applied to eukaryotic cells. Shape, size and texture features of a eukaryotic cell dividing has been classified using nearest neighbour classifiers or support vector machines [3]. El-Labben *et al.* [6] applied Hidden Markov Models to detect cell division using fluorescence microscopy images of eukaryotic cell nuclei. In Liu *et. al* [8] applied Hidden Conditional Random Fields to detect eukaryotic cell division from phase contrast sequences. When a eukaryotic cell divides, spindles form in the cell and its shape changes significantly. However, there is a lack of distinctive features in bacteria division. Vallaton *et al.* [13] recognized bacteria division by a small intensity gradient which represents the new cell wall formation. This technique may fail when two bacteria are close to each other. There is no known related work which classifies cell division in bacteria using temporal models.

II. Methods

Biofilms were grown in continuous-culture flow cells. The initial stages of *Pseudomonas aeruginosa* bacteria culture were captured with a Olympus IX73 microscope at $50\times$ magnification. In these bright-field images, acquired at 1 minute intervals, bacteria appear as ellipse shaped objets with bright intensity surround by a ring of dark intensity. An example is shown in Figure 3(a).

A. Bacteria Segmentation and Tracking

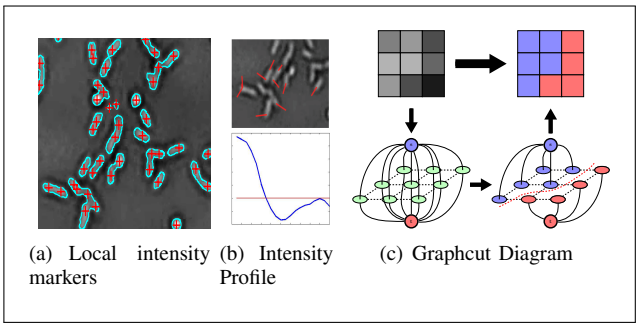
Vallaton *et al.* [13] applied edge detection to determine cell boundaries. The skeleton of connected regions were obtained to produce medial axes to determine bacteria boundaries. Xie *et al.* [14] applied active contours and an adaptive kernel-based technique to track non-proliferating bacteria. Both these related work were designed for phase contrast images. In bright field images, the intensity gradient between two bacteria is reduced when they are close to each other, resulting in segmentation errors as previously explained.

We address this issue by initially extracting the bacteria boundaries via level sets [2]. Bacteria which are close to each other will be extracted as a single cluster. Prior to cell division, each bacteria is recognized by a region of higher intensity at its center. Bacteria division is detected by an increase intensity gradient around its center, caused by the cell constricting. Hence, there will be two separate regions of higher intensity in a dividing cell.

The red $+$ in Figure 3(a) indicate the seed points, the maximum intensity regions inside the bacteria cluster. We create a reference pattern from intensity profile between the seed points to the background [5]. The red lines in top image of Figure 3(b) shows some examples of these intensity profile lines. The intensity profile of these lines are aligned and its reference pattern is shown in the lower image of Figure 3(b).

We denote X as the discrete intensity profile over a line segment of size N. At every signal lag m, the cross correlation image is

$$CC_{XM}(m) = \begin{pmatrix} \sum_{n=0}^{N-|m|-1} A, m > 0 \\ CC_{MX}(-m), m < 0 \end{pmatrix} \quad (1)$$

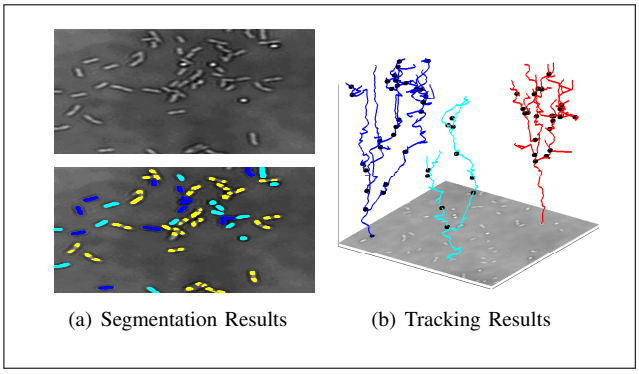


(a) Local intensity markers (b) Intensity Profile (c) Graphcut Diagram

Fig. 3. (a) shows a example bright field image. The cell boundaries (in cyan) were extracted with level sets [2]. The maximum regional intensity in each cluster is marked by the red crosses. In (b), the top figure shows example lines used to compute a reference intensity pattern shown in the bottom figure. (c) shows a graph-cuts diagram, where s is the source node (in blue) and t is the sink node (in red). The goal is to find the best "cut" or boundary, shown by the dashed red line.

where $A = ((X(n+m) - E(X))(M(n) - E(M)))$ and M is the discrete sequence of the reference pattern.

From a cropped image centered around each seed, we find its optimal boundary with a constrained graph-cut framework combined with the cross-correlation image [5].



(a) Segmentation Results (b) Tracking Results

Fig. 4. (a) Unconnected bacteria with one regional maxima shown in cyan. Early stages of bacteria division (blue) and separated bacteria (yellow) segmented from graphcuts. (b) Trajectories and lineages of three bacteria over space and time. Black ellipses indicated division events from tracking.

For seeds placed within the same cells, the segmentation results could be identical due to a smaller intensity gradient. Hence, two or more seeds with high degrees of overlap are merged and considered as a single cell. Figure 4(a) shows the segmentation results. The yellow labelled bacteria are the ones with overlap below a threshold. The blue have a high degree of overlap and are grouped as a single bacteria, and the cyan are single bacteria with one regional maxima. The segmented bacteria are tracked using Kalman Filtering. Figure 4(b) shows the trajectories of three bacteria families shown over space (x,y axis) and time (z-axis). The black ellipse detected division events from tracking. We refer the reader to our work on Kalman filtering in [9] for more details.

B. Hidden Conditional Random Fields

Figure 1 shows two tracked sequences where division was detected using methods in II-A. The top sequence is a division sequence but the bottom one is a incorrect

detection. By applying a classifier on the tracked results, we are able to detect the correct/incorrect labelled sequences. Given features from tracked (observed) sequence $\mathbf{X}$, we can compute the best corresponding label y^* by

$$y^* = \arg\max_y p(y|\mathbf{X}, \theta^*), \tag{2}$$

where y is a "division" or "no division" label. The conditional distribution $p(y|\mathbf{X})$ can be modelled in a probabilistic framework with Conditional Random Fields (CRFs) [7]. We specify the conditional dependency of each label on the sequence with an arbitrary number of feature functions. These feature functions may access the input sequence at any instance using the sequence. With CRFs, each frame will be assigned a label as the transitions between images is directly modelled. In our images, bacteria fission (division) lack distinctive cell features. A Hidden-state CRF (HCRF) predicts a label y to a given observations sequence $\mathbf{X} = x_1, x_2, \ldots, x_T$, where y is a member of a set $\mathbf{Y}$ for all possible labels [10], which is defined as:

$$p(y|\mathbf{X}, \theta) = \sum_h p(y, \mathbf{h}|\mathbf{X}, \theta) = \frac{\sum_h e^{\psi(y,\mathbf{h},\mathbf{X},\theta)}}{\sum_{y',h} e^{\psi(y,\mathbf{h},\mathbf{X},\theta)}}. \tag{3}$$

Here $\mathbf{h}$ is a vector of hidden variables globally conditioned on $\mathbf{X}$ and $\psi(y, \mathbf{h}, \mathbf{X}, \theta)$ is potential function, parameterized by θ as

$$\psi(y, \mathbf{h}, \mathbf{X}, \theta) = \sum_{j=1}^{m} \sum_{l \in L_1} f_{1,l}(i, y, h_j, \mathbf{X})\theta_{1,l} + \tag{4}$$

$$\sum_{(j,k)\in E} \sum_{l \in L_2} f_{2,l}(j, k, y, h_j, h_k, \mathbf{X})\theta_{2,l}. \tag{5}$$

L_1 is the set of node features and L_2 is the set of edge features. $f_{1,l}, f_{2,l}$ are the functions defining the features in the model and $\theta_{1,l}, \theta_{2,l}$ are the components of θ, corresponding to node and edge parameters. The feature function f_1 depends on a single hidden variable value in the model, while f_2 can depend on a pair of values. We learn the model parameters from training data by optimizing the objective function [10]

$$L(\theta) = \sum_{i=1}^{m} \log p(y_i|\mathbf{X}_i, \theta) - \frac{1}{2\sigma^2}||\theta||^2, \tag{6}$$

where m is the total number of training sequences, in which the first term is the data log-likelihood and the second term is the log of a Gaussian prior with variance σ^2. To search for the optimal parameter $\theta^* = \arg\max_\theta L(\theta)$, we use the gradient ascent algorithm.

III. Results

Figure 4(b) shows the trajectories of three bacteria tracked. We acquire the image frames to train a HCRF from the tracked trajectories. We acquired 30 sequences of cells division and 30 sequences of cells which are not dividing. These image frames vary between 7 and 30 frames long, where a potion of these are randomly selected as a training set. The remainder is used as a test set. To ensure that there are no biases, the classification of a random selection of training samples is repeated 50 times.

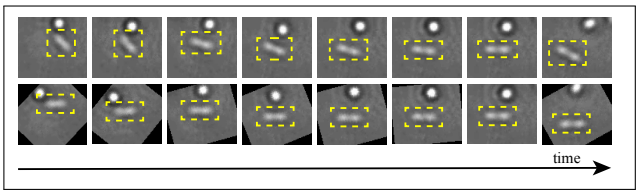


Fig. 5. Top row: Tracked bounding boxes of bacteria over time. Bottom row: Aligned sequence of the two row.

Number of Sequences Trained	HCRF	HMM
10 sequences	0.771 ± 0.124	0.755 ± 0.112
20 sequences	0.854 ± 0.110	0.802 ± 0.083
30 sequences	0.869 ± 0.095	0.825 ± 0.065
40 sequences	0.887 ± 0.099	0.834 ± 0.057
50 sequences	0.903 ± 0.136	0.839 ± 0.062

TABLE I

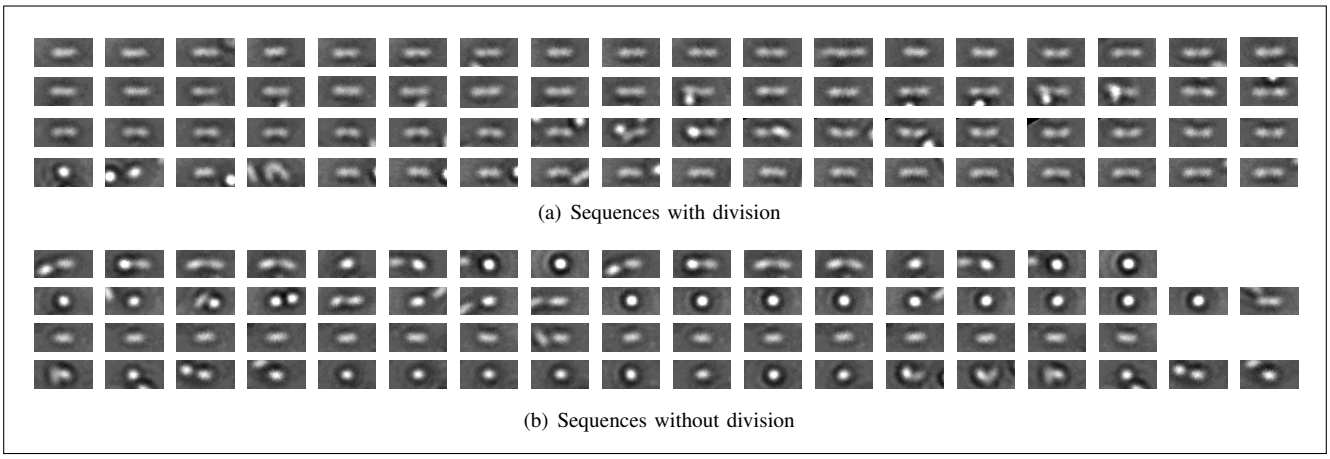
Sequence classification accuracy with HMMs and HCRFs.

The goal is the classify sequences to identify cell division among the tracked sequences. Bacteria constriction during division is detected by an increase intensity gradient around its center. However, bacteria in a tracked sequence could be oriented at different angles in the $2D$ frame. To build a HCRF model to classify this effect, we create a rotation invariant dataset by aligning each bacteria to the x-axis. The bottom row of images in Figure 5 shows the aligned sequence of the top row. The aligned images are cropped to a size of 20 pixels by 10 pixels to create a dataset. Figure 6(a) shows some examples in the dataset labelled as "division" and Figure 6(b) show image sequences labelled otherwise.

To create a feature vector to represent the image in a reduced dimension, we select a Histogram of Oriented Gradients [4] with a cell size of 4×4. Histogram of Orientated Gradients (HOGS) counts occurrences of gradient orientation, allow us to capture shapes and partial curves through gradient profile.

We compare the performance of HCRFs to classify our sequences with Hidden Markov Models (HMMs) [11]. A HMM jointly models set of hidden states and possible observations with a state transitional probability distribution and an observation probability distribution. Our results in Table I shows that HCRFs outperforms HMMs. A classification accuracy of 90% is achieved when trained with 50 image sequences using HCRFs.

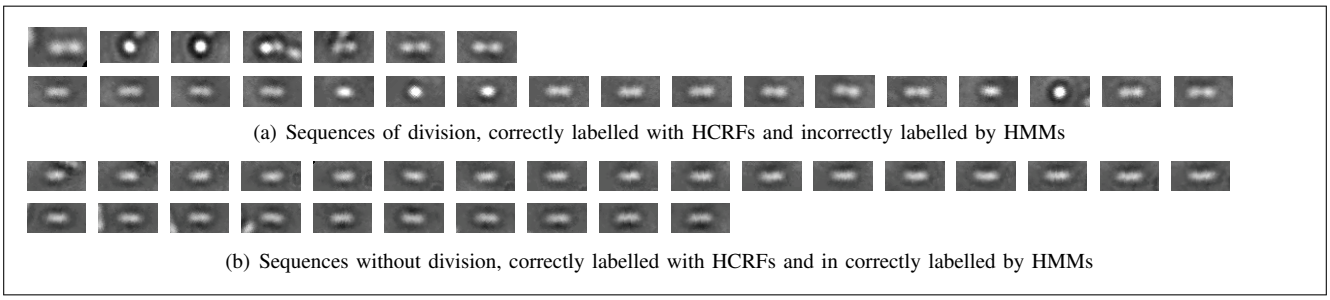
Figures 7(a) and 7(b) show sequences which were labelled correctly with HCRFs and incorrectly using HMMs. In Figure 7(a), the top sequence shows that another bacteria migrated over a proliferating bacteria (Frames $2-4$). The bottom sequence shows a dividing bacteria orientated itself perpendicular to the image plane between Frames $5-7$ and Frame 15. In these frames, the region of low intensity in the cell indicative of cell division is unclear. These sequences in 7(b) show bacteria in the growth stage, a stage prior to division. There are variations in intensity due to image noise resulting in some regions of low intensity in the cell.



(a) Sequences with division

(b) Sequences without division

Fig. 6. Top figure: The each row shows an example in the "division" labelled dataset. Bottom figure: Aligned image sequences labeled as "not division".



(a) Sequences of division, correctly labelled with HCRFs and incorrectly labelled by HMMs

(b) Sequences without division, correctly labelled with HCRFs and in correctly labelled by HMMs

Fig. 7. Example sequences successfully classified with HCRFs and incorrectly classified by HMMs.

These image sequences indicate that there is a long-range dependency among the observations. In HMMs, observations are assumed to be independent given the values of hidden labels, whereas HCRFs model this dependency. Hence, there is a higher likelihood that a sequence is incorrectly labelled using HMMs compared to HCRFs.

IV. CONCLUSIONS

We presented an approach to detect bacteria division with a two stage approach. After tracking each individual bacteria, we train a Hidden Conditional Random Field (HCRF) model to detect sequences with division. Future work includes exploring different feature descriptors and training a classifier to detect different stages of bacteria growth.

REFERENCES

[1] B. Carlson, *Principals of regenerative biology*, B. Carlson, Ed. Elsevier Academic Press, 2007.

[2] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277., 2001.

[3] X. Chen, X. Zhou, and S. Wong, "Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 762–766, April 2006.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 886–893.

[5] S. Dimopoulos, C. Mayer, F. Rudolf, and J. Stelling, "Accurate cell segmentation in microscopy images using membrane patterns." *Bioinformatics*, vol. 30, pp. 2644–51, 2014.

[6] A. El-Labban, A. Zisserman, Y. Toyoda, A. Bird, and A. Hyman, "Temporal models for mitotic phase labelling," *Medical Image Analysis*, vol. 18, no. 7, pp. 977–88, Oct 2014.

[7] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. IEEE Int. Conf. Machine Learning*, 2001, pp. 282–289.

[8] A.-A. Liu, K. Li, and T. Kanade, "A semi-markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations," *IEEE Trans. Med. Imaging*, vol. 31, no. 2, pp. 359–369, 2012.

[9] L.-L. S. Ong, M. H. A. Jr., and H. H. Asada, "Tracking of cell population from time lapse and end point confocal microscopy images with multiple hypothesis kalman smoothing filters," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2010)*, 2010, pp. 71–78.

[10] A. Quattoni, S. Wang, L.-S. Morency, M. Collins, and T. Darrell, "Hidden conditional random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 10, pp. 1848–1852, October 2007.

[11] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proc. of the IEEE*, 1989, pp. 257–286.

[12] P. Stoodley, J. Boyle, D. de Beer, and H. Lappin-Scott, "Evolving perspectives of biofilm structure," *Biofouling*, vol. 14, no. 1, pp. 75–90, 1999.

[13] P. Vallotton, C. Sun, D. Wang, L. Turnbull, C. Whitchurch, and P. Ranganathan, "Segmentation and tracking individual pseudomonas aeruginosa bacteria in dense populations of motile cells," in *24th International Conference Image and Vision Computing New Zealand (IVCNZ 2009)*, Sydney, NSW, Dec 2009, pp. 82 –86.

[14] J. Xie, S. Khan, and M. Shah, "Automatic tracking of escherichia coli bacteria," in *IEEE Transactions on Biomedical Engineering*, vol. 11, 2008, pp. 824–832.

ACKNOWLEDGMENT

This research was supported by the National Research Foundation Singapore through the Singapore MIT Alliance for Research and Technology's BioSystems and Micromechanics Inter-Disciplinary Research programme.