

Image Processing of Wild Type and Mutant Bacterial Cells

Sun-Hae Hong and Meng How Tan

ABSTRACT

Different types of bacteria have different cell shapes. In particular, *Caulobacter crescentus* is a crescent-shaped bacterium that undergoes asymmetric division to give a motile swarmer cell and a sessile stalked cell. However, when some genes are mutated, its morphology may be affected in a variety of ways. For example, the bacterium may grow filamentous or become rod-shaped. Our goal was to write a program that could decide whether a given image shows wild type or mutant cells. We used the discrete cosine transformation (DCT) and its principal components as the feature space to differentiate between wild type cells and mutant cells in general. We found that the projection vectors of abnormally shaped cells to those of wild type cells had different distributions. We also used locally weighted linear regression to successfully classify wild type crescent-shaped cells and mutant lemon-shaped cells. Future work includes extending the program to automatically identify protein mislocalization mutants from both phase contrast and fluorescence images.

INTRODUCTION

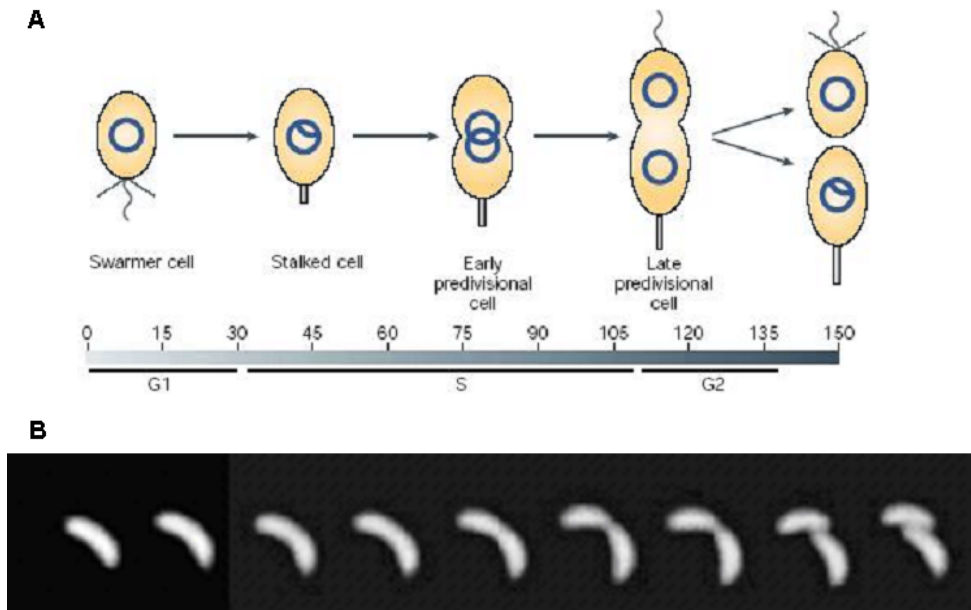


Figure 1. (A) A schematic of the different stages of the *Caulobacter* cell cycle (adapted from Skerker and Laub, 2004). (B) The cell's appearance under the microscope over the course of the cell cycle.

Caulobacter crescentus is an aquatic, non-pathogenic bacterium that has a distinct crescent shape. It serves as a model organism for studying both the prokaryotic cell cycle and how bacteria establish and maintain cellular asymmetry. As it progresses through the cell cycle, it elongates and pinches off in the middle to give two daughter cells - a smaller motile swarmer cell and a larger sessile stalked cell (Figure 1A). Since the bacterium

looks different under the microscope at different stage of the cell cycle (Figure 1B), we can use its morphology as a means for us to track its life progression.

We can use morphology to determine if the cells are normal or abnormal. Wild type *Caulobacter* cells have a stereotypical crescent shape and size (Figure 2A). In some mutants, cell division cannot occur. Instead, they grow filamentous and look abnormally long (Figure 2B). In other cases, the deletion of certain genes or the mislocalization of certain proteins causes polar defects (Figure 2C). Furthermore, when intracellular structural proteins are disrupted, the cells can lose its crescent shape and look like lemons (Figure 2D). Hence, we aimed to write a program, based on machine learning, that could take in phase contrast images and tell us if the image contains wild type or mutant cells.

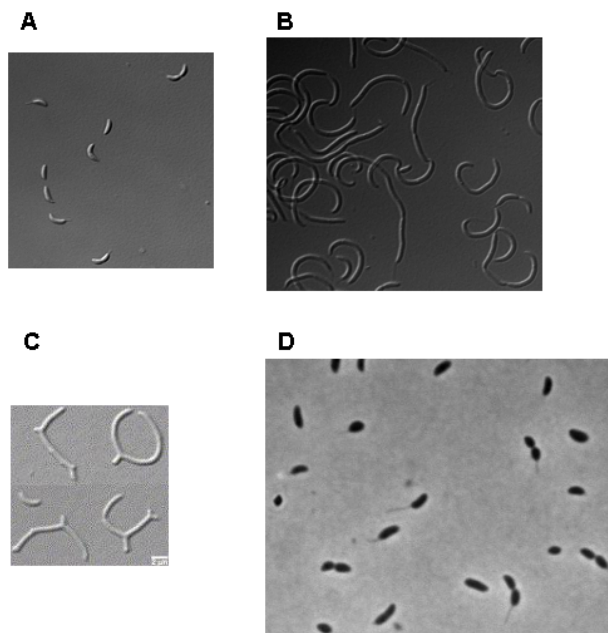


Figure 2. Cell morphology.

RESULTS AND DISCUSSION

Pre-processing of Images



Figure 3. How we obtain individual cells from a phase contrast image.

We pre-processed phase contrast images, each of which was 16-bit and of size 1000 by 1000, to obtain segmented cell images (Figure 3). We used a Matlab built-in function called `weiner2` for noise reduction. (We also tried the Canny edge detector, but it did not work very well on the input images.) Each image was segmented by thresholding. Cells were copied to individual boxes and oriented, so that the crescent side of each cell faced the right hand side. The brightness and contrast level were adjusted, so that the maximum brightness was 1 and the background level was 0.

Classification by principal components

We first tried to use Zernike moments to represent the segmented cell images. Zernike first proposed the Zernike polynomials in 1934. The moment formulation appeared to be very popular for image analysis and was able to outperform several alternatives in terms of noise resilience, information redundancy, and reconstruction capability (Teh and Chin, 1988.). However, we found that computation of the orthogonal Zernike moments was very slow.

Instead, we used the 2D DCT to represent our segmented cell images. Like any Fourier-related transform, DCTs express a function or a signal in terms of a sum of sinusoids with different frequencies and amplitudes (Ahmed et al., 1974). The DCT is often used in signal and image processing, especially for lossy data compression, because it has a strong “energy compaction” property: most of the signal information tends to be concentrated in a few low-frequency components of the DCT. Indeed, we discarded all the higher frequencies from the DCT to reduce the original matrix of size 101 by 101 to a 50 by 50 matrix (the error from discarding the higher frequencies was found to be insignificant). Then we calculated the principal components (PCs) of the feature vector. Figure 4 shows the first three PCs. The first PC seemed to be able to capture progression along the cell cycle fairly well.

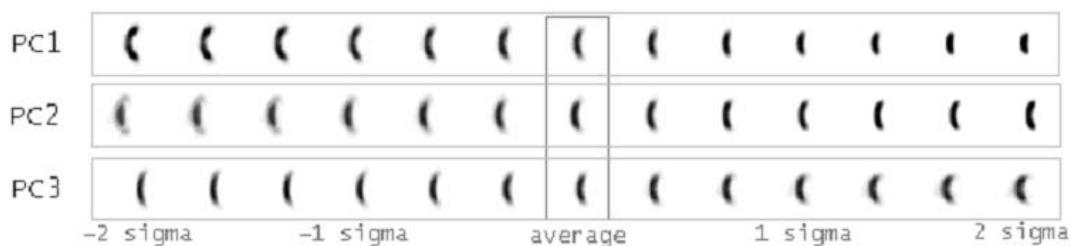


Figure 4. The first three PCs of wild type cell images.

We projected the feature vectors of both wild type and mutant cells to the first few principal vectors of wild type cells. The scatter plots of those projection vectors were different between wild type and mutant cells. For example, the distribution of $\Delta mreC$ cells was clearly distinguishable from that of wild type cells (Figure 5A). Cells expressing PleC-GFP were usually shorter than wild type, thereby resulting in a larger first PC (Figure 5B).

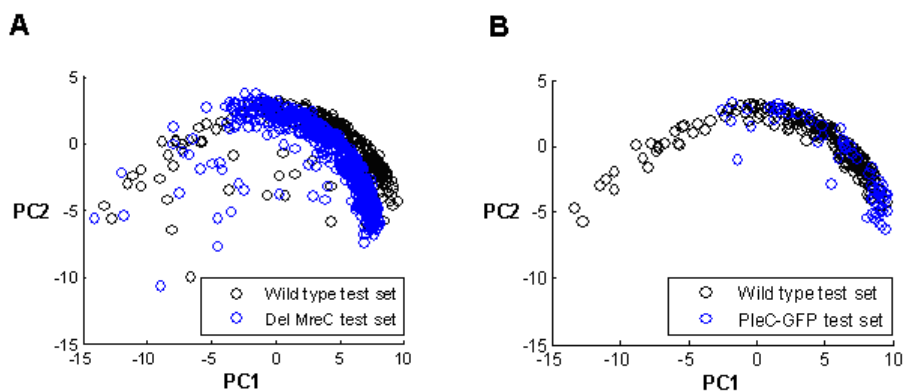


Figure 5. Scatter plots of the first two PCs for both wild type and mutant cells.

However, we were unable to use these principal components to classify wild type and mutant cells successfully. Although the first few PCs represented the cell shape of wild type cells quite well, they did not represent different types of mutant cells adequately, since the calculated PCs were optimized for expressing the cell shape of wild type cells. Figure 6 shows that the residual errors of the mutant test set are much bigger than those of the wild type test set.

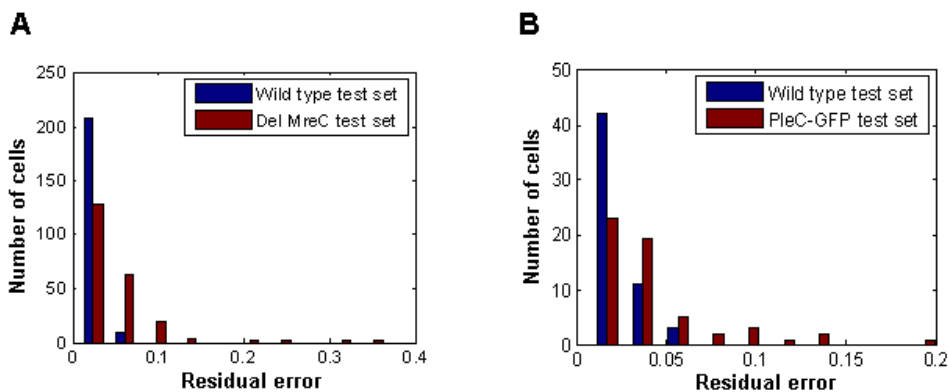


Figure 6. Histogram of residual errors.

Classification by Locally Weighted Linear Regression

We decided to focus on classifying wild type crescent-shaped cells and mutant lemon-shaped cells. Clearly, the width distribution along the cell's long axis was different between the wild type case and the mutant case (Figure 7). This formed the basis of our feature vector, which consisted of width measurements at 20 positions along the long axis. We also divided cells into two types – those that had started to pinch (swarmers and stalked cells) and those that were pinching (pre-divisional cells).

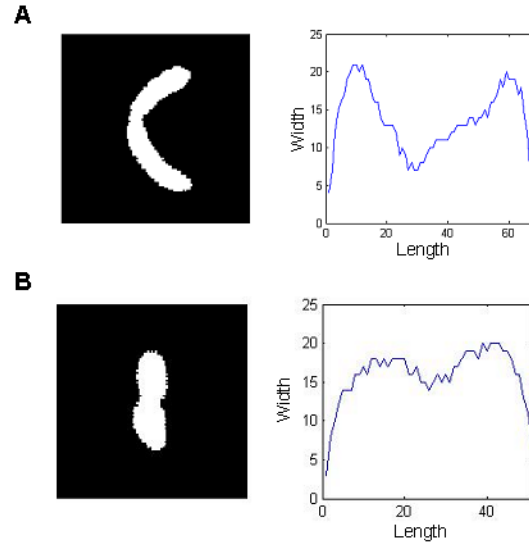


Figure 7. How the cell width varies along the long axis for both wild type and mutant cells.

We used locally weighted linear regression, so that the correct choice of features became less critical. Specifically, we wanted to minimize

$$\sum_i w^{(i)} (y^{(i)} - \theta^T x^{(i)})^2, \quad (1)$$

where the $w^{(i)}$'s were non-negative valued weights. We chose them to be

$$w^{(i)} = \exp\left(-\frac{(x^{(i)} - x)^2}{2\tau^2}\right), \quad (2)$$

where τ was the bandwidth parameter.

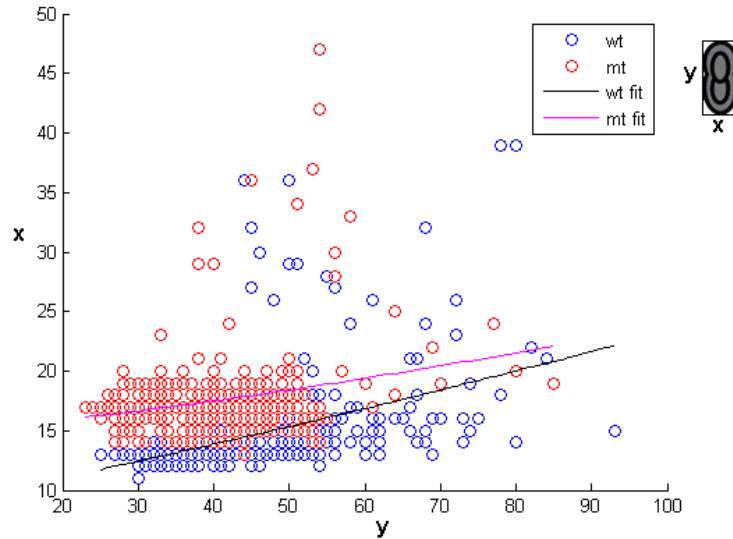


Figure 8. Size distributions of crescent-shaped and lemon-shaped cells.

We performed hold-out cross-validation. Our training set consisted of 402 wild type cells and 399 mutant cells. The results are given in Table 1, which shows that the program can differentiate between wild type and mutant cells. Furthermore, Figure 8 shows that crescent-shaped cells and lemon-shaped cells have different fits.

	Wild type cells	Mutant cells
Classified as wild type	235	76
Classified as mutant	7	233

FUTURE WORK

This research is still work-in-progress. Firstly, we want the program to be able to successfully classify other types morphology mutants, as shown in Figure 2. Secondly, we want to extend the program, so that it can automatically identify protein mislocalization mutants from both phase contrast and fluorescence images. A bacterial cell, although small, has well-defined compartments (for example midcell or the poles). Over the course of the cell cycle, different proteins localize to different positions in the cell (Figure 9). If we fluorescently tag the proteins, we can follow their localization patterns over the course of the cell cycle. We can then compare the pattern seen in wild type cells with the pattern in mutant cells.

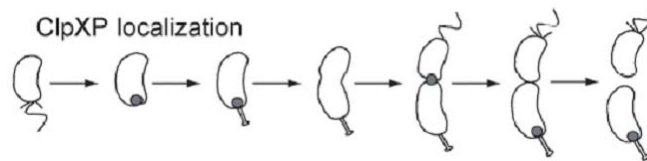


Figure 9. Schematic of the cell cycle localization pattern of ClpXP, a bacterial protease (adapted from Iniesta et al., 2006).

REFERENCES

1. Ahmed N., T. Natarajan, and K.R. Rao. 1974. Discrete cosine transform. *IEEE Trans. Computers* **C-23**: 90-93.
2. Iniesta A.A., P.T. McGrath, A. Reisenauer, H.H. McAdams, and L. Shapiro. 2006. A phosphor-signaling pathway controls the localization and activity of a protease complex critical for bacterial cell cycle progression. *Proc. Natl. Acad. Sci.* **103**: 10935-10940.
3. Skerker J.M. and M.T. Laub. 2004. Cell-cycle progression and the generation of asymmetry in *Caulobacter Crescentus*. *Nature Rev. Microbiol.* **2**: 325-337.
4. Teh C. and R.T. Chin. 1988. On image analysis by the method of moments. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **10**: 496-513.
5. Zernike F. 1934. Beugungstheorie des Schneidensverfahrens und seiner verbesserten Form, der Phasenkontrastmethode (Diffraction theory of the cut procedure and its improved form, the phase contrast method). *Physica* **1**: 689-704.