An image analysis tool in single molecule microscopy: Bacterial plasmid copy number control as a case study.

Author: Mohammad Intakhar Ahmad

Name of the institute: Institute of Systems and Synthetic Biology (iSSB)

Supervisor: Alfonso Jaramillo, PhD.

CR1 CNRS (Head of Synth-Bio team)

Genopole Campus 1, Bât. Genavenir 6

5 rue Henri Desbruères

Fax: +33(0)1 69 47 44 37

Phone: +33(0)1 69 47 44 44

F-91030 ÉVRY CEDEX, FRANCE

E-mail: alfonso.jaramillo@issb.genopole.fr

Master 2 in Systems and Synthetic Biology (mSSB)

Academic year: 2011-2012.

Abstract

Conventional bulk-scale molecular biology methods are incapable of tracking the copy number fluctuation of intracellular molecules. The potential impacts of such fluctuation have been demonstrated in a number of studies, ranging from cancer development to antibiotic resistance. On one hand, state-of-the-art microscopy techniques together with current developments on fluorescent tags have allowed us to probe single molecules in a living cell with millisecond time precision, and nanometer spatial resolution. On the other hand, with fast and improved imaging, the processing and analysis bottleneck have become even more challenging. The aim of this project is to develop an automated analysis tool for time-lapse fluorescence images. The tool then correlates the results with corresponding phase contrast images at pixel level accuracy to obtain quantitative data of single molecules from each individual cell in the image frames. Performance evaluation showed promising result on both the synthetic and real experimental data with 100% sensitivity and specificity until signal-to-noise-ratio (SNR) approaches 1.8. Preliminary dataset (n=20) generated by the plasmid copy number control experiment in *Escherichia coli* was analyzed using the tool we developed. The result suggests that plasmid copy number follows a linear function of cell volume, where the constant is 0.27.

Table of Contents

List of figures and schemes

List of Abbreviations

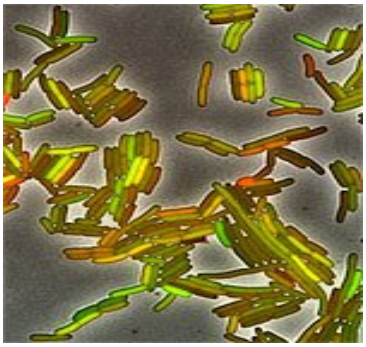
CNV	Copy number Variation
FPs	Fluorescent proteins
LacP	Lac Promotor
IA	Image Analysis
EM-CCD	Electron Multiplying Charge Coupled Device
RFP	Red florescent protein
YFP	Yellow fluorescent protein
IPTG	Isopropyl-β-D-thio-galactoside
GFP	Green fluorescent protein
NA	Numerical Aperture
CCD	Charge Coupled Device
SNR	Signal- to-noise-ratio
PDMS	Polydimethylsiloxane
TIFF	Tagged Image File Format (image file)
RMSE	Root Mean Square Error

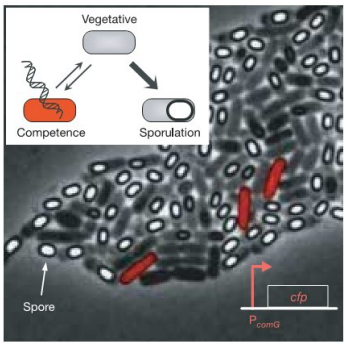
Chapter 1

Introduction

Biological processes deal with an intricate tradeoff between stability and variation. Stable heredity is important to all living organisms to pass deterministic traits to descendants. Even though, plasticity (room for fluctuation) fuels the adaptation strategy. Characteristic differences introduce individual variations which paves the way for natural selection. This is how; biological system heuristically guides variation towards an optimized output for better adaptation. Depending on their sources, variations or fluctuations are divided into two kinds, extrinsic noise and intrinsic noise(Swain, Elowitz, & Siggia, 2002). Internal noise has its origin inside the cell (e.g. probabilistic character of bio-chemical reactions, copy number fluctuation of molecular species). Extrinsic noise is generated by global variability (e.g. variation in temperature and pH). In our study, we focused on developing an analysis tool for the data generated from the experiment addressing the intrinsic noises like copy number variation (CNV) of plasmid in bacteria.

Random internal fluctuation of molecules inside the cell is almost not included in textbook biology. But in recent years, with the advancement of technologies, researchers have been exploring deeper inside the biological system and far more haphazard cellular features are being observed than expected. Questions therefore are being asked concerning observed noise. In the beginning of last decade, Michael Elowtiz and his colleagues experimentally showed that gene expression-flux varies significantly during transcription, even when, cells are genetically identical, growing in a homogeneous environment (Elowitz, Levine, Siggia, & Swain, 2002). Over the past decade, a great deal of data has been published related to biological noise at different levels in different biological systems.





A B

Figure 1.1: Experimental evidence of biological noise.

A. The two different fluorescent protein color variants were co-expressed in cells with identical promoters. Two-color reporter gene assay proves expression variation as an intrinsic noise (Elowitz et al., 2002). B. *Bacillus subtilis* micro-colony in nutrient-limited conditions illustrating random choice of developmental paths connecting the vegetative, spore forming and competent states (Süel, Garcia-Ojalvo, Liberman, & Elowitz, 2006).

At present, there is a well established biological value to study noise in a biological system. A good number of suggestions have published regarding how fluctuation arises and spreads(Paulsson & Elf, 2006). Those models are consistent with stochastic behavior of transcription, translation and other biochemical reactions. A dynamic range of molecular species or copy number is teaming inside a cell. In statistics, Square Root Law says fluctuation (noise) relies within a probability relative error of $1/\sqrt{n}$ (here, n represents sample size) (Schrödinger, 1992). Therefore, low copy number molecules in a cell are sensitive to natural stochasticity. The range of copy number variation within a cell is estimated to be about 1 to 10^6. But most of the molecular species are low copy number (less than 100 per cell) (Hachey & Chaurand, 2004). Many crucial biochemical process involve low copy number species, including cell signaling, regulation of gene expression, cell polarization, cytoskeleton dynamics etc (Huang et al., 2007).

Single-Molecule Experiments in Synthetic Biology are also getting more attention because synthetic biologists try to control gene expression in cells at the transcriptional level (Andrianantoandro, Basu, Karig, & Weiss, 2006). The possibility

to influence and control cell metabolism through modified synthetic genetic circuits offer fascinating prospects for molecular cell biology in the field of biomimetics and synthetic biology (Breslow, 1972). To be successful in these frameworks we need to know the extent to which a cell can have stocastic fluctuation of low copy of molecules. Then the knowledge about the origin of fluctuations can help us to handle the fluctuation.

To examine this feature experientially, we need very elegant technical solutions. We need to grow an isogenic population of cells in healthy conditions and in a homogeneous environment. A sophisticated system like a micro- fluidic device can help to supply cells with fresh medium continuously for many subsequent generations to be imaged with single molecule microscopy without extrinsic noise (Young et al., 2012).

Microscopy is one of the oldest techniques employed to study cells. Microscopes have gone a long way since its invention in the Netherlands in the early 1600's. Recent techniques can reach beyond the diffraction limit of light in terms of resolution (Huang, Babcock, & Zhuang, 2010). These days, modern microscopes are used in most studies of molecular, cell, and developmental biology. State-of-the-art microscopes and sensitive cameras paired with authoritative fluorescent probes allow for high-resolution real-time observation *in vivo* (Lichtman & Conchello, 2005). In fact, biological data have undergone a paradigm shift from largely non-quantitative representative observations in bulk-scale cells to high-throughput quantitative data in live individual cells (Xie, Choi, Li, Lee, & Lia, 2008).

Despite its cost and complexity, fluorescent imaging microscopy has been a vital tool for biologists, especially science the discovery of the green fluorescent protein and the possibility of tagging almost any protein with it (Chalfie & Kain, 1998). Probably the best indicator of the utility of GFP and GFP-like proteins is the 2008 chemistry Nobel Prize that was awarded to Profs. Shimomura, Chalfie and Tsien for the discovery and development of the green fluorescent protein, and the fact that in 2004 about 50%, 35%, 60% and 20% of the articles in the journals Cell, Development, Journal of Cell Biology, and Neuron mentioned or used GFP-like proteins (Lemay et al., 2008).

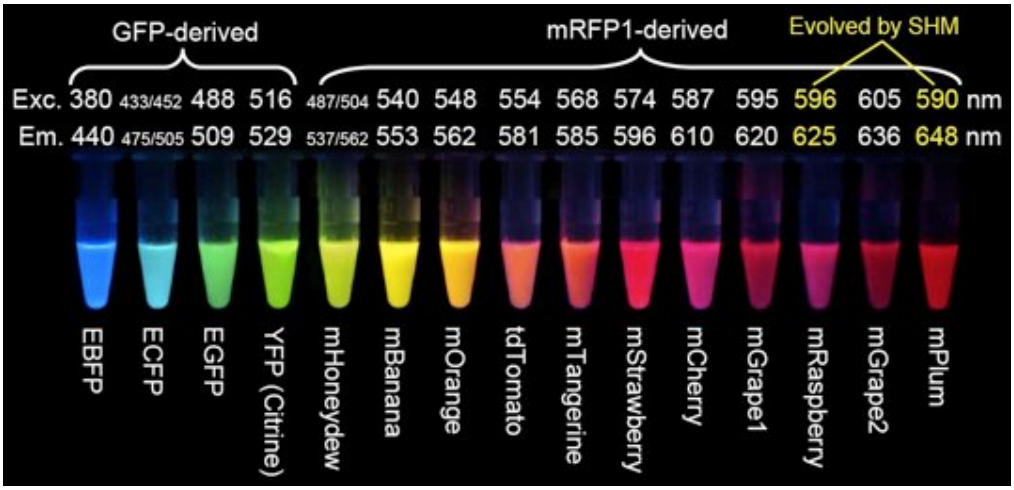


Figure 1.2: Color range of fluorescent proteins. GFP stands for Green fluorescent protein, RFP is for Red Fluorescent Protein, SHM for somatic hypermutation. E stands for enhanced versions of GFP, m are monomeric proteins, and td is a head-to-tail dimer (image from Roger Y. Tsien Nobel Lecture 2008.)

At the present time, a good number of biological laboratories are equipped with many techniques of modern microscopy. But, there are few standardized tools for the analysis of the images obtained (e.g. for cell segmentation, cell tracking and fluorescent signal analysis). Therefore, images are often analyzed by visual inspection alone, which is generally subjective and therefore engrosses the danger of erroneous conclusions. Visual inspection is further limited in the number of images analyzed. Therefore, computational image analysis is crucial in order to obtain consistent, statistically significant and reliable information from imaging, and thus deserves equal attention and effort as the image acquisition itself. In fact, image analysis is often equally time-consuming or even more tedious than the processes of sample preparation and imaging themselves.

Automated image analysis, which is highly needed for modern high-throughput studies in proteomics, functional genomics and drugs screening, is still a great challenge due to limitations in the acquisition process. The signal - to- noise ratio (SNR) is usually very low, as the illumination intensities are kept low during experiments to prevent photo bleaching and photo damage (Smal, Loog, Niessen, & Meijering, 2009).

All acquired images could be contaminated by noise from a variety of sources. Noise is astochastic phenomenon that cannot be compensated for as opposed to systematic distortions such as shading or some forms of image blur (Wilkinson & Schut, 1998). The noise sources that play a role in scientific CCD cameras are: photon noise, thermal noise (dark current and hot pixels), readout noise (i.e. amplifier noise, on-chip electronic noise), and quantization noise. Some of these noise sources can be made negligible by proper electronic design and careful operating conditions. One of them – photon noise – can never be eliminated and thus forms the limiting case when all other noise have become negligible compared to this one (Vliet, Sudar, & Young, n.d.).

The objects, labeled with fluorescent proteins, appear in the images as bright spots, each occupying only a few pixels. The identification of objects within an image can be a very difficult task. One way to simplify the problem is to change the grayscale image into a binary image, in which each pixel is restricted to a value of either 0 or 1. The techniques used on these binary images are known as morphological image processing. The foundation of morphological processing is in the mathematically rigorous field of set theory (Ritter & Wilson, 2000); however, that level of sophistication is seldom needed. Most morphological algorithms are simple logic operations. Each kind of images requires a custom solution developed by trial-and-error.

An array of tricks is used rather than standard algorithms and formal mathematical properties. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. By choosing the size and shape of the neighborhood, we can construct a morphological operation that is sensitive to specific shapes in the input image. There are mainly four kinds of operations frequently used in image processing. They are erosion, dilation, opening and closing. The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. Opening and closing can be done with their combination (Gil, Kimmel, & Member, 2002).

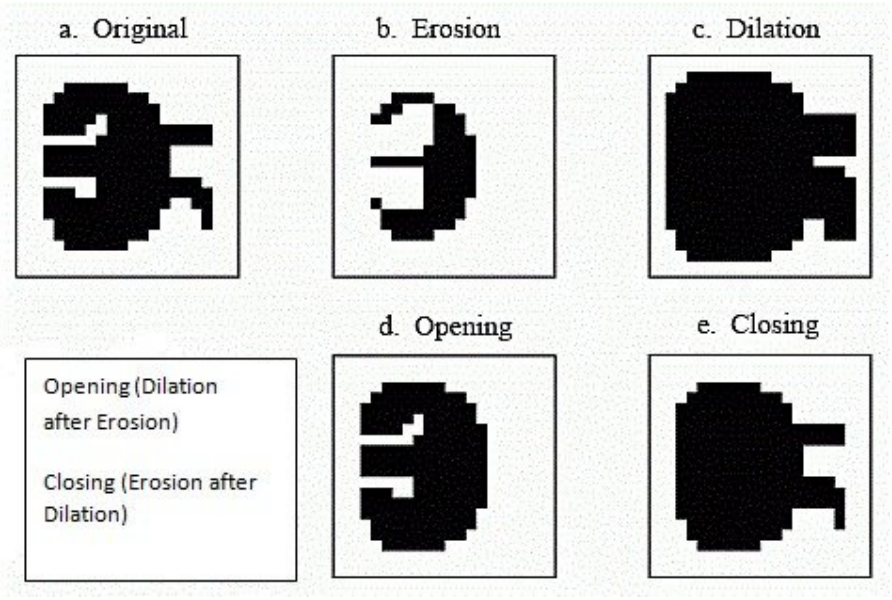


Figure 1.3: Major operations in image processing

Indeed, morphological image processing is an extremely powerful tool in image processing applications (e.g. filling holes, detect spots etc.). We can perform certain operations with applying structuring element (SE) on an image (Dougherty, 1992). For our purposes, we shall mainly be interested in three specific operations called morphological opening, closing and grayscale reconstruction by erosion.

Chapter 2

Materials and methods

2.1 Experimental model organism and fluorescent labeling of plasmids

We took *Escherichia coli* as our model organism**.** In our experimental setup, each bacterium measures roughly 1 µm in width and 5 µm in length. The generation time of *E. coli* varied between 30-40 minutes. We used an IncW family plasmid, R388. The plasmid lack *mob* region that is required for horizontal transfer of plasmid (Boyd, Archer, & Sherratt, 1989). This how, we stopped the horizontal transfer of plasmids. The plasmid had 30 binding sites for GFPs and the GFP expression was controlled by *lac* promoter (lacP) on the plasmid. So, the GFP production was induced by IPTG (manuscript in preparation).

2.2 Microfluidic experimental platform

In the experimental run time, the cells were uninterrupted by the fluctuation of physical parameters to avoid experimental noise. We kept them in a well defined physiological state and a stable chemical environment. To be able to do this we used microfluidic device. With that device we managed to keep precise control over the media and we manipulated fluid flow that was geometrically constrained to a very small scale. The chip was constructed with PDMS (Polydimethylsiloxane). The organic portion in PDMS foundation is the methyl group, which has one of the weakest intermolecular forces. The inorganic siloxane backbone is one of the most flexible polymer backbones available. It has almost no auto-fluorescence (Bélanger & Marois, 2001).

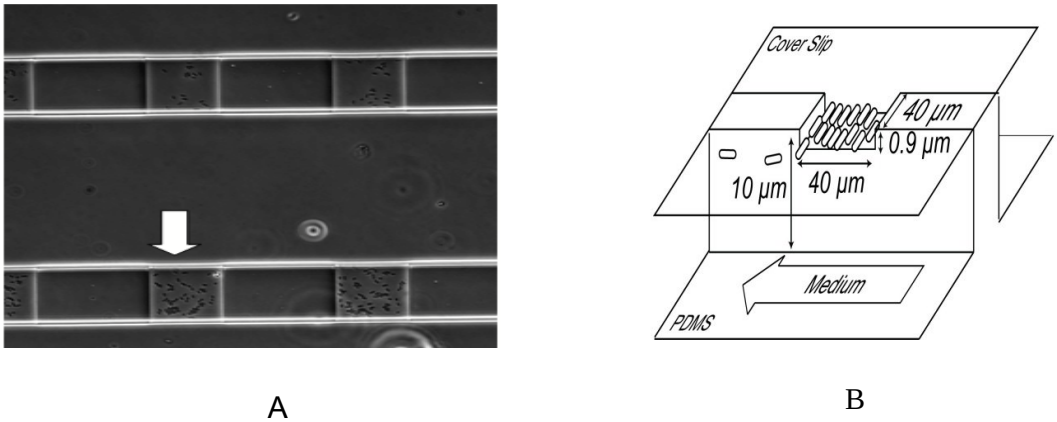


Figure 2.1: Illustration of the microfluidic platform

A.PDMS beds PDMS under phase contrast microscope (20x). B. Example of little traps of cells growing happily.

2.3 Single-molecule detection microscopy

We designed the plasmid such that a single plasmid is made visible in fluorescence imaging. We used epifluorescence microscopes, which means excitation and observation of the fluorescence are from above the specimen. We minimized photo bleaching by reducing the excitation light intensity and shortening the duration of illumination.

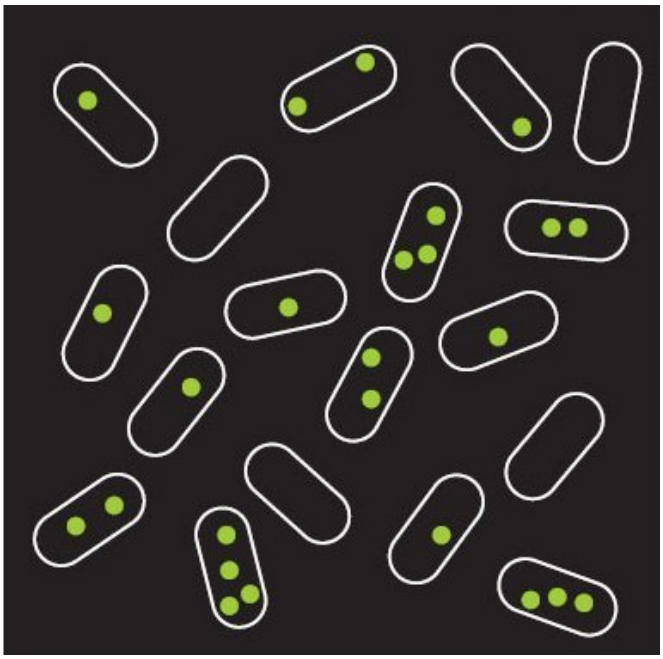


Figure 2.2: Single molecules with fluorescence microscope.

Single-molecule fluorescence microscopy using fluorescent protein tagged to protein of interest (a cartoon image,). White lines represent cell area and green dots represent fluorescent molecules.

For the particular experiment we used a DC powered mercury lamp (X-Cite® exacte) with 100 millisecond exposure time.

2.4 Optical set-up of the microscope

We employed an Axiovert 200 inverted fluorescent microscope from Zeiss. To capture image Hamamatsu EM-CCD Camera (C9100-02) was used. A schematic diagram of our optical setup is shown in **figure 2.3**. Approximately, 475 nm UV beams were hit to sample to excite. An excitation band-pass filter helped us to get only the peak of the excitation wavelength, and then there is a shutter to control the exposure time. A UV filter was there in front of the dichotic mirror. With the help of the dichroic mirror the beams were reflected upwards to the sample to excite the fluorophores. We have an epi-fluorescence oil immersion (100x) objective lens (numerical aperture, NA=1.4). Emitted wavelength was about 525nm which was passed through the long-pass dichotic mirror. Passed wavelength was re-filtered by an emission filter afterwards. Then the wavelength was focus on the EM-CCD (Electron Multiplying Charge Coupled Device) sensor. EM-CCD generated image of the signals (wavelength) and stored the data on a computer hard disk. We also take the phase contrast micrograph of the cells at 100x magnification. Definite Focus System (from Carl Zeiss) automatically maintained microscope on focus so that, the point of interest was always kept in sharp focus. For better resolution we used a 1.6x lens (NA 0.05) in fort of the EM-CCD.

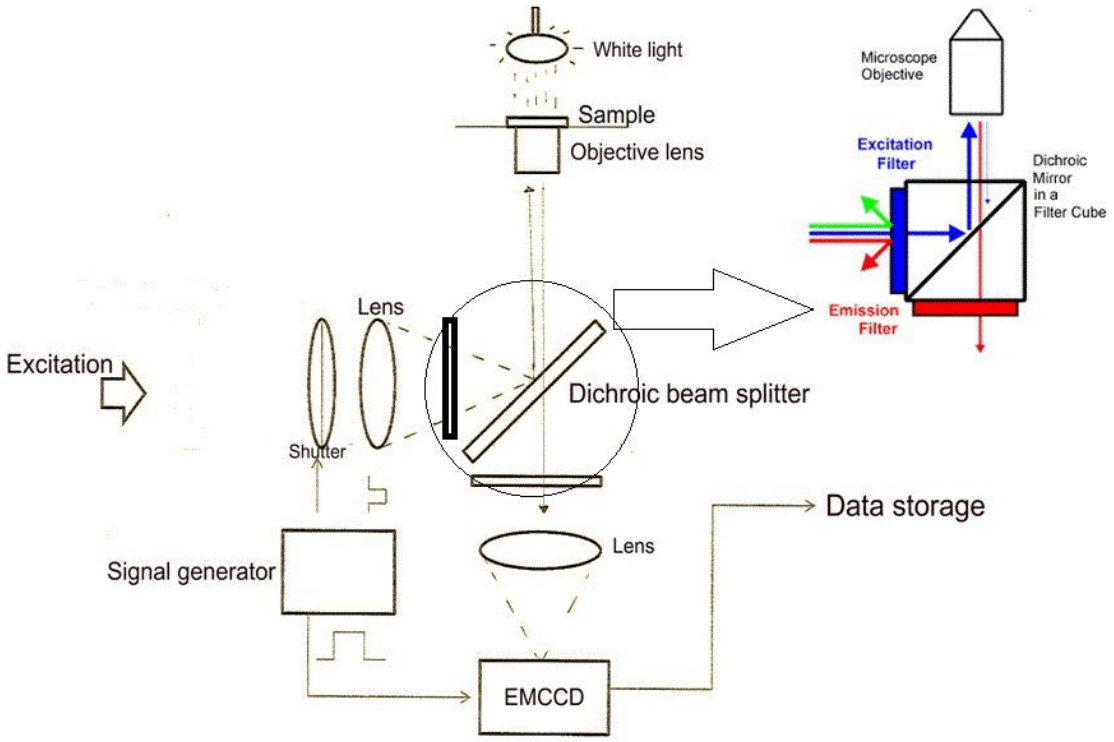
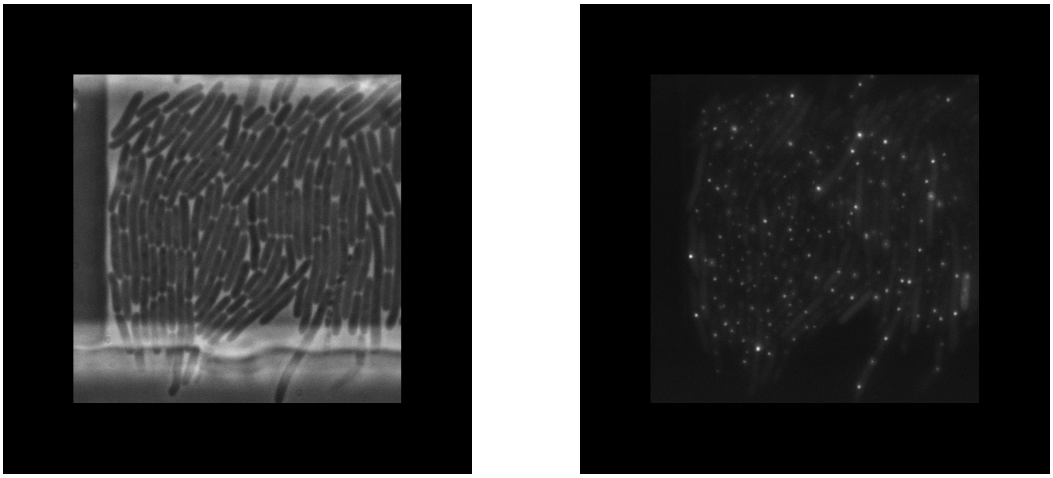


Figure 2.3: Schematic diagram of the optical setup

2.5 Data generation

Microscope Software AxioVision LE (Carl Zeiss) was in use to control the microscope. Two cameras, an 8-bit EM-CCD (1000x1000 resolution) and a CCD (1000x1000 resolution) for fluorescence and phase contrast images respectively were employed to record images. Images were stored as TIFF, a file format for bitmap -based images. Well controlled and well-timed image acquisitions were performed over the position-list of the selected traps in the bio-chip. Fourteen phase contrast images with 0.5 minute interval were taken and recorded with the CCD camera .Subsequently, one fluorescent image was taken after the first phase contrast and recorded with the EM-CCD camera (EM gain-150, exposure time 100 milliseconds).



A B

Figure 2.4: Sample images. A. Phase contrast image B. fluorescence image

A. Phase contrast taken with our optical system at 160x magnification. B. Fluorescent image at 160X magnification.

2.6 Spot detection on fluorescence images

We performed morphological top-hat algorithm implemented in a MATLAB script on the grayscale image to detect spots (the algorithm is available in the MATLAB image analysis tool box). Top-hat algorithm computes the morphological opening of the image and then subtracts the result from the original image. Top-hat uses the structuring element to determine the openings which is set by the user by setting a threshold value. We determined the minimum difference between the peaks of two separate dots by a mathematical formula, which gives us the highest resolving power of our optical system.

The resolution distance of a microscope is determined by the wavelength of the source of illumination (light or electrons) and the numerical aperture of the lens

used, such that:

$$R_D = \text{lambda}/2NA$$

R_D = resolution distance in nanometers, lambda = the wavelength of the source of illumination and NA = numerical aperture

We have two lens with 1.4 and 0.05 NA respectively and the wavelength in use was about 525nm. The R_D value of our optical system was measured in pixels, which is around 4 pixels (1pixel=50nm).We implemented a in-house method written in MATLAB to filter dots from false positives depending on this distance feature. The method fits 2D Gaussian (a built-in function in MATLAB) to area (7x7 pixel) around the detected spots by Tophat method and calculates the goodness of fit (Gof) for the fitted Gaussian.

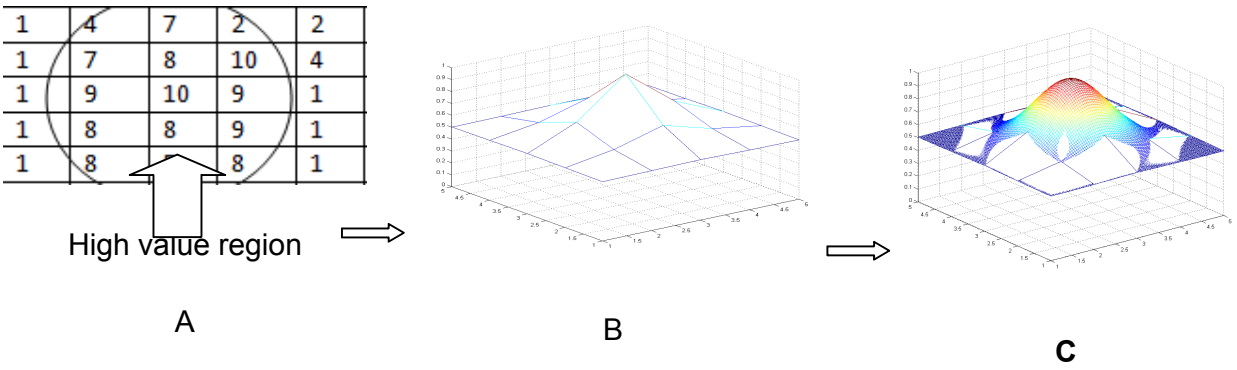


Figure 2.5: Schematic representation of Dot detection.

A. Dummy spot showing relative pixel values B. Surface plot of the values of pixels in a spot C. 2D Gaussian fitting to the spot

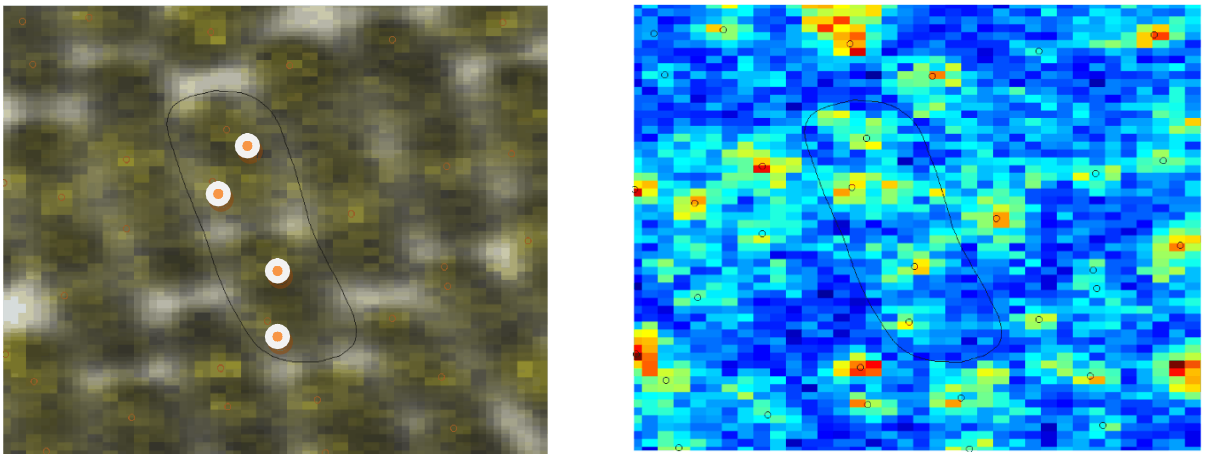
2.7 Cell segmentation of phase contrast images and correlate with the fluorescence images

In-house developed MATLAB scripts were used to segment and track the cells in the image frames. These scripts implement a few novel methods for automating cell segmentation and lineage reconstruction. Still it has some room to improve stability. Parameters for cell detection were determined by skilled scrutiny of cell features (shape and size). The fluorescent spots found in the fluorescence images are plotted

on the phase contrast images to assign their position in individual bacteria with pixel level precision.

2.8 counting of spots and result representation

We have an automatic pixel quadrants(x, y) compare script to correlate the position of the fluorescent singles in the phage contrast images. To be sure about the positions, we overlaid the spots found in the fluorescent image on the corresponding region of the phase contrast image and the cell contour found in phase contrast also overlaid on the fluorescent image.



A B

Figure 2.6: Overlay of Images (Phage contrast and corresponding fluorescence)

A. Fluorescent image is overlaid on the phase images and detected spots are marked B. A region of fluorescent image is marked (black line) which correspond a cell in the phase image and the detected spots are marked accordingly. False coloring is done according to pixel vale to have better contrast.

2.9 Measurement of signal to noise ratio (SNR)

Properly estimating the SNR is critical to image analysis application. Most often it is calculated incorrectly. Shot noise is proportional to the square root of the number of

photoelectrons. As a result, the noise level is higher in pixels containing a fluorescent object than in the background, even though the noise may be more visually apparent in the background. Failure to account for this relationship may lead to very large errors in estimated SNR. But there is a quite well known equation for SNR (Cheezum, Walker, & Guilford, 2001).

$$SNR = (I_o - I_b)/\sigma \ldots\ldots (1)$$

$$\sigma = \sqrt{[(\sigma b)^2 + (\sigma o)^2]} \ldots\ldots (2)$$

Here, mean intensity of object (I_o) and mean intensity of background (I_b),

Noise level (σ).

2.10 Algorithm development platform

A digital image can mathematically be defined as:

$$I = f(i,j), i = 1,2,\ldots,M, \quad j = 1,2,\ldots,N \ and \ I = 0,1,\ldots,2^k - 1$$

where (*i*,*j*) is the pixel coordinate of the image and *I* is the intensity levels of the image. For our experiment we used 8 bit images. A 8-bit grayscale image represents that the number of intensity levels that can be represented in the image is equal to 2^8=256 where a 0 equals to the color black, 255 is white and everything in between is different shades of gray. The plane {(i,j); i=1,2,…,M , j=1,2,..,N} is considered as the spatial plane of the image. An image containing the pixel values 1 or 0 is a binary image. MATLAB is especially good at developing algorithms and useful representation of a digital image in a matrix form. We took MATLAB as our platform of algorithm implementation.

Chapter 3

Results

3.1 Fluorescence signal detection algorithm

The first step: The more noise reduction is done the better detection we could have. But always there is tradeoff between noise reduction and signal loss. We used the structuring element (SE) with radius 3 pixels to perform image opening. This helped us to eliminate random bright spots smaller the structuring element.

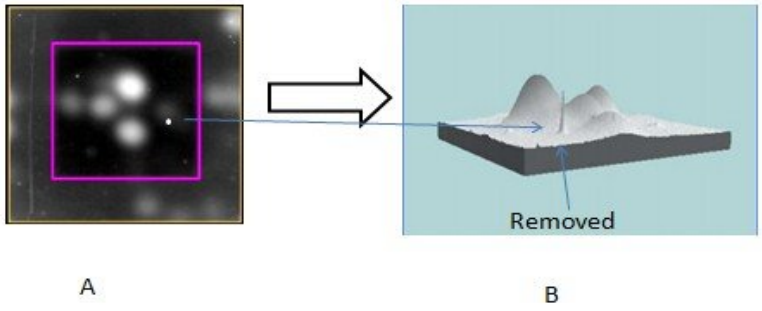
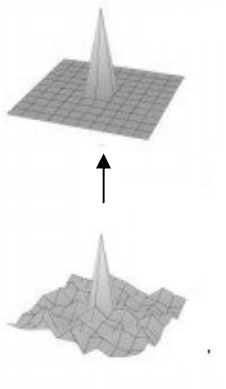


Figure 3.1: A. selected part of a fluorescence image B. Landscape feature of the selected part

Afterwards, we used a Mean filter for image smoothing. A disk size of 3 pixels was used for this average filter.



We used 3×3 averaging kernel for mean filtering. Since the shot noise pixel values which are very different from the surrounding values were removed using imopen (MATLAB command), there was no tend to significantly distort the pixel average calculated by the mean filter.

Figure 3.2 Average filtering of images

The second step: We need to make the initial guess of the spot position. We put a minimum difference of pixel value 10 from the background on the 8 bit gray scale picture (0 lowest and 255 highest) as our initial guess of the spot. We compared the central pixel value with its surround pixel values (5x5 pixels, average spot size) putting a threshold of 10. The program goes to each and every pixel and search for conditions declared in the scripts. This is how the first round of the search was done. The parameter values for the first round detection were selected by visual inspection on trial and error basis.

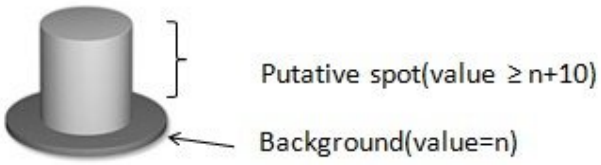


Figure 3.3: Putaive spot selection depending on Tophat method.

The third step: The preliminary selected spots are feed into a secondary sorting process that depends on the resolving power of our optical system. The theoretical resolving power of our system is about 4 pixels. We needed to eliminate spots closer than 4 pixels to another one. To do this we wanted to have a quality parameter for the discrimination. Our approach was to fit a 2D Gaussian function for each and every spots considering specific neighbor range (7x7 pixels) and took the goodness of fit in terms of RMSE (Root mean squared error).

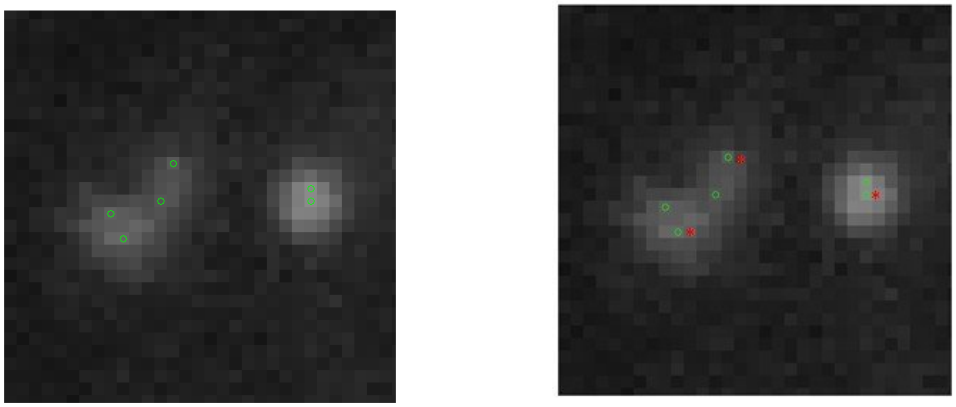
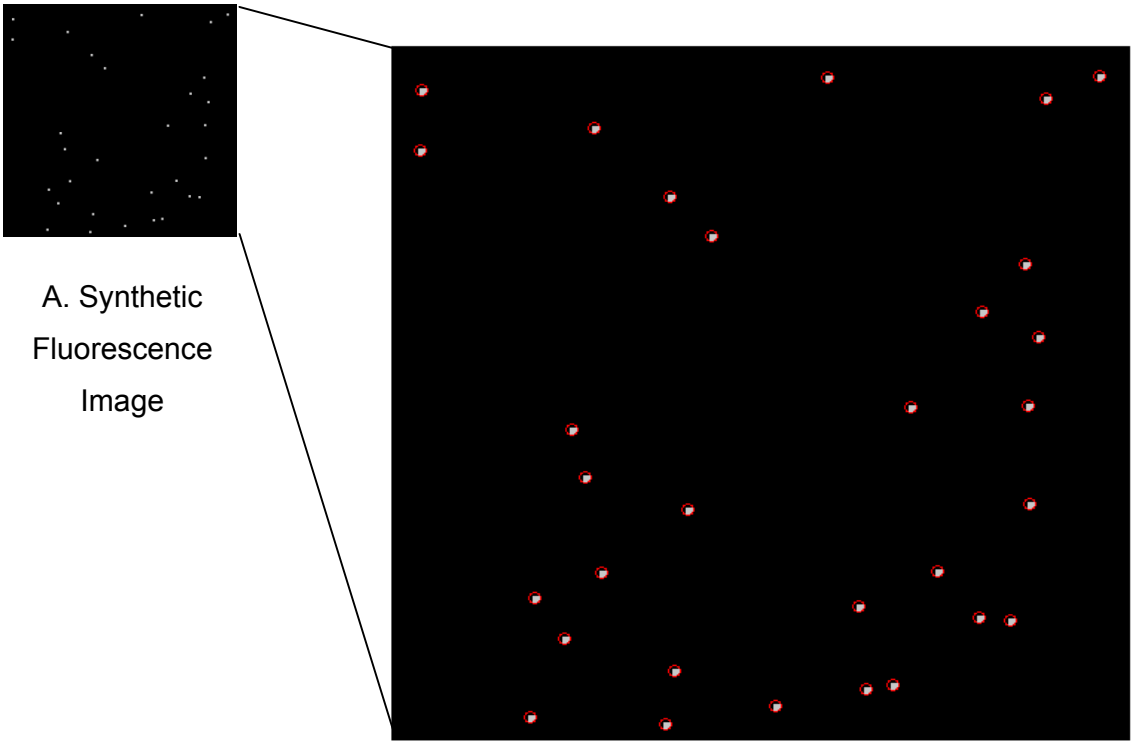


Figure 3.4: Left: First round selection of spots (green circle) Right: True spots after secondary sorting process (red star beside the green circle)

3.2 Evaluation on Synthetic images

We evaluated the accuracy of the spot detection methods using synthetic images of size 500 × 500 pixels, each containing 30 randomly generated spots. For each spot, the centroid and intensities are randomly selected, using parameters to define the range of random values from 100 to 255. Images were visually screened to count excessive spot overlap. We applied Gaussian noise with a range of standard deviation (σ = 0.000 …..0.015) to test the robustness limit of the detection method. The SNR of each image type was measured implementing the method described in chapter 2. We found perfect detection until the SNR of image approaches 1.8. Accuracy of the detection method was computed in terms of sensitivity and specificity.



A. Synthetic Fluorescence Image

B. Detected spots are surrounded by red circles

A. Sample image	B. Sample image	C. Sample image
SNR= 4	SNR= 2.5	SNR= 1.8

Figure 3.5: Sample images (Spot detection on synthetic image and SNR calculation)

3.3 Evaluation on real images

We performed evaluation on 10 real images for spot detection. The benchmark of true spot was defined by careful visual inspection. We computed the number of true positives (TP), false positives (FP), and false negatives (FN) to compute the accuracy of detection process. We also measured the noise level in each frame (SNR). The SNR of our fluorescent images was around 1.6.

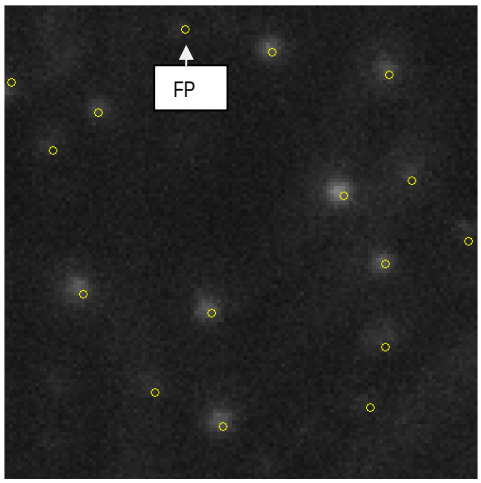
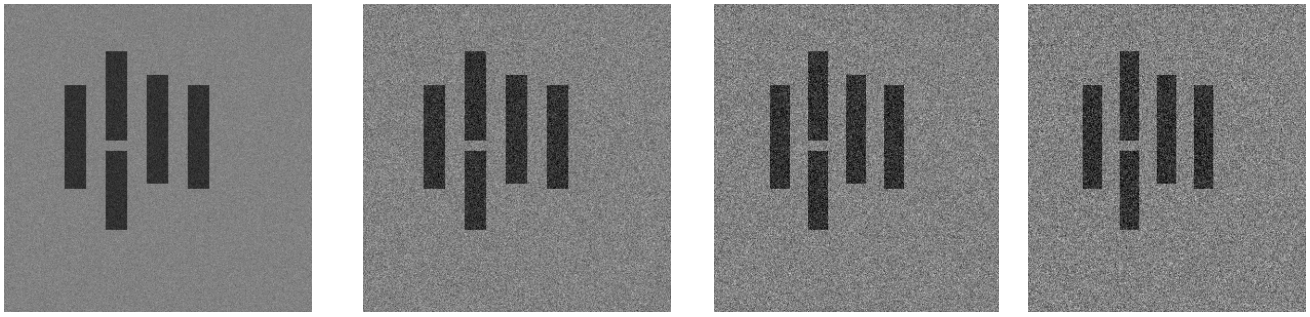


Figure 3.6: Detected Spots are circled in yellow and FP represents False positive

Calculated SNR of the particular image= 1.67

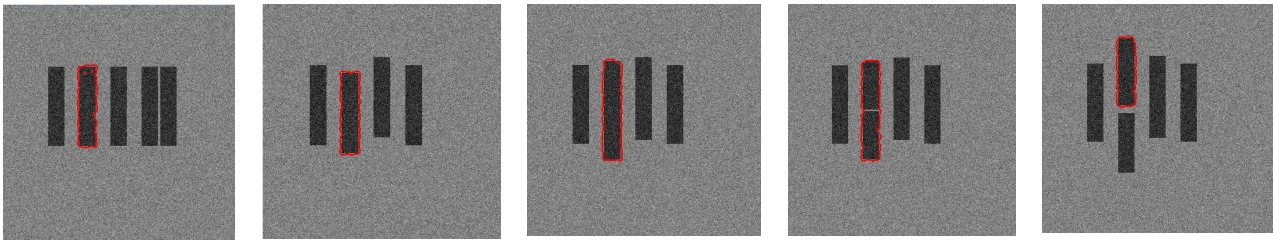
3.4 Evaluation of the cell detection and tracking algorithm on synthetic images

Cell detection and tracking was done on Time-lapse *Phase* Contrast Images with the help of a pre-existing MATLAB script. We also evaluated the performance of that pre-existing script. The algorithm of the script was based on the principle of detection-based tracking approach (i.e. first detects, then track). We generated synthetic Phase Contrast Images, where we obtained the image attributes (i.e. cell size, cell pixel intensity, background intensity) form real phase contrast images. We synthetically generated image sequence where we had the features of growth and division of bacterial cells. We tracked cells in the frames with our tracking algorithm to verify its tracking performance. We also applied Gaussian noise with a range of standard deviation (**σ** = 0.001, 0.05, 0.008. 0.01) to test the robustness of the cell detection process. We found that the algorithm works well until the noise (**σ**) = 0.008.



σ = 0.001 σ = 0.005 σ = 0.008 σ = 0.01

Figure 3.7: Synthetic phase contrast images with a range Gaussian noise. Each black box represents a bacterial cell (5x2 micron). Pixel values of cells and background resembles original phase contrast images.



Frame1(0min) Frame2(14min) Frame3(29min) Frame4(43min) Frame5(57min)

Figure 3.8: Cell tracking on a Synthetic phase contrast image (σ = 0.008). For example, the second cell (red border) was being tracked. Frame1: There are five cells in the frame; one is too close to another to make it more realistic Frame2: cells are randomly moved, one cell got out of the frame Frame3: the cell is growing fast Frame4: the cell was dividing Frame 5: the cell had multiplied.

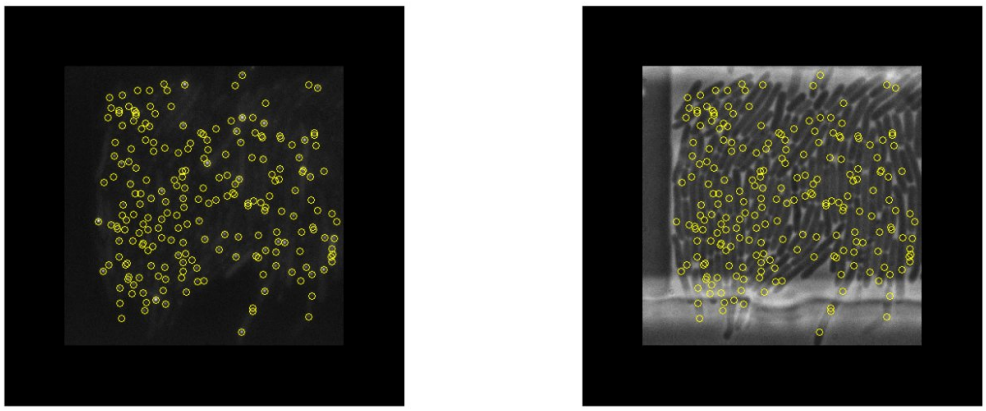
3.4 Spot matching algorithm

The objective of the algorithm is to find corresponding spot position of input fluorescence images with the corresponding phase contrast image. In our algorithm, we converted the (x, y) coordinate of pixels to an integer number with respect to a fame size. Then compute the maching pixels form cell segmentation output with the spot detection output.

$$n = A \cap B \text{ if and only if } n \in A \text{ and } n \in B.$$

n= number of spots belongs to cell, A= the pixels covered by a cell B= fluorescence spot position.

We plotted the spots on the phase contrast time to crosscheck the performance.



Detected spot position on fluorescence image(yellow circle)

A. Corresponding position on phase contrast image(yellow circle)

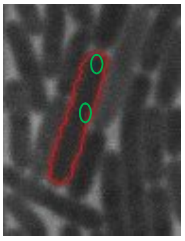
Figure 3.9. Position correspondence of spots on two related images

3.5 Biological experiment: Plasmid copy number per average cell volume

Bacterial plasmids maintain their number of copies by negative regulatory systems that adjust the rate of replication per plasmid copy in response to fluctuations in the copy number (Del Solar & Espinosa, 2000). We wanted to observe the relation

between plasmid copy number and cell volume. We found the there is correlation between cell volume and copy number of plasmid and the number increases with the cell volume

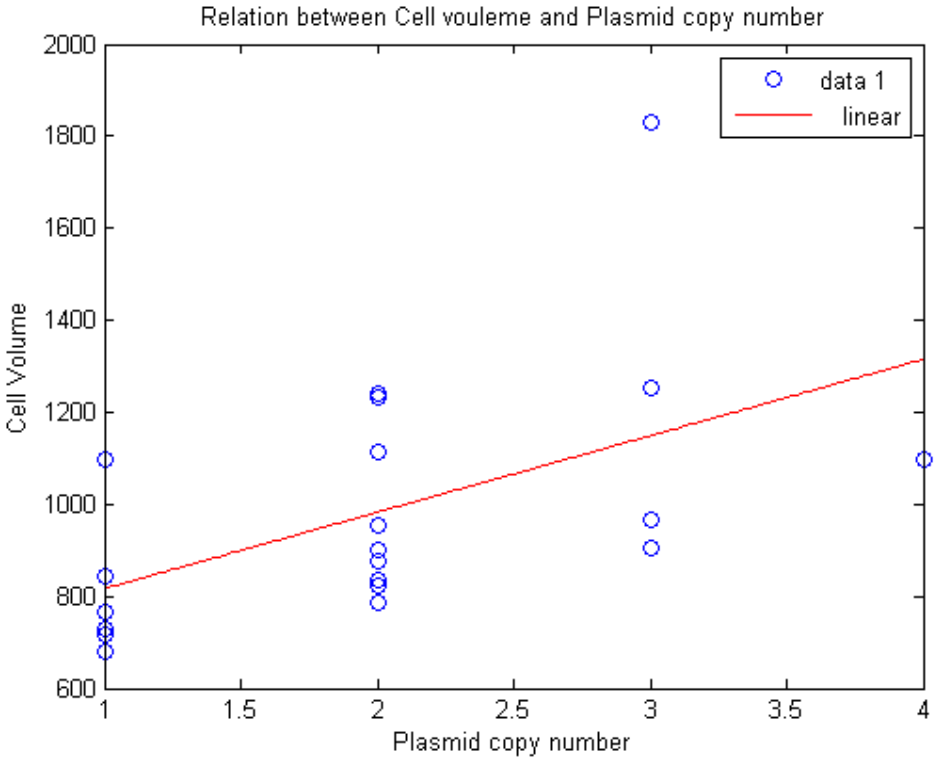
Copy number of plasmid= f(cell volume)= k ,In our case, K= 0.27 (sample size, n=20)



The selected region by red line is a cell (2D). Pixel size = 817 (1 pixel= 50x50 nm)

Green oval represent fluorescent spot in the corresponding fluorescence image

3.10 Figure: Example of cell segmentation and area calculation



Slop of the regression line = 0.27

3.11 Dot plot of the relation between plasmid copy number and cell volume

Chapter 4

Discussion

The Biological images are mostly noisy. The particular noise pattern existent in the images and the spot appearance is defined by a combination of environmental factors, stochastic biological events and the electronics of the imaging system. In live-cell fluorescence microscopy imaging, quantitative analysis of biological image data generally involves the detection of many subresolution objects, appearing as diffraction-limited spots. Due to acquisition limitations, the signal-to-noise ratio (SNR) can be extremely low, making automated spot detection a very challenging task.

We have presented a new spot detection method to detect bright spot in the cell nucleus. This method shows a clear improvement over the conventional Tophat method, because we used a secondary selection processes depending on the Gaussian fitting but it has a down side if we consider it especially with respect to computation time. Software is created giving the user the chance to choose deferent parts of image for spot detection because in some frame image quality defers in deferent parts of the image frame. This method can be applied to different kinds of images without the need of specifying application-related parameters. We have demonstrated its effectiveness with example images coming from biological microscopy applications.

Tophat algorithm was our choice for spot detection. The method of extracting spots using the Tophat filter is simple and straightforward utilizes the intensity and size information of the spots. We did not have to care for other parameters related to the apparatus. The idea is to define two circular regions D_{Top} and D_{Brim} which are related to the maximum expected spot radius and the shortest expected distance between spots, respectively. The purpose is to let the regions D_{Top} and D_{Brim} visit every pixel (i,j) in image I and for each (i,j) compute the average intensities I_{Top} and I_{Brim} after which we apply $C(i,j) = \begin{cases} 1 & x > \text{Threshold} \\ 0, & \text{otherwise} \end{cases}$. The connected components are finally taken to be the detected spots. The goal in computing the average intensities is to reduce the noise. The meaning of the parameter H is the minimum rise of a pixel region above its background for which we would consider that region a spot. So in summary, this method has three parameters (H,D_{Brim},D_{Top}). Tophat plays with the opening and closing operation on images. Since opening suppresses bright details smaller than the specified SE (structuring element), and closing suppresses dark details, they are often used in combination as morphological filters. As SE size

increases intensity contributions for larger dowels reduces , intensity contributions for larger dowels reduces. Comparing the opening by reconstruction to the regular opening, we notice that opening by reconstruction yields more uniform background. Subtracting the opening by reconstruction from the original image suppresses variations in background. For the comparison, a top-hat transformation (subtracting the standard opening from the original) does not suppress background variations.

The method we implemented works well for images with relatively high signal-to-noise (SNR). We have no option that would allow the user to edit spots after detection to optimize results. So, all of the results include only the spot lists automatically obtained by the algorithm using standard parameter settings. That is why, our result is absolutely objective.

In unraveling the underlying biology from the spot detection, the spots need to be detected in an accurate, unbiased and reproducible way. But, due to photobleaching effects, some image frames at the end the long experiments were noisy. So, the image analysis based on intensity thresholding is generally inadequate for those images. To deal with these artifacts we need to focus on the use of Bayesian model-based methods, which allow the incorporation of prior knowledge about object shape.

For our case study, we took a interest subject of biological noise. The life of plasmids of bacteria is a constant battle against fluctuations: failing to correct copy number fluctuations can increase the plasmid loss rate by many orders of magnitude, as can a failure to more evenly divide the copies between daughters at cell division. Plasmids are therefore long-standing model systems for stochastic processes in cells. Our preliminary data suggested a relation between cell volume and copy number but we need more experiments to come to a solid conclusion.

Chapter 5

Conclusion and future perspective

The method we developed is applicable to different kinds of images without specifying application-related parameters. We have demonstrated its effectiveness with example images coming from biological microscopy applications. We used it to measure plasmid copy number but this method can be useful in broad range of experiments like exploring protein localization and dynamics for genetic screening, tracking of individual fluorescent probes, Protein subcellular localization, Spatial-temporal dynamics. In the near future, these biological experiments could answer a number of unsolved questions related to biological noise. For example: Does a particular noise pattern exist for a particular gene family? Does the system ensure that genes or a category of genes always function within a definite noise threshold? What happens when the cells cross the boundary condition of noise? Are biological complexity and emergence products of noise? Does noise play a role in speciation? Knowing the answer to these questions, bio-engineers could make use of noise signatures and design novel genomes delicately harmonized to an explicit noise pattern. We anticipate that the future work will be the continuation of revealing noise and its role in diverse bio-systems at different levels.

References

Andrianantoandro, E., Basu, S., Karig, D. K., & Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, *2*(1), 2006.0028. EMBO and Nature Publishing Group. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1681505&tool=pmcentrez&rendertype=abstract

Boyd, A. C., Archer, J. A., & Sherratt, D. J. (1989). Characterization of the ColE1 mobilization region and its protein products. *Molecular general genetics MGG*, *217*(2-3), 488-498. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2671664

Breslow, R. (1972). Centenary Lecture. Biomimetic chemistry. *Chemical Society Reviews*, *1*(4), 553. doi:10.1039/cs9720100553

Bélanger, M. C., & Marois, Y. (2001). Hemocompatibility, biocompatibility, inflammatory and in vivo studies of primary reference materials low-density polyethylene and polydimethylsiloxane: a review. *Journal of Biomedical Materials Research*, *58*(5), 467-477. John Wiley & Sons, Inc. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11505420

Chalfie, M., & Kain, S. (1998). *Green fluorescent protein: properties, applications, and protocols. Methods of Biochemical Analysis* (Vol. 47, pp. 45-75). Wiley-Liss. Retrieved from http://books.google.com/books?id=v8Y4zrEofpIC&pgis=1

Cheezum, M. K., Walker, W. F., & Guilford, W. H. (2001). Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophysical journal*, *81*(4), 2378-88. doi:10.1016/S0006-3495(01)75884-5

Del Solar, G., & Espinosa, M. (2000). Plasmid copy number control: an ever-growing story. *Molecular Microbiology*, *37*(3), 492-500. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10931343

Dougherty, E. R. (1992). *An Introduction to Morphological Image Processing*. *Tutorial Texts in Optical Engineering* (Vol. TT9, p. 161). A publication of SPIE - the Int. Society for Optical Engineering (SPIE Press). Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=60790

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic Gene Expression in a Single Cell. *Science*, *297*(5584), 1183-1186. American Association for the Advancement of Science. doi:10.1126/science.1070919

Gil, J. Y., Kimmel, R., & Member, S. (2002). Efficient Dilation , Erosion , Opening , and Closing Algorithms, *24*(12), 1606-1617.

Hachey, D. L., & Chaurand, P. (2004). Proteomics in reproductive medicine: the technology for separation and identification of proteins. *Journal of Reproductive Immunology*, *63*(1), 61-73. Retrieved from http://www.sciencedirect.com/science/article/pii/S0165037804000658

Huang, B., Babcock, H., & Zhuang, X. (2010). Breaking the diffraction barrier: super-resolution imaging of cells. *Cell*, *143*(7), 1047-1058. Elsevier Inc. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21168201

Huang, B., Wu, H., Bhaya, D., Grossman, A., Granier, S., Kobilka, B. K., & Zare, R. N. (2007). Counting low-copy number proteins in a single cell. *Science*, *315*(5808), 81-84. AAAS. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17204646

Lemay, N. P., Morgan, A. L., Archer, E. J., Dickson, L. A., Megley, C. M., & Zimmer, M. (2008). The Role of the Tight-Turn, Broken Hydrogen Bonding, Glu222 and Arg96 in the Post-translational Green Fluorescent Protein Chromophore Formation. *Chemical Physics*, *348*(1-3), 152-160. Elsevier. doi:10.1016/j.chemphys.2008.02.055

Lichtman, J. W., & Conchello, J.-A. (2005). Fluorescence microscopy. (B. Hoffmann, Ed.)*Nature Methods*, *2*(12), 910-919. Nature Publishing Group. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16299476

Paulsson, J., & Elf, J. (2006). Stochastic Modeling of Intracellular Kinetics. (Z. Szallasi, J. Stelling, & V. Periwal, Eds.)*Differential Equations*, (8), 149-175. MIT Press,. Retrieved from http://www.vanderbilt.edu/viibre/members/documents/16091-Paulsson_In_SMCB-2006.pdf

Ritter, G., & Wilson, J. (2000). Handbook of Computer Vision Algorithms in Image Algebra. *Computer*, 417 pages. CRC Press. doi:10.1201/9781420042382

Schrödinger, E. (1992). *What is life?: The physical aspect of the living cell□; with Mind and matter□; & Autobiographical sketches*. *Mind and Matter* (p. 204). Cambridge University Press. Retrieved from http://books.google.be/books?id=dg2bYMwdaBwC

Smal, I., Loog, M., Niessen, W., & Meijering, E. (2009). Quantitative comparison of spot detection methods in live-cell fluorescence microscopy imaging. (W. C. Karl, B. Rosen, & D. Brooks, Eds.)*2009 IEEE International Symposium on Biomedical Imaging From Nano to Macro*, 1178-1181. Ieee. doi:10.1109/ISBI.2009.5193268

Swain, P. S., Elowitz, M. B., & Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12795-800. doi:10.1073/pnas.162041399

Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M., & Elowitz, M. B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, *440*(7083), 545-550. Nature Publishing Group. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16554821

Vliet, L. J. V., Sudar, D., & Young, I. T. (n.d.). Digital Fluorescence Imaging Using Cooled CCD Array Cameras invisible, *III*(Castleman 1996), 109-120.

Wilkinson, M. H. F., & Schut, F. (1998). Digital Image Analysis of Microbes: Imaging, Morphometry, Fluorometry and Motility Techniques and Applications. (M. Goodfellow, Ed.)*Modern Microbiological Methods*. John Wiley & Sons. Retrieved from http://books.google.com/books?id=8JJwuU13YPAC&pgis=1

Xie, X. S., Choi, P. J., Li, G.-W., Lee, N. K., & Lia, G. (2008). Single-molecule approach to molecular biology in living bacterial cells. *Annual review of biophysics*, *37*(1), 417-44. Annual Reviews. doi:10.1146/annurev.biophys.37.092607.174640

Young, J. W., Locke, J. C. W., Altinok, A., Rosenfeld, N., Bacarian, T., Swain, P. S., Mjolsness, E., et al. (2012). Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature protocols*, *7*(1), 80-8. doi:10.1038/nprot.2011.432