MSc Project Proposal: Cluster Analysis of PALM and STORM Generated Data-Sets as an ImageJ Plugin

Josh Wainwright
Supervisor: Iain Styles

June 2014

1 Background

In medical imaging, viewing individual features of single cells is essential to learn about how the processes in the cells take place, but also extremely difficult. To view something as small as a cell, fluorescent dyes are attached to the cell. When light of a certain frequency is shone on these dyes, the dye molecule is excited. When de-excitation occurs, a photon of light is released with a different wave length. This allows the dye molecules to be imaged.

Imaging objects becomes more difficult as they get smaller because of the wavelength of light. Once two objects are separated by a distance of an order similar to that of the wavelength (λ) of the light used to view them, it is no longer possible to resolve these two objects apart, instead all that can be seen is a blur of the two objects together. The dyes that are used usually respond in the visible frequency range, thus limiting the resolution of separate points to around 300 nm.

There are a number of techniques that have been developed to avoid this problem. Some of them employ different wavelengths of light. X-ray and electron microscopy use x-rays and electrons respectively which have shorter wavelengths than visible light and so can resolve smaller distances. These have problems, however, since the shorter wavelengths imply higher energies and so often cause damage or require destroying to the sample.

Two examples that are used in medical imaging are PALM [Owen et al., 2010] and STORM [Rust et al., 2006]. These both employ special dyes that allow molecules to be imaged at different times, a different subset of all the dye molecules are activated for each image. Using the point spread function (PSF) of the imaged molecule, the precise location can be estimated and combined with the locations of other molecules from other images of the cell. This allows the diffraction limit to be circumvented without affecting the sample material, producing a large number of points, each representing a single molecule attached to a position on the cell.

2 Project

This project aims to interpret the data that is produced by this analysis in a more efficient way than currently used methods and to provide certain quantitative statistics regarding the data; size of structure, number of points, density of points, etc. Current analysis methods have proven to be inefficient and are often not able to cope well with noisy data. This project will investigate alternative methods for identifying clusters in the data and ignoring erroneous points.

2.1 Uniform Discrete Cell Method

An initial method for identifying structure in the data points will involve splitting the image space into discretized square cells and treating each cell as a grey-scale pixel. The value that will be assigned to each pixel will depend on the number of data points that it contains and noise will be removed by thresholding the image to a predetermined limit value. This will require defining a cell size before starting and then analysing the image space as a whole.

2.2 Quadtree Method

A second method that will be investigated will be to use a quadtree abstract data type to arrange the data points such that structure emerges naturally. Each datum will be placed into the quadtree such that only up to a maximum number of points in each node is allowed. This should have the benefit of being much more dynamic, so being faster and less resource demanding to achieve. It will require that the data structure that is used can be traversed efficiently in order to determine adjacent nodes when extracting statistics.

With both of these initial approaches, a method for classifying the resultant information will need to be produced and tested. These strategies will be adapted through the development and testing stages depending on their performance with large real data-sets.

3 Deliverables

The final goal deliverable will be an easily usable and intuitive plugin for the ImageJ program. ImageJ [Rasband, 1997], developed by the Natural Institutes of Health, is used as the industry standard for analysis and manipulation of biological or medical images. Since it is a public-domain program with an open Java plugin architecture, this should be achievable.

4 Software Development Model

There is a distinction in steps between the initial design of the data structres and algorithms and the development of the program and the plugin for ImageJ. To make the most use of this, the project will follow an Agile development model. There will be stages of development of the algorithms and the plugin and these will be revisited as necessary during the development.

5 Preliminary Timescale

Stage	Tasks	Date
1	Research existing methods of cluster analysis and identify short comings.	20th June 2014
2	Build implementation of Uniform Discrete Cell method.	27th June 2014
3	Build implementation of Quadtree method.	11th July 2014
4	Test and compare previous algorithms. Perform timing and resource usage analysis.	18th July 2014
5	Build first iteration ImageJ plugin using chosen method.	25th July 2014
6	Using chosen method, implement cluster analysis algorithms.	1th August 2014
7	Add cluster analysis to ImageJ plugin.	8th August 2014
8	Performance and ease of use testing of plugin.	15th August 2014
8	Write-up of background research and current implementations investigation	18th July 2014
10	Write-up of data structure algorithms.	1st August 2014
11	Write-up of cluster analysis algorithms.	22nd August 2014
12	Final write-up of processes, improvements and end results of project.	29th August 2014

References

[Owen et al., 2010] Owen, D. M., Rentero, C., Rossy, J., Magenau, A., Williamson, D., Rodriguez, M., and Gaus, K. (2010). PALM imaging and cluster analysis of protein heterogeneity at the cell surface. *Journal of biophotonics*, 3(7):446–454.

[Rasband, 1997] Rasband, W. (1997). ImageJ, US National Institutes of Health. *Bethesda, Maryland, USA*, 2012.

[Rust et al., 2006] Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793–796.