# 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 ($\lambda$) 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 several techniques that have been developed to avoid this problem. 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 subset of all the dye molecules is imaged at a time. 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 and produces 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 from this analysis in a more efficient way, and to provide certain quantitative statistics regarding the data; size of structure, number of points, density, 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 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.

### 2.2 Quadtree Method

A second method that will be investigated will be to use a quadtree abstract data type to hold the data points. Each datum will be placed into the quadtree such that a maximum number of points in each node is allowed. This should have the benefit of being much more dynamic, so respond better to the types of structure expected.

With both of these initial approaches, a method for classifying the resultant information will need to be produced and tested.

## 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 an public-domain program with an open Java plugin architecture, this should be achievable.

## 4 Preliminary Timescale

| Stage | Tasks | Date |
|---|---|---|
| 1 | Research existing methods of cluster analysis and identify short comings. | 20th June 2014 |
| 1 | Build implementation of Uniform Discrete Cell method. | 24th June 2014 |
| 1 | Build implementation of Quadtree method. | 30th June 2014 |
| 1 | Test and compare previous algorithms. Perform timing and resource usage analysis. | 8th June 2014 |
| 1 | Using chosen method, implement cluster analysis algorithms. | |
| 1 | Write-up of background research and current implementations investigation | |
| 1 | Write-up of data structure algorithms. | |
| 1 | Write-up of cluster analysis algorithms. | |
| 1 | Final write-up of processes, improvements and end results of project. | |

## 5 Software Development Model

Since there is a distinction in steps between the initial design and development of the strategy to use to analyse the data and writing the plugin for ImageJ, the steps involved in this 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.

## 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.