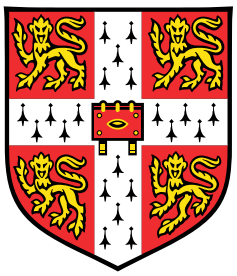
Using 3D image data to improve 2D Brightfield image segmentation

A method of 3D projection



Nicholas Piano

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Master of Philosophy

St Edmund's College December 2015

I would like to dedicate this thesis to my long-suffering supervisor, Y. Y. “Shery" Huang, for her help and advice. Without her guidance, this would not have been possible.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 15,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Nicholas Piano
December 2015

Acknowledgements

And I would like to acknowledge the support and patience of my parents, my colleague Cristina Bertulli, and Xiaohao Cai.

Abstract

Table of contents

List of figures

Chapter 1

Introduction

1.1 Cell segmentation in microbiology

An important part of live cell microbiology is the accurate measurement and tracking of cell morphology during an experiment. Using a microscope, there are many different ways of observing the cells ranging from brightfield microscopy to 3D fluorescence reconstructions. Both 2D and 3D shape data from the cells along with their speed and directionality can provide information on the effectiveness of drugs or other agents in the experiment. The processing of cell data often relies on the quality of Cell Segmentation, or the automatic or manual differentiation of Objects of Interest, such as cells, from the background. Many algorithms and software packages, such as Cellprofiler and ImageJ, are used to segment cells automatically, yielding variable quality.

A key limitation in widely used software that this paper seeks to address is the inability to account for consistent features that cannot be easily located in 3D data. 3D image data, such as from a confocal microscope, contains information about an environment on many focal planes [ref]. Objects can appear blurred or in focus depending the current focal plane. Consequently, features that are useful for segmentation; dark edges, uniform bright interiors, and other features such as fluorescent markers placed within cells are subject to any fluctuations in focus or the movement of objects vertically in the environment. This prevents consistent segmentation of the cell.

The environment used in this study is a microfludics chip built to simulate a human blood vessel. This type of chip, a microchannel framework printed on a PDMS substrate, is widely used in the medical industry to mimic body tissue [ref]. The 3D nature of this setup requires the use of confocal microscopy or similar methods to observe objects in all parts of the environment. In this case, a confocal microscope was used to record data in both a brightfield channel and a fluorescent GFP channel. A number of limitations on the

image quality, discussed in Chapter 2, prevent more powerful direct 3D methods to be used for segmentation. This necessitates a deeper investigation into how the currently available brightfield and GFP information can be used to segment the cells consistently.

1.2 Current work in context

The problem of consistently recognizing objects in a 3D environment using the brightfield is partially addressed in the 2009 study by Selinummi et al. [ref] They attempted to remove the problem of finding objects in 3D by studying the variance of vertical brightness profiles in the brightfield. This had the effect of simplifying 3D data into a 2D plane containing relevant object shapes, from which segmentation could be done more easily. Although this was effective in their case, it has several disadvantages that the current work aims to solve. Notably, it performs poorly in a multicellular environment, necessary for this and many other studies. This is discussed in more detail in the method in Chapter [ref, method].

The current work builds on the idea of studying vertical intensity profiles by applying this concept instead to the GFP. In the current data, the low quality GFP does not show the outlines of cells accurately. This is due to the internal cellular distribution of the GFP, described in Chapter 5. The central regions of the cell (not including the nucleus) are highlighted, so it can be used to locate the general bulk of the cell. The brightfield, on the other hand, shows the edges of the cell more clearly, but only if the cell is in focus (at the correct level). Edges of the cell start to fade along protrusions. In this study, any useful segmentation done must include accurate outlines of long cell protrusions to determine cell motility and behaviour.

The method described here uses the bulk of the cell visible in the GFP to locate the level in the data needed to ensure consistently clear dark edges in the brightfield, yielding accurate cell shape. The GFP and brightfield data represent the same physical space. This correspondence is exploited to allow information about the GFP channel to aid searching for features in the brightfield. The correct edges are located by building a vertical GFP intensity distribution for each pixel and selecting pixels from the brightfield data that match the level of the distribution peak. Several other properties of this GFP "profile" can be used to find different information about the cells. This is described in more detail in Chapter 6. This new method is one of image pre-processing; to prepare images for segmentation by software, in this case, Cellprofiler. It is not a method of segmentation, although as outlined in Chapter 7, it has the potential to become one.

1.3 Thesis outline

Firstly, a brief description of the microfluidics environment, its purpose, and its limitations will be given in Chapter 2. The system used and its properties are the original reason this study into image processing was undertaken. The 3D nature of the system prevents conventional 2D microscopy, and the live cell environment meant that more detailed cell fixing techniques could not be used. The lack of image quality and coarseness of the data led to the need to investigate if any useful information could be extracted reliably from it.

The current method depends heavily on the distribution of GFP within a cell. This, along with an understanding of cell shape and its optical properties and responses to the light used in the microscope is important for determining how useful this method can be in changing conditions. If the type of microscope or environment is changed, this method should be able to operate under a new set of parameters. A description of the places GFP can be found within a cell can inform adaptations of this method and help find conditions where the method could fail or have reduced effectiveness. An example of this is cytosolic versus cytoskeletal GFP, found in the central cytosol and the external cytoskeleton respectively. These two different distributions of GFP can effect this method which relies on peak positions of the profiles. Peaks in different distributions can give misleading answers regarding the positions of parts of the cell. A description of the brightfield reveals similar situations. A key example is the superposition of two objects in the environment. The brightfield light is mixed in the final data, so data representing superimposed objects cannot be divided into specific shapes of both objects. This limitation cannot be resolved with the GFP data. This and other problems will be described in Chapter 3.

To explain the context of this type image pre-processing, a background in image processing and some techniques used by Cellprofiler will be given in Chapter 4. Basic image features can be combined into larger features that can delimit objects and parts of objects. Software like Cellprofiler can use this information to measure shapes and other properties of cells in an image. It can also be used to track cells from one frame to another by looking for similar cell shapes and proximity. The method described here depends on the way a recognizer like Cellprofiler works, although with further work, the information used for pre-processing in this method could allow a more powerful segmentation algorithm to be created. This will be investigated finally in Chapter 7.

Chapter 6

Finally, testing will be used to compare this method to previous experimental work and commonly used methods of image pre-processing. Although this is not yet a fully capable segmentation method, the pre-processing can be compared by segmentation. Data from several experiments was collected and cancer cells within them segmented. Several properties

of the cancer cell shapes, such as the projected 2D area, can be used to compare images on their ability to represent the cells.

Chapter 2

The microfludics environment

2.1 Purpose and the need for segmentation

The reason this environment was developed is to determine how cancer cells cross the endothelium (blood vessel wall cell barrier) during extravasation, or the stage of metastasis where cancerous cells that have already entered the blood stream from some initial location, make their way around the body, break through the blood vessel wall, and embed themselves in another type of tissue. Extravasation is a poorly understood process in cancer biology [ref], and this type of system could help to understand it by providing accurate data on the morphodynamics of the cells as they move in the simulation of a blood vessel.

To provide useful data, information about a great many cells must be gathered. This requires visually identifying cells and parts of cells and assessing their changing shapes, sizes, and speeds over time and in distinct regions of the environment. This can be done manually by keeping track of each cell and recording its progress over the course of the experiment. Unfortunately, this is impractical for the amount of data needed and would be too time consuming for a human to do. As such, it must be done automatically. This type of recognition done by a computer is known as segmentation, where cell shapes are separated and differentiated from the background. Many algorithms, techniques, and pieces of software exist to solve this with varying levels of success. They require a mathematical description of a set of images, and a way to consistently find features of the objects of interest that can be matched and measured.

Certain aspects of the environment used in the experiment and the methods by which it was imaged pose problems for accurate cell segmentation by disrupting the consistency that cell segmentation algorithms depend on. Traditional cell microscopy using glass slides and chemically fixed cells allows for high resolution images of cells that are consistently in focus. Factors affecting the quality in this case are the 3D structure of the channel, the

live cell environment where cells are free to move and interact, and the movement of the microscope hardware and cancer cells vertically (relative to the microscope objective) that causes object to fluctuate in and out of focus in the final data.

2.2 Description of the environment

The microfludics environment used for this study consisted of a microchannel framework printed using soft lithography on a PDMS chip. PDMS, or Polydimethylsiloxane, is a a silicone gel that can be molded and used to create microscopic structures, such as microscopic channels to use as environments for cells [ref]. This microchannel technique is widely used in the biomedical industry to mimic body tissue for complex in vitro [italic] studies. In this case, the channel was set up to model a human blood vessel. The diagram in Figure [ref] below shows part of the channel framework with two PDMS pillars on either side of a gap. On one side of the gap, a liquid medium meant to simulate blood plasma is pumped in where it can remain static, or be used to simulate the flow inside a blood vessel. On the other side of the gap, collagen gel, used to simulate the extra-cellular matrix surrounding a blood vessel, provides an anchor point for the endothelial cells, or cells found in the blood vessel wall, to attach to. These endothelial cells are added to the environment prior to the experiment. The experiment is monitored when cancer cells that are marked with a fluorescent GFP, or Green Fluorescent Protein, are added to the environment in the medium channel, from which they can cross the endothelial cell barrier into the collagen gel, as they might in vivo [italic] during the extravasation stage of metastasis [ref].

The entire channel is approximately 100 microns thick [ref]. this is roughly an order of magnitude greater than the width of a typical cancer cell used in the experiments [ref]. During the experiment the cells may be attached (flattened) and thus much thinner vertically [ref]. The width of the gap horizontally containing the endothelial cell barrier is between 100-200 microns [ref]. A typical blood vessel would be tubular, but this setup, while possible [ref], would be difficult to study and to image with a microscope. Opting instead for the simplified setup, the vertical wall of cells can be observed much more easily and still be used to provide information on how the cancer cells cross the barrier. More advanced setups are under development at the time of writing.

2.3 Confocal microscope and 3D structure

Image data were gathered using a confocal microscope. This allows a 3D space to be imaged by scanning over a volume and recording data at a number of discreet levels. Data can

also be recorded in a number of different channels. The data for these experiments used brightfield, or transmission, images and GFP fluorescence images. The light in brightfield images from different levels is subject to mixing due to the optics of the microscope objective, resulting in blurred representations of objects unless their level is in focus. In contrast, light from GFP inside cells can be isolated to a particular level using the pinhole mechanism of the microscope. The size of the pinhole determines the maximum resolution of the image. This isolation of the GFP channel levels can be used to build up a 3D model of the objects marked with GFP, which in this experiment is only the cancer cells. Cancer cells can then be distinguished from other types of objects in the experiment.

The 3D information from the GFP can be used to locate cancer cells within the environment, but the limits on the resolution prevent accurate outlines of the cells from being found. This is also dependent on the cells' internal distribution of GFP. The cancer cells in the current experiments had GFP embedded within the cytosol, or the fluid interior of the cell. From this, the general bulk of the cell and to a certain extent its general shape could be made out, but finer edges and protrusions could not be determined since the GFP did not penetrate the cytoskeleton or the outer wall of the cell. The method used to find cells and prepare the images for segmentation is directly dependent on the this arrangement of the GFP. modifications can be made to account for other sources of fluorescence inside the cell. This is discussed more in Chapter [ref].

2.4 Live cell imaging limitations

Imaging live cells imposes limits on the techniques available since the cells must remain alive and be able to behave naturally for the entire duration of the experiment. This means they cannot be subjected to unnatural amounts of illumination from the laser used to stimulate fluorescence lest they absorb too much energy and overheat, causing erratic movements and finally death, making any data on their movements unusable. This limit on the amount of light that one is able to put into the cells to stimulate the GFP limits the amount of light that can be gathered from the cells. This lowers the brightness and contrast of the final data, making the location of finer details in a cell less certain.

The cells are also free to move about the environment. Tracking a single cell means it must remain in the field of view. The time period necessary to allow the cells to cross the endothelial barrier is between 10-15 hours, and the distance a cell can move in this time sets a lower limit on the size of the viewport, or the view of the microscope that is finally recorded as an image. A large viewport means a lower resolution, since the objective must move away from the sample. This sacrifices image quality per cell for information about

many cells. For this trade-off to be worthwhile, useful data must be extracted from the lower quality cell images. This is difficult for simple cell segmentation algorithms.

The movement of the cells is not restricted to the two dimensions of the viewport. The cells can move up and down in the environment (towards and away from the objective). This will move them between different focal planes. While all this data is recorded, cells cannot be guaranteed to stay on one level for the whole experiment. As a cell moves vertically, its representation in the image will change and properties of its edges will change their colour and texture. This breaks the feature consistency that segmentation relies on.

2.5 Autofocus and focal fluctuations

Apart from the behaviour of the cells, the microscope hardware must also operate reliably for the 10-15 hour time period. This is not always possible, and small changes in the alignment of different components will cause systematic errors to build up and compound over time. The result of the microscope imaging in one alignment while the instruments report another causes errors in the consistency of the image sequence. The focus will fluctuate visibly while an index representing the focal plane remains the same. This means such an index cannot be relied on the accurately report the location in the environment.

Software that is meant to solve this problem is known as Autofocus [ref]. Most commercial microscope setups come with this software preinstalled. It attempts to analyse the images gathered at regular intervals, for example, every frame, and determine the current focal plane in the environment from simple numerical properties. Such a property is the minimum entropy of the image [ref]. By keeping the minimum entropy constant, the image plane with the most similar entropy in the next frame can be found. This is set as the corresponding level so that representations of the objects remain constant. Other methods such as measuring the reflection of light from the cover slip of the sample are also used [ref]. Highly variable environments such as live cells can cause misleading calculations of a constant entropy, hence the unreliability of the level index. This makes segmentation more complicated.

Chapter 3

Background

3.1 Introduction

A comprehensive background on the key concepts that the method presented in this thesis is based on is essential for understanding its context and significance. The previous chapter outlined some of the problems faced when attempting to segment cancer cells given a 3D live cell environment. These included autofocus fluctuations of the microscope and limits on light absorption by cells.

3.2 Cell microscopy, optical structure, and GFP distributions

When imaging cells, the visible parts of the cell and the optical properties of any edges and colours observed depend on the materials that make up the cell.

Transparency and visible cell edges Cell edges and interiors Cell features: protrusions, attachment, cell edge and halo effect GFP injection and different distributions of GFP inside a cell

3.3 Image processing and segmentation

Edge detection Blob detection Watershed and distance transforms: separating objects 3D segmentation

3.4 Cellprofiler and segmentation software

Cellprofiler secondary objects ImageJ lasso tool Lack of 3D support

3.5 Cell tracking

The LAP tracking algorithm Errors caused by long time delay between frames

3.6 The Selinummi brightfield profile method

The 3D brightfield profile Mathematical properties of the profile Disadvantages of the Selinummi method

Chapter 4

Method

4.1 The GFP profile

Properties of the vertical GFP distribution Z position Variance image compared with absolute GFP intensity

4.2 Optimum features for cell recognition

Dark edges and uniform object interiors Vertical difference between maximum GFP and ideal features

4.3 zMod and zBF

zMod: use information about Z position of each pixel from profile zBF: map zMod to brightfield data and sample corresponding pixels into a final 2D image

4.4 Artificial edges for segmentation: zDiff and zEdge

Cellprofiler secondary object segmentation cannot be bounded Impose boundaries using GFP halo to denote maximum extent of the cell

Chapter 5

Results and discussion

5.1 zMod parameters: R, Sigma, and Delta-z

Several constants were used to make zMod and zBF. They can be varied to produce different results.

5.2 zDiff and zEdge

Artificial edges can cut off parts of the cell that do not contain enough GFP.

5.3 Comparison with common methods

Previous methods cannot account for the inconsistencies in the focus fluctuations and so are not really comparable in their quality.

5.4 Errors and limitations

Lack of GFP Cross-level focus artifacts

Chapter 6

Conclusion

6.1 Summary

New method can exploit the 3D representation possible with GFP to correct focus fluctuations in the brightfield and greatly improve segmentation accuracy.

6.2 Observations

Segmentation accuracy increased by an order of magnitude.

6.3 Further work

Could be used to help locate seeds in images to simplify microscope imaging. Currently the method only does pre-processing. Can be used instead to do segmentation directly using more information than a current segmentation algorithm is able to use.

References