

# **A Dynamical Systems Modelling Framework for Breast Cancer Cell Motility and Morphology Analysis**

A thesis submitted to the University of Sheffield for the degree of Doctor of  
Philosophy

**Yang Zhang**

Department of Automatic Control and Systems Engineering

January 2017





*To my family*



## Acknowledgements

I would like to express a special thanks to my supervisor Professor Visakan Karikamanathan. His tremendous guidance and mentor on my research not only helped me get through my PhD study, but also broaden my research views, where I discovered my career directions. A particular thanks for the support and opportunities he has given me at the beginning of my research career.

I would like to thank my second supervisor Dr. Sean Anderson, who encourages me all the time and gives me guidance and suggestions on both research and general life. I have learnt a lot research management and presenting skills from the way how he works efficiently and how he cheers up the people around him.

I also want to thanks to Dr. Andrew Hills who introduced the La, Python, Iceberg, Git and most of the tools I have used in my research. Thanks for that he shared his experiences and encouraged me to get through the toughest time during my PhD.

I would like to mention Dr. Julia E. Sero and Dr. Chris Bakal for the data collections for the project and assisting the biological background knowledge about cancer cells.

Thanks to Dr. Geoffrey Holmes for his guidance on helping me with the tracking of cell migration.

Additionally, I thank Dr. Parham Aram and Dr. Veronica Biga for all the help and advice on my studies in the UK over the years.

I also want to thank my best friends Will Jacobs, Tara Baldacchino for helping me get used to the culture and life in the UK and for all the time we spend together which makes my study in UK more enjoyable.

Finally, I especially thank my parents, my mother Dongping and father Saishi, for all their love, support and understanding.

## Abstract

Cancer is a worldwide disease and, in the UK, breast cancer is the most common. Compared to healthy cells, cancer cells migrate abnormally, associated with alterations in cell motility and morphology. The development of biomedical imaging techniques result in the production of large amounts of data. The analysis of such large data, the variety of cancer cell shapes and the potential links between cell motility and morphology present a challenge for cell migration study: how to analyse cell motility and morphology simultaneously.

This thesis proposes a computational framework to address integrated cancer cell migration analysis. Firstly, automated tracking of cell boundaries is undertaken by a discrete white noise acceleration (DWNA) kinematic model of cell boundaries, described by B-spline active contours. The tracked cell states intrinsically links cell morphology to motility features. As a result, cell centroid and boundary dynamics are successfully tracked, followed by quantitative motility analysis.

A module to quantitatively analyse cell morphology is proposed after tracking. Cell shapes are described by a 2D descriptor. Accordingly, cell morphodynamics are modelled as a hidden Markov process, along with three shape states: round, elongated and teardrop. In order to explore the potential interactions between cell shapes and motility, cell centroid motility characteristics are associated to the identified shape states. When the analysis was applied to breast cancer control cells, the identified shape states showed distinct motility characteristics.

Finally, the proposed framework is adapted to the comparison of MDA-MB-231 cell behaviours with regulating migration-associated proteins: i) Blebbistatin and Y-27632, which are chemical inhibitors of two different proteins working on the same pathway, showed identical, but different degrees of effects on the motility and morphology characteristics of MDA-MB-231 cells. ii) The absence of FA-associated genes, including focal adhesion kinase (FAK), RhoE and  $\beta$ -PIX, respectively showed distinct effects on cell migrations.

# Contents

<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Algorithms</b>	<b>ix</b>
<b>Nomenclature</b>	<b>xi</b>
<b>Acronyms</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	2
1.2 Aims and objectives . . . . .	3
1.3 Structure of the thesis . . . . .	4
1.4 Contributions . . . . .	5
<b>2 Computational Methods in Cell Behaviour Study</b>	<b>7</b>
2.1 Cancer cell migration . . . . .	8
2.1.1 Biomedical imaging techniques . . . . .	8
2.1.2 Cancer metastasis and cell migration . . . . .	9
2.1.3 Lamellipodia in cell migration . . . . .	12
2.1.4 Cell adhesion dynamics and migration . . . . .	12
2.2 Estimation and discrete state-space models . . . . .	13
2.2.1 Non-Bayesian parameter estimators . . . . .	15
2.2.2 Bayesian parameter estimators . . . . .	18
2.2.3 Discrete-time kinematic models . . . . .	20
2.2.4 State estimation in discrete-time dynamic systems . . . . .	23
2.2.5 Hidden Markov models and expectation maximisation . . . . .	27
2.2.6 Smoothing for discrete-time state-space models . . . . .	29
2.3 Methods for motility and morphology analysis . . . . .	30
2.3.1 Overview of object tracking . . . . .	30

2.3.2	Image segmentation . . . . .	34
2.3.3	Active contour models for shape representation . . . . .	38
2.3.4	Morphology description techniques . . . . .	40
2.3.5	Cluster analysis . . . . .	42
2.3.6	Clustering algorithms . . . . .	44
2.4	Summary . . . . .	49
<b>3</b>	<b>High-throughput Tracking of Cancer Cells Using Parametric Active Con- tour and Kalman Smoothing</b>	<b>51</b>
3.1	Automatic cell boundary segmentation using marker-controlled wa- tershed algorithm . . . . .	53
3.2	Cell boundary representation by B-spline active contours . . . . .	55
3.2.1	The construction of B-spline active contours . . . . .	56
3.2.2	The parameter selection of B-spline active contours . . . . .	57
3.3	Cell boundary tracking with Kalman filter estimation and smoothing	60
3.3.1	A discrete white noise acceleration model of cell migration . . . . .	60
3.3.2	Cell tracking with Kalman estimation and smoothing . . . . .	62
3.4	Cell motility analysis . . . . .	67
3.4.1	Cell centroid motility features . . . . .	68
3.4.2	Cell boundary motility features . . . . .	69
3.5	Breast cancer migration . . . . .	71
3.5.1	Cell segmentation results . . . . .	71
3.5.2	Cell tracking results . . . . .	73
3.5.3	Statistical analysis of cell motility . . . . .	76
3.6	Conclusion . . . . .	79
<b>4</b>	<b>A Computational Module for Cell Morphology Analysis</b>	<b>81</b>
4.1	Cell shape description . . . . .	82
4.1.1	Fourier descriptors of cell contour . . . . .	83
4.1.2	Feature extraction of cell boundary descriptors . . . . .	84
4.2	Cell shape clustering . . . . .	88
4.2.1	Cell shape clustering using K-means . . . . .	89
4.2.2	Cell shape clustering using Gaussian mixture model . . . . .	90
4.3	Cell morphodynamics analysis . . . . .	94
4.3.1	Description of cell shape evolution based on Gaussian mix- ture model . . . . .	94
4.3.2	Cell morphodynamics modelling using hidden Markov model	95
4.4	Morphology analysis of breast cancer cells . . . . .	100
4.4.1	Quantitative morphodynamics analysis . . . . .	101

---

4.4.2	Integrated motility and morphology analysis . . . . .	105
4.4.3	Morphodynamics analysis of single cell migration . . . . .	106
4.5	Conclusion . . . . .	112
<b>5</b>	<b>Comparison of MDA-MB-231 Cell Behaviours with Regulating Migration-Associated Proteins</b>	<b>113</b>
5.1	Experimental procedures . . . . .	115
5.1.1	Introduction of experimental plasmids . . . . .	115
5.1.2	Experimental data . . . . .	116
5.2	Quantitative motility and morphology analysis of MDA-MB-231 cells treated by blebbistatin and Y-27632 . . . . .	118
5.2.1	Analysis of cell centroid motility . . . . .	118
5.2.2	Analysis of cell boundary motility . . . . .	121
5.2.3	Integrated analysis of cell morphology and motility . . . . .	123
5.2.4	Analysis of cell morphodynamics . . . . .	124
5.3	Quantitative motility and morphology analysis of MDA-MB-231 cells in the absence of focal adhesion associated genes . . . . .	126
5.3.1	Analysis of cell centroid motility . . . . .	128
5.3.2	Analysis of cell boundary motility . . . . .	129
5.3.3	Integrated analysis of cell morphology and motility . . . . .	132
5.3.4	Analysis of cell morphodynamics . . . . .	133
5.4	Conclusion . . . . .	136
<b>6</b>	<b>Conclusions and Future Work</b>	<b>139</b>
6.1	Summary and conclusions . . . . .	139
6.2	Future work . . . . .	142
	<b>Bibliography</b>	<b>144</b>

# List of Figures

2.1	Major steps of cell migration. . . . .	10
2.2	Different migration modes of cancer cells. . . . .	11
2.3	Diagrams of lamellipodia. . . . .	12
2.4	Cell adhesion dynamics. . . . .	14
2.5	Flow diagram of the Kalman filter (KF). . . . .	25
2.6	The algorithm of the interacting multiple model (IMM) estimator. . . . .	26
2.7	Diagram of hidden Markov model (HMM). . . . .	28
2.8	The demonstration of a Rauch-Tung-Striebel (RTS) smoother. . . . .	31
2.9	The basic idea of 1-D watershed segmentation. . . . .	38
2.10	An example of cell contour representation by geometric active contour. . . . .	39
2.11	An example of B-spline contours. . . . .	40
2.12	Examples of cell shape descriptors. . . . .	42
2.13	Procedure for cluster analysis. . . . .	43
3.1	Overview of the cell tracking framework. . . . .	52
3.2	An example of landscape view of a grey-scale image. . . . .	54
3.3	An example of fitting the same cell boundary using B-spline basis functions of different spline orders. . . . .	58
3.4	The mean squared error (MSE) and algorithm running time of B-spline estimation with different number of spans. . . . .	59
3.5	Illustration of cell contour correspondence. . . . .	66
3.6	An example of cell centroid trajectory. . . . .	68
3.7	Illustration of cell orientation and protrusion identification. . . . .	70
3.8	Image segmentation process. . . . .	72
3.9	Examples of the image segmentation algorithm working on the images with cell shape variety and different image quality. . . . .	74
3.10	Tracking of cell trajectories and velocities. . . . .	75
3.11	Analysis of cell centroid motility. . . . .	77
3.12	An example of cell orientation identification. . . . .	78



---

3.13	Example of cell protrusion identification. . . . .	78
3.14	Polar histogram of protrusion distributions associated with cell orientation (J21-F1). . . . .	79
4.1	Flowchart of cell morphology analysis. . . . .	82
4.2	The normalisation of shape descriptors. . . . .	85
4.3	Feature extraction from the Fourier descriptors (FDs) . . . . .	87
4.4	Cell shape description. . . . .	88
4.5	Cell shape clustering using K-means. . . . .	90
4.6	The identification of a Gaussian mixture model of cell shape distribution by the expectation maximisation (EM) algorithm. . . . .	93
4.7	Lattice diagram of cell shape transitions. . . . .	96
4.8	Flowchart of cell morphology analysis. . . . .	100
4.9	Results of cell clustering based on HMM. . . . .	103
4.10	The comparison of cell migration speed of different shapes. . . . .	107
4.11	The comparison of cell migration angle change for different shapes. . . . .	108
4.12	Cell speed with respect to direction change for different shape state. . . . .	109
4.13	Tracking of single cell migration. . . . .	110
4.14	Motility and morphodynamic characteristics of single cell migration. . . . .	111
4.15	Shape tracking in single cell migration. . . . .	111
5.1	Diagram of the integrated cell migration analysis framework. . . . .	114
5.2	Quantitative analysis of cell centroid speed and direction change. . . . .	119
5.3	Cell centroid speed in respect to migration direction change. . . . .	120
5.4	Quantitative analysis of cell directional migration. . . . .	121
5.5	Quantitative analysis of cell boundary motility. . . . .	122
5.6	Population analysis of cell shape state. . . . .	124
5.7	Quantitative centroid motility analysis of breast cancer cells subjected to knockdown of focal adhesion associated genes. . . . .	129
5.8	Cell directional migration analysis. . . . .	130
5.9	Quantitative analysis of cell boundary motility. . . . .	131
5.10	Cell migration images extracted from time-lapse image sequence. . . . .	132
5.11	Ratio of cells of different shape states. . . . .	133

# List of Tables

2.1	The comparison of cancer cells in different migration modes . . . . .	11
3.1	Experimental information for identification of the number of spans on B-spline contours. . . . .	60
4.1	Initialisation of Gaussian mixture model using K-means. . . . .	91
4.2	The parameters of converged Gaussian mixture model. . . . .	94
4.3	Experimental information. . . . .	101
4.4	Count of shape transitions between adjacent frames. . . . .	102
4.5	Parameters of shape feature distribution in the identified HMM. . .	102
4.6	The Sojourn time of being each state. . . . .	104
5.1	Data used for comparison of blebbistatin and Y-27632. . . . .	117
5.2	Data used for the comparison of focal adhesion associated genes. . .	117
5.3	Count of shape transitions between adjacent frames (Control cells). .	125
5.4	Count of shape transitions between adjacent frames (Blebbistatin). .	125
5.5	Count of shape transitions between adjacent frames (Y-27632) . . . .	125
5.6	The shape transition relations of breast cancer cells treated with chemical inhibitors blebbistatin and Y-27632. . . . .	127
5.7	Count shape transitions between adjacent frames (Control cells for gene tests). . . . .	134
5.8	Count of shape transitions between adjacent frames (FAK knock-down cells). . . . .	135
5.9	Count of shape transitions between adjacent frames (RhoE knock-down cells). . . . .	135
5.10	Count of shape transitions between adjacent frames ( $\beta$ -PIX knock-down cells). . . . .	135
5.11	The shape transition relations of breast cancer cells with different gene knocked-down. . . . .	137

# List of Algorithms

2.1	Hierarchical clustering: Agglomerative approach. . . . .	45
2.2	Partitional clustering: K-means algorithm. . . . .	47
3.1	Cell Tracking Algorithm. . . . .	63



# Nomenclature

A list of the variables and notation used in this thesis is defined below. The definitions and conventions set here will be observed throughout unless otherwise stated. For a list of acronyms, please consult page xiii.

$(x, y)$	Cartesian coordinates
$*$	Convolution operation
$\langle, \rangle$	Inner product
$[a, b]$	Closed interval between $a$ and $b$
$\arg \max$	Argument that maximises
$\arg \min$	Argument that minimises
$\cap$	Logical “AND” operation
$\cup$	Logical “OR” operation
$\Delta T$	Sampling period
$\hat{x}$	Estimate of $x$
$\in$	Element of
$\ \cdot\ $	Euclidean norm operator
$\mathbb{E}(\cdot)$	Expectation operator
$\mathbb{R}^L$	Real numbers of dimensions $L$
$\mathbf{0}$	Zero matrix
$\mathbf{I}$	Identity matrix
$\mathbf{X}$	Matrix

---

$\mathbf{x}$	Vector
$\mathcal{N}(\mu, \Sigma)$	Normal distribution with mean $\mu$ and covariance $\Sigma$
$\mathcal{Z}$	Set
$\tilde{x}$	Estimation errors
$\top$	Transpose
$\triangleq$	Defined as
$\emptyset$	Empty set
$\{z_n\}_{n=1}^N$	The sequence $z_n$ , where $n = 1, \dots, N$ .
$p(\cdot   \cdot)$	Conditional probability operator
$p(\cdot)$	Probability operator
$\ll$	Much less than
$\mathbb{Z}$	Integers

# Acronyms

**ARMA** autoregressive moving average. 49

**CA** constant acceleration. 22

**CSS** curvature scale-space. 41, 42

**CT** constant turn. 22

**CV** constant velocity. 21, 22

**DMEM** Dulbecco's Modified Eagle's Medium. 101

**DNA** deoxyribonucleic acid. 9, 48

**DWNA** discrete white noise acceleration. ii, 21, 30, 31, 60, 73, 79, 140

**DWPA** discrete Wiener process acceleration. 21, 22

**ECM** extracellular matrix. 9–12

**EGF** epidermal growth factor. 12

**EKF** extended Kalman filter. 24–26, 30

**EM** expectation maximisation. vii, 29, 48, 49, 91, 93, 96–98, 102

**FA** focal adhesion. 4, 115, 116, 126, 127, 134, 136, 138, 141

**FAK** focal adhesion kinase. ii, viii, 4, 5, 115–117, 126–138, 141, 142

**FBS** Fetal Bovine Serum. 101

**FD** Fourier descriptor. vii, 4, 41, 42, 82–87, 93, 112, 140, 142

**GFP** green fluorescent protein. 12

**GPS** Global Positioning System. 26

- HMM** hidden Markov model. vi–viii, 4, 5, 15, 27–29, 49, 94–96, 99–105, 112, 124, 125, 133, 134, 140, 143
- HTS** High-throughput screening. 9
- IMM** interacting multiple model. vi, 25, 26, 51, 143
- JPDAF** Joint Probability Data Association Filter. 142
- KF** Kalman filter. vi, 4, 24–26, 30–32, 60, 62, 80, 140
- LS** least squares. 17, 18, 20, 56, 57, 62
- MAP** maximum a posteriori. 18–20
- MHT** Multiple Hypothesis Tracking. 142
- MISE** mean integrated squared error. 85, 87
- ML** maximum likelihood. 15–17, 19, 20
- MMSE** minimum mean square error. 19, 20
- MSE** mean squared error. vi, 57, 59
- NN** nearest-neighbour. 62, 73, 80, 140, 142
- PC** principal component. 86, 88
- PCA** principal component analysis. 4, 41, 140
- PDF** probability density function. 15, 18, 19, 23
- ROCK** Rho-associated protein kinase. 13, 114, 115, 118, 136, 138
- RTS** Rauch-Tung-Striebel. vi, 4, 30, 31, 62, 140
- UKF** unscented Kalman filter. 25



# Chapter 1

## Introduction

Cancer has existed for as long as human history, but in medicine, the first written description was about breast cancer in approximately 3000 BC [55]. After 5,000 years, cancer has become a disease that affects everyone – male and female, wealthy and poor, young and old – and is one of the leading causes of mortality throughout the world. In the year 2012, cancer deaths worldwide were approximately 8.2 million, accounting for 14.6% of all human death. Around 70% of cancer deaths occurred in middle-income and low-income countries [162]. There are more than 100 types of cancer. The top 5 most common causes of cancer death worldwide are the lung, liver, stomach, colorectal and breast cancers [161]. If detected at an early stage, most types of cancer, such as colorectal and breast cancers, can be cured with adequate treatments [162]. Thus, large amounts of scientific research efforts have been invested into developing cancer diagnosis and treatments.

In the UK, the most common cancer is breast cancer. 53,696 new cases were diagnosed in 2013, of which 99% are females and 50% are over the age of 65 [27]. In developed countries, breast cancer is a common type of cancer, but has high survival rates with almost 78% of women surviving over 10 years after their diagnosis of breast cancer in the UK (2010–2011) [27]. Meanwhile, a list of breast cell lines have been established since the first line in 1958, BT-20 [84], which acts as a source for experimental testing of treatments, and as databases for scientific study of the behaviour of breast cancers.

Cancer, which starts as an abnormal cell, is capable of travelling to multiple sites through human blood and the lymphatic system through metastasis [109]. The migration of breast cancer cells has not been fully understood. A wide range of research has been proposed based on the studies on the established breast cancer cell lines, such as studies of cells migration mechanisms [29], cell adhesion

network [174] and to test different treatments that affect cell behaviour [120].

The challenges in studying breast cancer cell migration are in extracting information efficiently from the cell lines to analyse the cell features. The cell lines can involve multiple cell migrations, which show wide variation on cell quantities, motilities and morphologies. Different cell lines can be cultivated under different conditions and the screening presented in various formats, such as phase contrast images, fluorescent images and quantitative phase contrast images [182]. The data complexity and diversity that emerges from the screening experiments requires a framework that can extract information from cell lines and process the extracted information to analyse cell migration characteristics. This thesis proposes a novel integrated framework that consists of a module of automated multiple cell migration tracking based on time-lapse greyscale image sequences and a module of statistical analysis of cell motility and morphology features.

This chapter begins with the motivation of the project in Section 1.1 and aims and objectives in Section 1.2. The thesis structure is outlined in Section 1.3 and the original contributions are summarised in Section 1.4.

## 1.1 Motivation

The development of time-lapse fluorescence microscopy techniques enables the observation of cellular dynamics over time, which is primarily used for fundamental research of cell biology process or clinical *in vitro* fertilisation [37]. The large amount of data in the image sequences give rise to the necessity of automated visual tracking and computational analysis methods for describing the dynamic cellular phenotypes.

Cell morphology contains information pertaining to particular cell behaviours. For example, cellular senescence is depicted as cells of large flat morphology, whereas proliferating cells maintain a balanced morphology through cell division [88]. Most animal cells present elongated or spindle morphologies in interphase, but round up to spherical shapes during mitosis [26]. In epithelial-mesenchymal transition, the epithelial cells have apical-basal polarity, tight cell-cell junctions and lose adhesions and polarisation. After transition, the invasive isolated mesenchymal stem cells are morphologically characterised by spindle-like shapes [83]. Notably, a collection of cell morphological signatures can indicate the activities of specific signalling networks [6]. When combined with microenvironment read-outs, cell shapes can be used to explore signal transduction [138].

Distinct cell states corresponding to different motility patterns underlie some of the hypotheses for cell behaviour studies. For example, compared to epithe-

lium, mesenchymal stem cells show much stronger migration ability [83]. In comparison with healthy cells, tumour cells present continuous migratory behaviour in the absence of stop signals [48]. As such cell motility is generally used to quantify the effects of chemical drug treatments. Various motility measurements have been proposed, such as the migration distance, confinement ratio and migration direction change [10].

To date, automated visual tracking and computational analysis methods have provided various studies of cell motility and morphology independently. Automated visual tracking of cell centroids and subcellular molecular dynamics have been widely employed to gain insight into understanding cell migration, lineage structures and proliferation [10, 86, 92]. Cell morphology is studied separately through computational analysis of morphology signatures that can contribute to the investigation of chemical substance testing or gene expressions [174, 181]. These strategies for cell properties analysis are limited in either focusing on cell motility descriptions that neglect cell morphology characteristics, or on statistical analysis of cell morphology signatures, disregarding motility characteristics. This separate analysis can limit the insight into understanding cell behaviours, such as cell cycle and epithelial-mesenchymal transition, which are schemes that inherently link cell motility and morphology. Therefore, this thesis is aimed at developing an integrated framework that includes dynamical system descriptions, combined estimation of cell motility and morphology and characterisation of cell migration states.

## 1.2 Aims and objectives

The aim of this thesis is to develop a dynamical systems modelling and estimation framework to analyse the migration of breast cancer cells through the quantitative analysis of cell motility, cell morphology and integrated analysis. According to the aims of the framework, the objectives of the project are given as follows:

- Develop an automated visual tracking algorithm to perform the automated tracking of multiple cells motility and morphology dynamics. The tracking algorithm will be illustrated by applying to the time-lapse grey-scale image sequences of breast cancer cells.
- Develop a quantitative motility analysis module based on the tracking results. The module will include the analysis of cell global dynamics, described by cell centroid motility features, and local skeleton dynamics, described by boundary motility features.

- Develop a morphology analysis module by using cell morphology tracking results. The objective of this module is to identify different cell states according to their shapes and compare the motility characteristics of different cell states.
- Apply the developed computational framework to quantitatively compare the effects of chemical inhibitors on breast cancer cell migration. The computational framework consists of tracking, motility analysis and morphology analysis modules. The databases include the wild breast cancer cells, cells treated by blebbistatin and Y-27632, respectively.
- Apply the developed computational framework to analyse the effect of focal adhesion (FA) associated genes. The breast cell lines, MDA-MB-231, were cultivated by depletion of genes: focal adhesion kinase (FAK), RhoE and  $\beta$ -PIX.

### 1.3 Structure of the thesis

This thesis proposes an integrated framework for cell migration analysis, including cell tracking, motility and morphology analysis. The structure of this thesis is summarised as follows:

- Chapter 2 introduces the theory and techniques associated to computational cell behaviour study. The mechanisms of cell migration are introduced in this chapter. The basic estimation theory and discrete kinematic models are reviewed. Additionally, a number of computational analysis methods, including the image segmentation, shape representation and clustering algorithms, are described.
- Chapter 3 proposes a framework for automated multiple cell boundary tracking, combined with statistical motility analysis. Cell boundaries are represented by B-spline active contours. Subsequently, the boundary tracking is performed by using Kalman filter (KF) and Rauch-Tung-Striebel (RTS) smoothing to track cell contours. The cell centroid and boundary motility characteristics are analysed quantitatively based on the B-spline representations.
- Chapter 4 presents a cell morphology module based on cell boundary tracking results. Cell shapes are described by two dimensional descriptors that are obtained by the application of Fourier descriptor (FD) and principal component analysis (PCA). In addition, hidden Markov model (HMM) is applied

to the shape descriptors. Three different cell shape states and their corresponding transition characteristics are identified.

- Chapter 5 presents the application of the proposed framework. The computational framework is used to perform quantitative comparisons of motility and morphology of breast cancer cells that were cultivated by chemical inhibitors: blebbistatin and Y-27632. In addition, the computational framework is utilised to quantitatively analyse the functions of adhesion-associated genes: FAK, RhoE and  $\beta$ -PIX.
- Chapter 6 summarises the developed cell migration analysis framework and the conclusions of the applications of the framework. Future directions for this framework including its improvement and potential applications are presented.

## 1.4 Contributions

The contributions of this work are summarised as follows:

- An automated multiple cell boundary tracking algorithm is presented and illustrated by application to a time-lapse image sequence of wild breast cancer cells (Chapter 3). The work was presented at the 37-th International Conference of the IEEE Engineering in Medicine and Biology Society as a poster 'High-throughput Tracking of Cancer Cells using Parametric Active Contours and Kalman Smoothing', co-authored with Julia Sero, Geoffrey Holmes, Chris Bakal, Sean Anderson, Visakan Kadirkamanathan.
- A framework for cell migration analysis is developed (Chapter 3 and Chapter 4). The framework is constructed on a cell-centric scenery where the motility and morphology information of cell migration is organised based on single cell track. Intrinsic link between cell motility and morphology is performed through tracking the dynamics of cell shapes represented by parametric active contours. Cell states are explored by applying hidden Markov model (HMM) algorithm to cell morphology characteristics. The cell-centric scenario not only provides the information for single cell migration study, but can also be aggregated for population level analysis.
- The proposed framework is used to analyse the effects of different biochemical inhibitors on cell migration and the functions of different focal adhesion-associated genes (Chapter 5). The framework for cell migration analysis and

the application are organised as a manuscript in draft form for submission to an international journal,

- Yang Zhang, Julia Sero, Geoffrey Holmes, Andrew Hills, Sean Anderson, Stephen Renshaw, Chris Bakal, Visakan Kadiramanathan. Integrated Motility and Morphodynamics Analysis of Breast Cancer Cells.

## Chapter 2

# Computational Methods in Cell Behaviour Study

The first written description of breast cancer occurred in approximately 3000 BC [55]; since then, especially in modern times, cancer has become one of the leading causes of deaths worldwide [55]. Cancers are capable of spreading to different organs and tissues from an initial cancerous site through the lymph and blood systems. This invasion process is called cancer metastasis, which is the major cause of cancer-related deaths [161]. Therefore, cancer cell migration, which is the basis of cancer metastasis, is a key mechanism that needs to be identified for understanding particular forms of cancer progression.

Bioinformatics, which integrates techniques from the fields of computer science, engineering, statistics and mathematics to analyse large biological databases, has become an essential tool for biological systems study since its introduction in the 1970s [59]. Specifically, signal processing techniques, such as mathematical models linked with estimation theory, can assist the study of cell behaviours by enabling the description of cell dynamics over time. These mathematical descriptions can give efficient insight into key attributes of cell migration.

Furthermore, visualisation and pattern-recognition methods can be used to statistically measure cell behaviours based on large-scale databases. The measurements of dynamic cellular features can be used to assess and compare the behaviours of different cells.

The computational methods used in this study to process large amounts of information about cell migration aim to give insight into cell dynamics. After an introduction to cancer cell migration, including the basic mechanisms, cell lamellipodia structures and adhesion dynamics in migration, are provided in Section 2.1; an introduction to computational models used in this study is provided:

- Estimation theory and discrete state-space models, which are necessary for mathematical description of cell migration, are described in Section 2.2.
- Statistical methods for motility and morphology analysis are presented in Section 2.3.

## 2.1 Cancer cell migration

Cancer, namely breast cancer, was first recorded in the Egyptian Edwin Smith Papyrus, around 3000 BC [55]. In the 15<sup>th</sup> to the 17<sup>th</sup> centuries, doctors began to determine the cause of death by dissecting bodies, and they regarded cancer as a contagious disease [55]. Due to the development of microscope techniques in the 18<sup>th</sup> century, cancers were discovered to be capable of spreading from one site to another by travelling through the lymph nodes, which is known as metastasis [53]. Today, cancer is defined as a set of diseases caused by the abnormal growth of cells that are capable of metastasis to multiple organs [146].

Cancer, as a group of diseases, has caused deaths worldwide. In 2012, 8.2 million people died because of the diseases related to cancers and 70% of these deaths occurred in Africa, Asia and Central and South America [161]. In addition, most cancer patients have cancer metastasis, which is the primary cause of mortality. Therefore, a wide range of research focuses on understanding cancer metastasis. This section first gives an overview of cancer cell migration, including the development of biomedical imaging techniques, features of cancer cell migration, as well as protrusions and adhesion dynamics in cell migration.

### 2.1.1 Biomedical imaging techniques

Biomedical imaging techniques are used to acquire images of biological and medical phenomena in the interior regions of the body, such as organs, tissues and cells [8].

Traditional fluorescent microscopy refers to an optical microscope that uses fluorescence and phosphorescence to mark proteins and generate fluorescence micrographs for the visualisation of either the cells or sub-cellular molecular activities [182]. In 2014, Eric Betzig, William Moerner and Stefan Hell introduced fluorescence microscopy into the nano-dimension and were awarded the Nobel Prize in Chemistry [31]. Fluorescent microscopy usually requires the use of a set of sophisticated equipment, such that most operations heavily rely on the meticulously trained research operators.

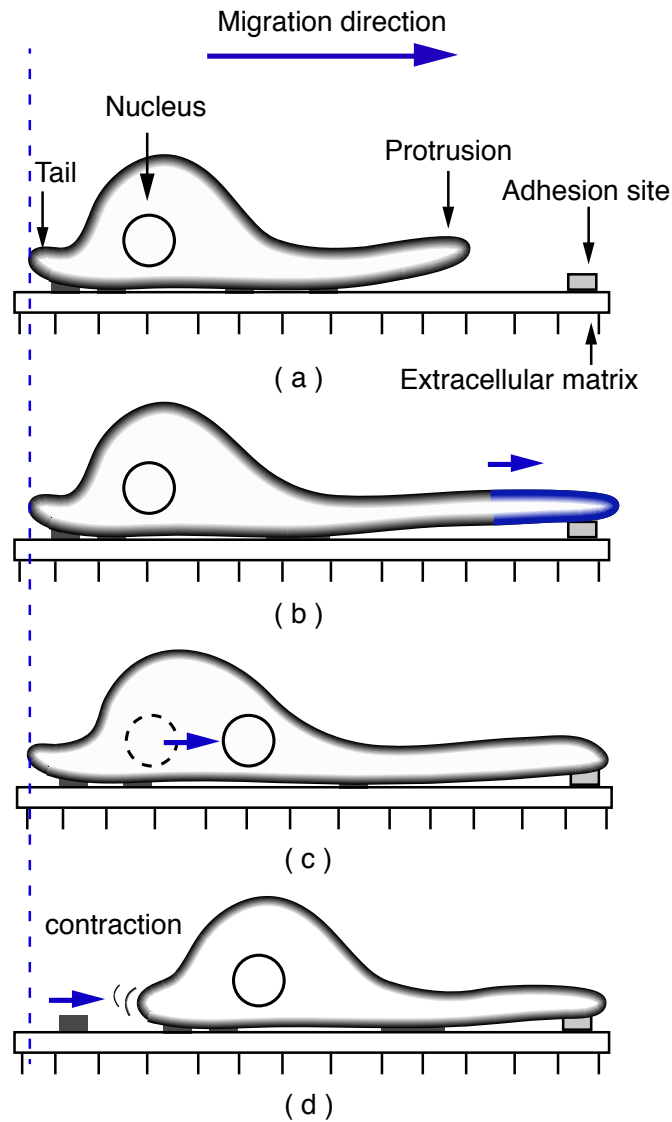


High-throughput screening (HTS) was developed in the early 1990s at about the same time as the first automated deoxyribonucleic acid (DNA) sequencers were generated. HTS achieves high throughput by using high-content screening and automated motorised microscopes. For example, an ultra-high-throughput screening can screen over 100,000 compounds every day [155]. In 2010, an HTS process with screening 1000 times faster than previous conventional processes was developed [1]. Recently, time-lapse data acquisition techniques were added to HTS to monitor intracellular or cellular activities [57, 182]. Time-lapse high-throughput microscopy, which was used in this project to capture the breast cancer cell migration, provided a large amount of precise and informative digital image sequences. The large scale and complexity of the data is a challenge for further processing.

### 2.1.2 Cancer metastasis and cell migration

Cancer is originally caused by one or more DNA mutations that alter the normal cell cycle and lead to unregulated proliferation with unusual cytological features. Cancer metastasis starts from angiogenesis, which is the formation of new blood vessels [167]. Angiogenesis supplies the necessary condition for the primary tumour to find access to the bloodstream. Tumour cells attach to healthy cells and invade new organs or tissues via the body's blood circulatory system or lymphatic system. The newly arrived cancer cells proliferate in the new site and may lead to further metastasis [30, 164]. General cells, such as leucocytes and blood cells, migrate in tissues following a sequence of temporally integrated steps [104]. Cell migration is first activated by the extracellular matrix (ECM), which causes the cells to be polarised along the migrating direction (see Figure 2.1a). The extracellular matrix (ECM) exists in the extracellular space and consists of water and molecules that are secreted by cells. Also, the ECM contributes to maintaining cell structure and chemical balance and plays an essential role in regulating cell dynamics [46]. Next, a protrusion of the cell membrane extends in the migration direction and the cell generates new adhesion at the cell's leading edge (see Figure 2.1b). Traction forces are generated because of the protrusion extension, which then carry the cell body moving forwards (see Figure 2.1c). In the last step, the tailing region of the cell breaks its original adhesion and contracts against the rest of the cell body to complete the forward migration (see Figure 2.1d). During migration, cell motility and morphology vary continuously, along with the changes in the extracellular environment [48, 127].

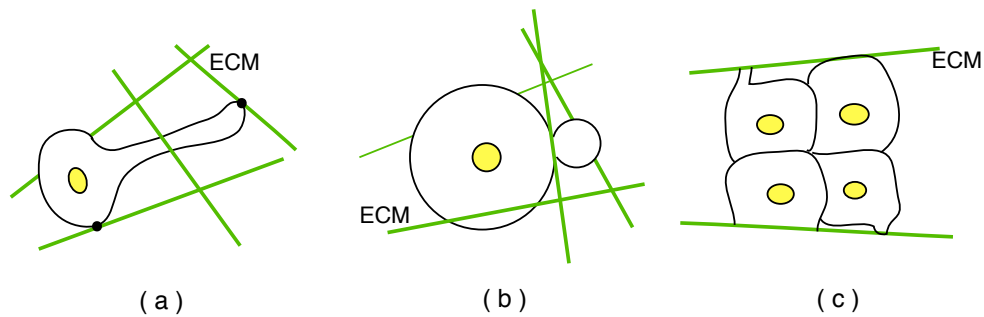
Invasive cancer cells migrate following the same basic mechanism as general cells, but have their own characteristics, such that different migration modes are



**Figure 2.1: Major steps of cell migration [104].** (a) Cell migration is driven by the ECM; (b) A cell protrusion extends toward the migration direction and generates new adhesions at cell leading edge; (c) Cell body is carried forward; (d) The cell breaks the original adhesion on cell trailing edge, which is retracted against the cell body.

observed during the cancer cell movement [151]. Depending on the interaction between cells, the migration of cancer cells is divided into two types: collective migration, in which cells have tight junctions, and individual cell migration, with weak junctions between cells.

The individual migration mode is divided into two types according to the mechanisms used by the cells to migrate. One is mesenchymal migration, where a



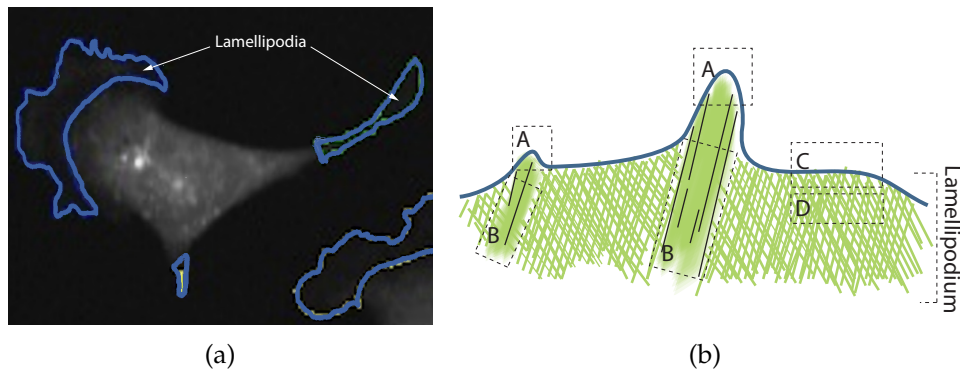
**Figure 2.2: Different migration modes of cancer cells [47, 48].** (a) Mesenchymal; (b) Amoeboid; (c) Collective cells. The green lines around cells represent collagen fibres, which are the major component of ECM.

**Table 2.1: The comparison of cancer cells in different migration modes.**

Modes	Mesenchymal	Amoeboid	Collective
Diagram	Figure 2.2a	Figure 2.2b	Figure 2.2c
Morphology	<ul style="list-style-type: none"> <li>• Elongated or spindle-shaped</li> <li>• Length: 50–200 <math>\mu\text{m}</math></li> <li>• Cell-substrate adhesive</li> </ul>	<ul style="list-style-type: none"> <li>• Rounded-shaped</li> <li>• Length: 0–30 <math>\mu\text{m}</math></li> <li>• Weak cell-substrate adhesive</li> </ul>	<ul style="list-style-type: none"> <li>• Sheets, strands, clusters or ducts</li> <li>• Large cell mess</li> <li>• Cell-cell adhesive</li> </ul>
Motility	<ul style="list-style-type: none"> <li>• Smaller velocity: 0.1–1 <math>\mu\text{m}/\text{min}</math></li> </ul>	<ul style="list-style-type: none"> <li>• Velocity in range: 0.1–20 <math>\mu\text{m}/\text{min}</math></li> </ul>	<ul style="list-style-type: none"> <li>• Higher invasion efficiency</li> </ul>

cell is elongated and needs an ECM gradient difference to inspire the movement. The other one is amoeboid migration where cells have rounded shapes and move independently of ECM degradation [96].

With respect to cells in collective migration mode, they move in groups and have strong communications between adjacent cells [5, 48]. A comparison of cell motility and morphology characteristics is provided in Table 2.1 [47, 48]. Moreover, recent studies have proved that cancer cells are able to transition between different migration modes under certain treatments [96, 171]. For example, mesenchymal-amoeboid transitions can be realised via changing the extracellular environment by applying five different protease inhibitors simultaneously [163].



**Figure 2.3: Diagrams of lamellipodia.** (a) The lamellipodia of a breast tumour cell in a grey scale image; The light grey parts, highlighted by blue lines are lamellipodia and the region of dark grey is the main body of cell; (b) A schematic representation of lamellipodia [139]: (A) filopodium tips; (B) bundle; (C) lamellipodium tip; (D) actin meshwork.

### 2.1.3 Lamellipodia in cell migration

The Lamellipodium is a layer of broad and light cytoplasm structure and plays an essential role in cell migration. Figure 2.3a shows the lamellipodia of a breast cancer cell observed under a fluorescence microscope and Figure 2.3b depicts the compositions of a lamellipodium.

Lamellipodia are formed through actin polymerisation (see Section 2.1.4), this is used to monitor the behaviour of lamellipodia. Cell migration is led by lamellipodia through the generation of protrusions, with a balance between lamellipodia assembling at a cell's front and disassembling at rear [85, 110, 139, 148]. Currently, researchers have been focused on how to regulate the actin polymerisation, so as to affect lamellipodium-based cell motility. It was proposed that epidermal growth factor (EGF)-activated chemotaxis can be used to stimulate lamellipodia protrusion activities by using green fluorescent protein (GFP) tagging [36, 58]. At the same time, lamellipodia extension can be controlled by restraining cellular traction force and cell shapes, which indicates that lamellipodia not only relate to cell motility but also to morphological characteristics [115, 137].

### 2.1.4 Cell adhesion dynamics and migration

Cell migration is generally activated by extracellular signals, such as ECM proteins, changes in chemical gradients and mechanical forces [180]. Initially, cells polymerise the actin filaments at the leading edge to generate protrusions on the cell membrane towards the extracellular signal. The adhesions that connect the cytoskeleton to ECM proteins then stabilise the protrusions, as lamellipodia and

traction forces are generated at the same time. These forces contribute to the adhesion disassembly at the trailing edge of the cells so that the cells are able to travel forwards. Accordingly, cell migration is driven by the formation and disassembly of cell adhesions [127].

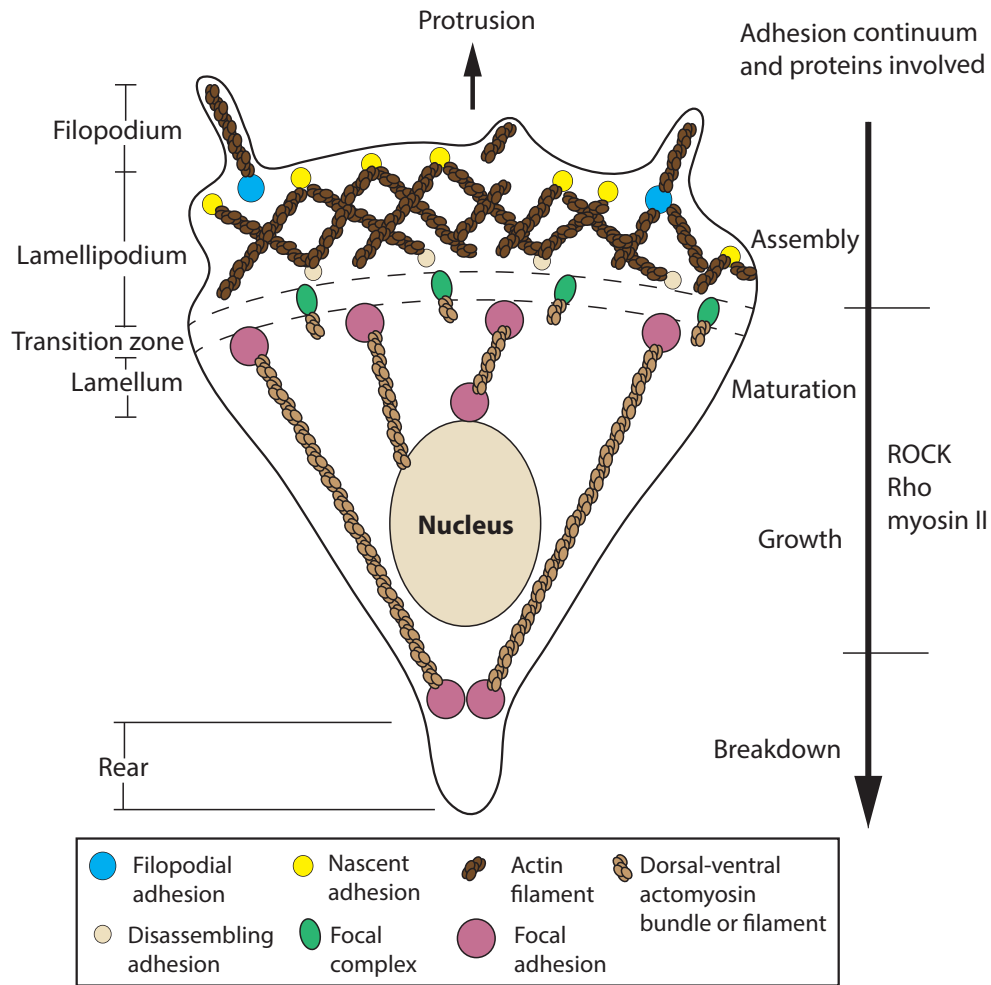
Cell adhesion is a dynamic continuum structure, depicted in Figure 2.4. The adhesion cycle consists of adhesion assembly, maturation, growth and disassembly [116]:

- Adhesion assembly: Cell migration begins with an extension of the protrusion at the leading edge, where nascent adhesions are formed. At this stage, the protrusions are not stabilised, but the adhesions strengthen the polymerisation of actin at the leading edge, so as to generate the protrusions.
- Adhesion maturation and growth: The nascent adhesions are linked to actomyosin that exist in the lamellum area so that they become mature as adhesion complexes and focal adhesions stabilise lamellipodia. Although it is still not clear how adhesion maturation is accomplished, myosin II is shown to be able to generate the tension to sustain adhesion maturation by the Rho activation, which is regulated by the actions of Rho-associated protein kinase (ROCK) [33].
- Adhesion disassembly: At this stage of adhesion maturation, some nascent adhesions, are not able to merge into the lamellipodium and lamellum. Instead, they disassemble at the cell front when the lamellipodium moves. Disassembly also happens at cell retracting areas, including the cell front and rear.

As introduced above, cell migration is driven by adhesion dynamics, which are regulated by protein expressions, such as myosin II actin polymerisation and ROCK. Hence, cancer cell migration is expected to be regulated by the control of the adhesion cycle or adhesion-associated proteins. The cell adhesion network, as a highly informational and complex system, incorporates over 180 nodes of protein to protein interactions [178]. However, most of those proteins and the way they cooperate are still not fully explained, which leads to the necessity for these protein-related experiments and for an integrated analysis framework for the results analysis.

## 2.2 Estimation and discrete state-space models

Systems biology combines the techniques from system engineering, statistical analysis, biophysics and other fields for the quantitative or qualitative study of biolog-



**Figure 2.4: Cell adhesion dynamics.** Adhesion dynamic is a continuous and cyclical system, which consists of adhesion assembly, adhesion maturation and growth and breakdown of adhesion [116].

ical systems [40, 149]. Mathematical and computational models are commonly constructed to explain and characterise the underlying mechanisms [45, 123].

Parameter estimation is a typically static estimation problem. Generally, the system is represented by regression type equations, where parameters are constant over time [7]. Cell boundary contour estimation can be thought of as a parameter estimation problem. State estimation is a dynamic estimation problem. The state space model representation is required and model parameters need to be known. The variables are dynamic and harshly measured, requiring reconstruction from observed variables [75, 80]. Cell velocity and position estimation as well as cell shape tracking can be considered as problems of state estimation. Slowly changing

parameter estimation can be represented as a state estimation problem with a suitable model characterising the nature of slow variation. In some problems of state estimation, model parameters may be unknown. In this case, the problem becomes a state and parameter estimation problem, such as in a HMM.

### 2.2.1 Non-Bayesian parameter estimators

Parameter estimation (static estimation) problems are described as follows [7].

The measurements  $z_n$  are given by

$$z_n = h(n, \theta, \omega_n) \quad n = 1, \dots, N, \quad (2.1)$$

where  $h(\cdot)$  is the function relating the parameters to the measurements.  $n \in \mathbb{Z}$  is the index of measurements and is in the range  $(0, N]$ ,  $\theta$  is an unknown parameter that is assumed to be time invariant, and  $\omega_n$  is the disturbance. The estimation task is to find a function of the  $N$  observations

$$\hat{\theta}_N \triangleq \hat{\theta}(N, \mathcal{Z}) \quad (2.2)$$

This function is called the estimator and  $\mathcal{Z}$  is a set of observations,

$$\mathcal{Z} \triangleq \{z_n\}_{n=1}^N \quad (2.3)$$

with the estimation error given by,

$$\tilde{\theta} \triangleq \theta - \hat{\theta} \quad (2.4)$$

#### Maximum likelihood estimator

In the case of non-Bayesian parameters, a likelihood function,  $\Lambda_{\mathcal{Z}}(\theta)$ , which is obtained from a probability density function (PDF) of the observations given by the parameters, is constructed to measure how likely the observations are to have arisen from the the chosen value of parameters,

$$\Lambda_{\mathcal{Z}}(\theta) \triangleq p(\mathcal{Z} | \theta) \quad (2.5)$$

The maximum likelihood (ML) estimator is then computed by maximisation of the likelihood function [7],

$$\hat{\theta}^{ML} = \arg \max_{\theta} \Lambda_{\mathcal{Z}}(\theta) \quad (2.6)$$

For example, for a set of measurements with independent and identically zero-

mean Gaussian distributed noise  $\omega \sim \mathcal{N}(0, \sigma^2)$ .

$$z_n = \theta + \omega_n \quad n = 1, \dots, N \quad (2.7)$$

where  $\theta$  is an unknown constant parameter. Thus,

$$z_n \sim \mathcal{N}(\theta, \sigma^2) \quad (2.8)$$

According to the definition (2.5), the likelihood function is constructed as:

$$\begin{aligned} \Lambda_{\mathcal{Z}}(\theta) &\triangleq p(z_1, \dots, z_N | \theta) = \prod_{n=1}^N \mathcal{N}(z_n; \theta, \sigma^2) \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^N} e^{-\frac{1}{2\sigma^2} \sum_{n=1}^N (z_n - \theta)^2} \\ &= c e^{-\frac{1}{2\sigma^2} \sum_{n=1}^N z_n^2} e^{-\frac{1}{2\sigma^2} (N\theta^2 - 2\theta \sum_{n=1}^N z_n)} \\ &= c f_1(\mathcal{Z}) f_2(\mathcal{Z}, \theta) \end{aligned} \quad (2.9)$$

where,

$$c = \frac{1}{(\sqrt{2\pi}\sigma)^N} \quad (2.10)$$

$$f_1(\mathcal{Z}) = e^{-\frac{1}{2\sigma^2} \sum_{n=1}^N z_n^2} \quad (2.11)$$

$$f_2(\mathcal{Z}, \theta) = e^{-\frac{1}{2\sigma^2} (N\theta^2 - 2\theta \sum_{n=1}^N z_n)} \quad (2.12)$$

The ML estimation of  $\theta$  is then computed by maximising the likelihood function through setting the first derivative equal to zero.

$$\frac{d\Lambda_{\mathcal{Z}}(\theta)}{d\theta} = 0 \Rightarrow \frac{df_2(\mathcal{Z}, \theta)}{d\theta} = 0 \quad (2.13)$$

$$\Rightarrow \hat{\theta}^{ML} = \frac{1}{N} \sum_{n=1}^N z_n = \bar{z} \quad (2.14)$$

In (2.14), it is seen that the ML estimation is equal to the mean value of the measurements. The expectation of estimation error is:

$$\mathbb{E}(\tilde{\theta}) = \mathbb{E}(\theta - \hat{\theta}) = \mathbb{E}(z_n - \omega_n - \bar{z}) = \mathbb{E}(\omega_n) = 0 \quad (2.15)$$

It shows that the ML estimate is unbiased when the noise  $\omega_n$  is zero-mean Gaussian distributed [7, 118].



### Least squares estimator

The least squares (LS) method is another common non-Bayesian estimation. Given the measurements

$$z_n = h(n, \theta) + \omega_n \quad n = 1 \dots N \quad (2.16)$$

the corresponding LS estimator is derived by the minimisation of quadratic estimation errors [91, 145, 145]:

$$\hat{\theta}_N^{LS} = \arg \min_{\theta} J = \arg \min_{\theta} \left( \sum_{n=1}^N [z_n - h(n, \theta)]^2 \right) \quad (2.17)$$

If  $\omega_n$  is assumed to be independent and identical zero-mean Gaussian distributed random noise of variance  $\sigma^2$ , the measurements satisfy  $z_n \sim \mathcal{N}(h(n, \theta), \sigma^2)$ . The likelihood function can be calculated as

$$\Lambda(\theta) = p(z_1, \dots, z_N | \theta) = c \exp \left( -\frac{1}{2\sigma^2} \sum_{n=1}^N [z_n - h(j, \theta)]^2 \right) \quad (2.18)$$

Comparing (2.17) with (2.18) indicates that minimising the quadratic estimation errors is equivalent to maximising the likelihood function. The LS estimator, hence, coincides with the ML estimator [3, 7].

Additionally,  $h(n, \theta)$  can be either a linear or non-linear function of parameter  $\theta$ , corresponding to either a linear or non-linear LS estimation problem, respectively. Because of the linear static models involved in the project, the linear LS problem is described here. Given a linear measurement vector

$$\mathbf{z} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\omega} \quad (2.19)$$

where  $\mathbf{z} \in \mathbb{R}^N$  is the vector of measurements,  $\mathbf{H} \in \mathbb{R}^{N \times M}$  is the measurement matrix,  $\boldsymbol{\theta} \in \mathbb{R}^M$  is the parameter vector and  $\boldsymbol{\omega} \in \mathbb{R}^N$  is the vector of measurement errors. The quadratic error

$$J(\boldsymbol{\theta}) = (\mathbf{z} - \mathbf{H}\boldsymbol{\theta})^\top (\mathbf{z} - \mathbf{H}\boldsymbol{\theta}) \quad (2.20)$$

is minimised by setting the first derivative of  $J(\boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta}$  to 0:

$$\frac{d(J(\boldsymbol{\theta}))}{d\boldsymbol{\theta}} = 2\boldsymbol{\theta}^\top \mathbf{H}^\top \mathbf{H} - 2\mathbf{z}^\top \mathbf{H} = 0 \quad (2.21)$$

The linear LS estimator is then calculated as

$$\hat{\boldsymbol{\theta}}^{LS} = \left( \mathbf{H}^\top \mathbf{H} \right)^{-1} \mathbf{H}^\top \mathbf{z} \quad (2.22)$$

The LS algorithm, furthermore, has a recursion feature such that it can be written in a recursive form, and this leads it to be capable of solving the problems of time series models. For example, the iterative least squares and recursive least squares methods are derived according to the LS estimation and have been applied to the parameter identification for the controlled autoregressive moving average models [166] and output error moving average systems [60].

### 2.2.2 Bayesian parameter estimators

#### Maximum a posteriori estimator

In Bayesian estimation, parameters are viewed as random variables and hence admits distributions. Hence, a prior distribution  $p(\theta)$  for the parameter  $\theta$  is assumed as a Gaussian variable with mean value  $\bar{\theta}$  and variance  $\sigma_0^2$  (i.e.,  $\theta \sim \mathcal{N}(\bar{\theta}, \sigma_0^2)$ ). The maximum a posteriori (MAP) estimator can be derived by evaluating the posterior PDF from the Bayes' rule,

$$p(\theta | \mathcal{Z}) = \frac{p(\mathcal{Z} | \theta)p(\theta)}{p(\mathcal{Z})} \quad (2.23)$$

which involves the likelihood function in (2.9) and where the *posteriori* PDF is,

$$\begin{aligned} p(\theta | \mathcal{Z}) &= \frac{cf_1(\mathcal{Z})f_2(\mathcal{Z}, \theta)p(\theta)}{p(\mathcal{Z})} \\ &= \frac{cf_1(\mathcal{Z})}{p(\mathcal{Z})} \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[ -\frac{1}{2\sigma^2} \left( N\theta^2 - 2\theta \sum_{n=1}^N z_n \right) - \frac{1}{2\sigma_0^2} (\theta - \bar{\theta})^2 \right] \\ &= \frac{cf_1(\mathcal{Z})}{p(\mathcal{Z})} \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[ -\frac{N\sigma_0^2 + \sigma^2}{2\sigma^2\sigma_0^2} \left( \theta^2 - 2\frac{\sigma_0^2 \sum_{n=1}^N z_n + \sigma^2\bar{\theta}}{N\sigma_0^2 + \sigma^2} \theta + \frac{\sigma^2\bar{\theta}}{N\sigma_0^2 + \sigma^2} \right) \right] \\ &= f_3(\mathcal{Z}) \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left[ -\frac{1}{2\sigma_1^2} (\theta - \zeta(\mathcal{Z}))^2 \right] \end{aligned} \quad (2.24)$$

where,

$$\sigma_1^2 = -\frac{\sigma^2\sigma_0^2}{N\sigma_0^2 + \sigma^2} \quad (2.25)$$

$$\zeta(\mathcal{Z}) = \frac{\sigma_0^2 \sum_{n=1}^N z_n + \sigma^2\bar{\theta}}{N\sigma_0^2 + \sigma^2} \quad (2.26)$$

$$f_3(\mathcal{Z}) = \frac{cf_1(\mathcal{Z})\sigma_1}{p(\mathcal{Z})\sigma_0} \exp \left[ -\frac{1}{2\sigma_1^2} \left( -\zeta(\mathcal{Z})^2 + \frac{\sigma^2\bar{\theta}}{N\sigma_0^2 + \sigma^2} \right) \right] \quad (2.27)$$

Hence, the MAP estimator that seeks to find the mode of the *a posteriori* distribution gives,

$$\hat{\theta}^{MAP}(\mathcal{Z}) = \arg \max_{\theta} p(\theta | \mathcal{Z}) \quad (2.28)$$

$$\frac{d(p(\theta | \mathcal{Z}))}{d\theta} = 0 \Rightarrow \hat{\theta}^{MAP} = \zeta(\mathcal{Z}) \quad (2.29)$$

If the prior PDF has  $\sigma_0 \rightarrow \infty$ , MAP estimator is

$$\hat{\theta}^{MAP}(\mathcal{Z}) = \zeta(\mathcal{Z}) = \frac{1}{N} \sum_{n=1}^N z_n = \bar{z} = \hat{\theta}^{ML} \quad (2.30)$$

The assumption of  $\sigma_0 \rightarrow \infty$ , which causes the prior PDF being equivalent to a uniform distribution, makes the MAP estimator coincide with the ML estimator. The prior adds a bias to the maximum likelihood estimate but improves by reducing its variance, which can be useful in solving ill-posed problems [7].

### Minimum mean square error estimator

The minimum mean square error (MMSE) estimator is applied in the Bayesian setting when a prior PDF of parameter is given, and the estimator is computed by minimising the mean squared errors:

$$\hat{\theta}^{MMSE} = \arg \min_{\hat{\theta}} \mathbb{E} [(\theta - \hat{\theta})^2 | \mathcal{Z}] \quad (2.31)$$

The first derivative of the above function is

$$\frac{d\mathbb{E} [(\theta - \hat{\theta})^2 | \mathcal{Z}]}{d\hat{\theta}} = 2\hat{\theta} - 2\mathbb{E}(\theta | \mathcal{Z}) = 0 \quad (2.32)$$

The MMSE estimator, hence, is

$$\hat{\theta}^{MMSE} = \mathbb{E}(\theta | \mathcal{Z}) \quad (2.33)$$

which is the conditional expected value of  $\theta$ , given the observations.

With the assumption of Gaussian distributed  $\theta$ , the MAP estimator equals the mean value of the conditional Gaussian PDF in (2.24). Therefore, the MMSE estimator coincides with the MAP estimator.

As mentioned above, an estimate computed from different approaches with different constraints can be the same under specific cases. The ML estimator, as a non-Bayesian approach, coincides with the Bayesian MAP estimator with a diffused Gaussian prior PDF that indicates the ignorance of prior information.

For the non-Bayesian approaches, the LS estimate of a system with independent and identical zero-mean Gaussian disturbance coincides with the ML estimate. For the Bayesian approaches, the MMSE estimate of a Gaussian variable coincides with the MAP estimate.

In practice, non-Bayesian approaches are commonly applied to the cases lacking reliable prior knowledge; otherwise Bayesian approaches are used. The ML estimator is applicable when large amounts of data are available so that unbiased estimation can be carried out. Compared to the ML estimator, the LS estimator does not require any assumptions about noise distributions and hence needs less parameters to estimate. This results in the data being used more effectively [51]. The derivation of the MAP estimator is completely model-dependent (Section 2.2.2) so that errors can be easily caused by inadequately selected models and distributions. The MMSE estimate given in (2.33) is actually the expected value of the posterior parameter, and it is generally calculated according to Bayes' rule (2.23), instead of being given directly.

### 2.2.3 Discrete-time kinematic models

Kinematic models can be defined directly in discrete time and correspond to physical characters of the described motions. Discrete-time linear stochastic systems can be represented by a discrete-time state-space model,

$$\mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{G}_t \mathbf{u}_t + \mathbf{\Gamma}_t \omega_t \quad (2.34)$$

where  $\mathbf{x}_t$  is the state at time index  $t$ ,  $\mathbf{F}_t$  is the system state matrix and  $\mathbf{G}_t$  is the gain of input,  $\mathbf{u}_t$ , and  $\mathbf{\Gamma}_t$  is the gain of process noise,  $\omega_t$ , which is defined as a sequence with zero-mean and variance  $\sigma_\omega^2$ . The covariance of process noise then can be calculated as

$$\mathbf{Q}_t = \mathbb{E}[\mathbf{\Gamma}_t \omega_t \omega_t \mathbf{\Gamma}_t^\top] = \mathbf{\Gamma}_t \sigma_\omega^2 \mathbf{\Gamma}_t^\top \quad (2.35)$$

where,

$$\mathbb{E}(\omega_t \omega_j) = \sigma_\omega^2 \delta_{tj} \quad (2.36)$$

The discrete-time measurement  $\mathbf{z}_t$  is given in (2.37), which is represented by a measurement matrix  $\mathbf{H}_t$  and a zero-mean measurement noise  $\mathbf{v}_t$  of covariance  $\mathbf{R}_t$ .

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t \quad (2.37)$$

The measurement and the process noise sequences are assumed to be uncorrelated.

According to (2.34), a kinematic system without an input sequence is directly defined in (2.38) with a zero-mean white noise sequence  $\omega_t$ .

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{\Gamma}\omega_t. \quad (2.38)$$

In practice, kinematic models are derived from Newton's laws of motion with the assumption of constant velocity or constant acceleration. These models are known as discrete white noise acceleration (DWNA) and discrete Wiener process acceleration (DWPA) models and are related to second-order and third-order models, respectively [7].

### Discrete white noise acceleration model

The DWNA model is defined by assuming that the object is moving with constant acceleration  $a_t$  from sampling time  $t\Delta T$  to  $(t+1)\Delta T$ , where  $\Delta T$  is the sampling period, so that it is also known as constant velocity (CV) model. According to this assumption, velocity  $v_t$  has following relations:

$$v_{t+1} = v_t + a_t\Delta T \quad (2.39)$$

and the position is

$$p_{t+1} = p_t + \frac{(a_t + a_{t+1})\Delta T}{2} = p_t + v_t\Delta T + a_t\frac{\Delta T^2}{2} \quad (2.40)$$

Therefore, the second-order kinematic model is constructed as follows:

$$\begin{bmatrix} p_{t+1} \\ v_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_t \\ v_t \end{bmatrix} + \begin{bmatrix} \frac{\Delta T^2}{2} \\ \Delta T \end{bmatrix} a_t \quad (2.41)$$

Compared to (2.38), the state  $\mathbf{x}_t = \begin{bmatrix} p_t & v_t \end{bmatrix}^\top$  and process noise is the assumed acceleration (i.e.,  $\omega_t = a_t$ ). The state matrix  $\mathbf{F}$  and noise gain matrix  $\mathbf{\Gamma}$  are given as follows:

$$\mathbf{F} = \begin{bmatrix} 1 & \Delta T \\ 0 & 1 \end{bmatrix} \quad \mathbf{\Gamma} = \begin{bmatrix} \frac{\Delta T^2}{2} \\ \Delta T \end{bmatrix} \quad (2.42)$$

The covariance of the process noise, according to (2.35), is then computed as:

$$\mathbf{Q} = \mathbf{\Gamma}\sigma_\omega^2\mathbf{\Gamma}^\top = \begin{bmatrix} \frac{\Delta T^4}{4} & \frac{\Delta T^3}{2} \\ \frac{\Delta T^3}{2} & \Delta T^2 \end{bmatrix} \sigma_\omega^2 \quad (2.43)$$

where,  $\sigma_\omega$  has the same physical dimensions as acceleration. Practically,  $\sigma_\omega$  is set up in the range of  $[0.5a_M, a_M]$ , where  $a_M$  is the maximum acceleration [7].

### Discrete Wiener process acceleration model

The DWPA model is also known as constant acceleration (CA) model. The model is constructed based on the assumption that the acceleration increases from sampling time  $t\Delta T$  to  $(t+1)\Delta T$  with zero-mean white noise sequence  $\omega_k$ . The model is then constructed as follows:

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{\Gamma}\omega_t \quad (2.44)$$

where

$$\mathbf{x}_t = \begin{bmatrix} p_t & v_t & a_t \end{bmatrix}^\top \quad (2.45)$$

$$\mathbf{F} = \begin{bmatrix} 1 & \Delta T & \frac{\Delta T^2}{2} \\ 0 & 1 & \Delta T \\ 0 & 0 & 1 \end{bmatrix} \quad (2.46)$$

$$\mathbf{\Gamma} = \begin{bmatrix} \frac{\Delta T^2}{2} \\ \Delta T \\ 1 \end{bmatrix} \quad (2.47)$$

The covariance of noise term  $\mathbf{\Gamma}\omega_t$  is computed as shown in (2.48). In applications,  $\sigma_\omega$  is generally set up in the range of  $[0.5\Delta a_M, \Delta a_M]$ , where  $\Delta a_M$  is the maximum value of acceleration increments during sampling interval.

$$\mathbf{Q} = \mathbf{\Gamma}\sigma_\omega^2\mathbf{\Gamma}^\top = \begin{bmatrix} \frac{\Delta T^4}{4} & \frac{\Delta T^3}{2} & \frac{\Delta T^2}{2} \\ \frac{\Delta T^3}{2} & \Delta T^2 & \Delta T \\ \frac{\Delta T^2}{2} & \Delta T & 1 \end{bmatrix} \sigma_\omega^2 \quad (2.48)$$

The direct discrete kinematic models became popular because of their links to the actual physical processes (i.e., position, velocity and acceleration), especially for fixed sampling intervals. The second-order and the third-order models can be applied as either isolated or combined in the same system. For example, in the prediction of pedestrian path, multiple models, such as CV, CA, constant turn (CT) models, are applied in the same system to predict pedestrian position, velocity, acceleration and turning, from the image sequences [134].

### 2.2.4 State estimation in discrete-time dynamic systems

A discrete-time dynamic system is described as follows:

$$\mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{G}_t \mathbf{u}_t + \mathbf{\Gamma}_t \omega_t \quad (2.49)$$

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t \quad (2.50)$$

where  $\mathbf{x}_t$ ,  $\mathbf{u}_t$ ,  $\mathbf{\Gamma}_t$  and  $\omega_t$  denote the vectors of system state, input, gain of process noise and the noise, respectively.  $\mathbf{z}_t$  and  $\mathbf{v}_t$  are the vectors of measurement and measurement noise. This dynamic system is assumed to have following properties:

- The process noise satisfies  $\omega \sim \mathcal{N}(0, \sigma_\omega)$ , and therefore  $\mathbf{\Gamma}\omega_t$  is zero-mean Gaussian variable with covariance  $\mathbf{Q}_t$ , derived in (2.35).
- The measurement noise  $\mathbf{v}_t$  is zero-mean Gaussian variable with covariance  $\mathbf{R}_t$ .
- The noise  $\omega_t$  and  $\mathbf{v}_t$  are mutually independent.

The fundamental equations of linear estimation are derived from the posterior PDF  $p(\mathbf{x} | \mathbf{z})$  of a static system, where  $\mathbf{x}$  and  $\mathbf{z}$  are assumed to be jointly Gaussian [7]. The equations include a conditional mean and the corresponding covariance:

$$\hat{\mathbf{x}} \triangleq \mathbb{E}(\mathbf{x} | \mathbf{z}) = \bar{\mathbf{x}} + \mathbf{P}_{\mathbf{xz}} \mathbf{P}_{\mathbf{zz}}^{-1} (\mathbf{z} - \bar{\mathbf{z}}) \quad (2.51)$$

$$\mathbf{P}_{\mathbf{xx}|\mathbf{z}} \triangleq \text{cov}(\mathbf{x} | \mathbf{z}) = \mathbf{P}_{\mathbf{xx}} - \mathbf{P}_{\mathbf{xz}} \mathbf{P}_{\mathbf{zz}}^{-1} \mathbf{P}_{\mathbf{zx}} \quad (2.52)$$

where  $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P}_{\mathbf{xx}})$ ,  $\mathbf{z} \sim \mathcal{N}(\bar{\mathbf{z}}, \mathbf{P}_{\mathbf{zz}})$  and  $\mathbf{P}_{\mathbf{xz}} = \mathbb{E}[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{z} - \bar{\mathbf{z}})^\top] = \mathbf{P}_{\mathbf{zx}}^\top$

#### Kalman filter

The dynamic system described in (2.49) and (2.50) can be regarded as recursive linear static estimation such that the state at time  $t + 1$  is estimated from previous moment  $t$ . The  $\bar{\mathbf{x}}$  is substituted by  $\hat{\mathbf{x}}_{t+1|t}$  that can be calculated through

$$\begin{aligned} \hat{\mathbf{x}}_{t+1|t} &= \mathbb{E}(\mathbf{x}_{t+1} | \mathcal{Z}^t) = \mathbb{E}(\mathbf{F}_t \mathbf{x}_t + \mathbf{G}_t \mathbf{u}_t + \mathbf{\Gamma}_t \omega_t | \mathcal{Z}^t) \\ &\Rightarrow \hat{\mathbf{x}}_{t+1|t} = \mathbf{F}_t \hat{\mathbf{x}}_{t|t} + \mathbf{G}_t \mathbf{u}_t \end{aligned} \quad (2.53)$$

where  $\mathcal{Z}^t = \{\mathbf{z}_1, \dots, \mathbf{z}_t\}$ . The prediction error and the corresponding covariance are

$$\tilde{\mathbf{x}}_{t+1|t} = \mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t} = \mathbf{F}_t \tilde{\mathbf{x}}_{t|t} + \mathbf{\Gamma}_t \omega_t \quad (2.54)$$

$$\begin{aligned}\mathbb{E}(\tilde{\mathbf{x}}_{t+1|t}\tilde{\mathbf{x}}_{t+1|t}^T | \mathcal{Z}^t) &= \mathbf{F}_t\mathbb{E}(\tilde{\mathbf{x}}_{t|t}\tilde{\mathbf{x}}_{t|t}^T)\mathbf{F}_t^T + \mathbb{E}(\mathbf{\Gamma}_t\omega_t\omega_t^T\mathbf{\Gamma}_t^T) \\ \Rightarrow \mathbf{P}_{t+1|t} &= \mathbf{F}_t\mathbf{P}_{t|t}\mathbf{F}_t^T + \mathbf{Q}_t\end{aligned}\quad (2.55)$$

The predicted measurement  $\bar{\mathbf{z}}$  is then substituted by  $\hat{\mathbf{z}}_{t+1|t}$ ,

$$\begin{aligned}\mathbb{E}(\mathbf{z}_{t+1} | \mathcal{Z}^t) &= \mathbb{E}(\mathbf{H}_{t+1}\mathbf{x}_{t+1} + \mathbf{v}_t | \mathcal{Z}^t) \\ \Rightarrow \hat{\mathbf{z}}_{t+1|t} &= \mathbf{H}_{t+1}\hat{\mathbf{x}}_{t+1|t}\end{aligned}\quad (2.56)$$

The measurement prediction error and the covariance that is the substitute of  $\mathbf{P}_{\mathbf{x}\mathbf{x}}$  are

$$\tilde{\mathbf{z}}_{t+1|t} = \mathbf{H}_{t+1}\tilde{\mathbf{x}}_{t+1|t} + \mathbf{v}_{t+1}\quad (2.57)$$

$$\mathbf{S}_{t+1} = \mathbf{H}_{t+1}\mathbf{P}_{t+1|t}\mathbf{H}_{t+1}^T + \mathbf{R}_{t+1}\quad (2.58)$$

The substitution of  $\mathbf{P}_{\mathbf{x}\mathbf{z}}$  is the covariance between state and measurement that is

$$\begin{aligned}\mathbb{E}(\tilde{\mathbf{x}}_{t+1|t}\tilde{\mathbf{z}}_{t+1|t}^T | \mathcal{Z}^t) &= \mathbb{E}(\tilde{\mathbf{x}}_{t+1|t}[\mathbf{H}_{t+1}\tilde{\mathbf{x}}_{t+1|t} + \mathbf{v}_{t+1}]^T | \mathcal{Z}^t) \\ &= \mathbf{P}_{t+1|t}\mathbf{H}_{t+1}^T\end{aligned}\quad (2.59)$$

Subsequently, according to (2.51), the estimated gain is

$$\mathbf{K}_{t+1} = \mathbb{E}(\tilde{\mathbf{x}}_{t+1|t}\tilde{\mathbf{z}}_{t+1|t}^T | \mathcal{Z}^t)\mathbf{S}_{t+1}^{-1} = \mathbf{P}_{t+1|t}\mathbf{H}_{t+1}^T\mathbf{S}_{t+1}^{-1}\quad (2.60)$$

In dynamic systems, the estimated state  $\hat{\mathbf{x}}_{t+1|t+1}$  and covariance  $\mathbf{P}_{t+1|t+1}$ , according to (2.51) and (2.52), are represented in the form

$$\hat{\mathbf{x}}_{t+1|t+1} = \hat{\mathbf{x}}_{t+1|t} + \mathbf{K}_{t+1}(\tilde{\mathbf{z}}_{t+1|t})\quad (2.61)$$

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|t} - \mathbf{K}_{t+1}\mathbf{H}_{t+1}\mathbf{P}_{t+1|t} = \mathbf{P}_{t+1|t} - \mathbf{K}_{t+1}\mathbf{S}_{t+1}\mathbf{K}_{t+1}^T\quad (2.62)$$

The above linear dynamic state estimation is the KF, whose applications include the evolution of state and measurement, and it is overviewed in Figure 2.5 [7].

If the system is represented as a non-linear model, given in (2.63) and (2.64), the KF is extended to the non-linear estimation. For example, the extended Kalman filter (EKF) follows the same process as the KF and linearises the non-linear function  $f$  and  $h$  by Jacobian matrices and current estimation to perform the covariance computation [43, 72]. Since the non-linearity becomes higher, the Jacobian matrices calculation becomes challenging and the large errors generated from linearisation



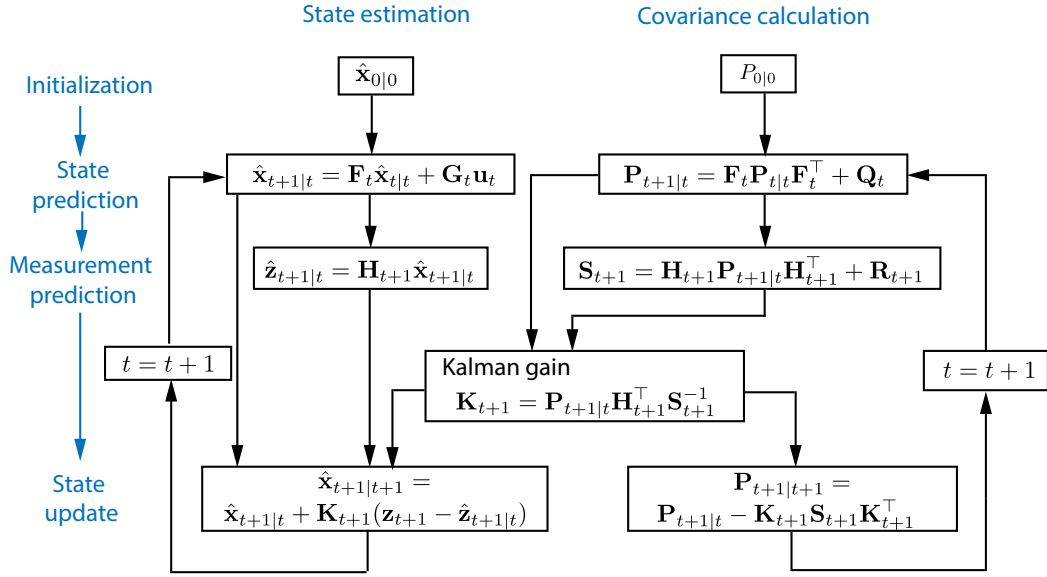


Figure 2.5: Flow diagram of the KF

results in the poor performance of EKF.

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t) + \boldsymbol{\omega}_t \quad (2.63)$$

$$\mathbf{z}_t = h(\mathbf{x}_t) + \mathbf{v}_t \quad (2.64)$$

The unscented Kalman filter (UKF) applies an unscented transform to generate the samples of the same mean and variance with  $\mathbf{x}_{t|t}$ , then recovers the estimate mean and covariance by propagating these samples through the non-linear transition and measurement functions. The UKF has been demonstrated to be able to provide more accurate estimates than EKF for highly non-linear cases [43, 70, 71]. In this thesis, application of the non-linear estimation was not required.

### Interacting multiple model estimators

When a system's behaviour cannot be characterised by a single model, multiple models can be adapted appropriately with respect to the observations to obtain a better system description; this is called the multiple model approach.

The interacting multiple model (IMM) algorithm is able to estimate states of a system including different dynamic modes. It runs filters under finite number of possible models in parallel and combines the results according to the previous models for the state estimation [20, 98].

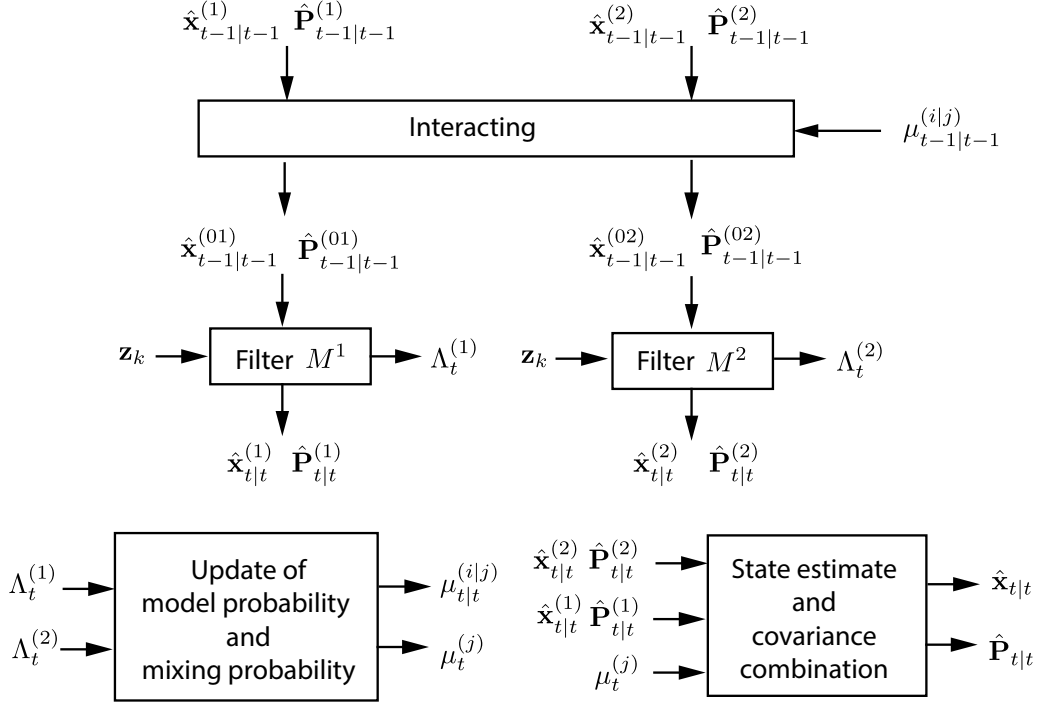


Figure 2.6: The algorithm of the IMM estimator [7].

An IMM estimator of a dynamic system of two models ( $M^{(i)}$   $i = 1, 2$ ) is demonstrated in Figure 2.6. The previous estimate  $\hat{\mathbf{x}}_{t-1|t-1}^{(i)}$  and covariance  $\mathbf{P}_{t-1|t-1}^{(i)}$  are selected as the input. Firstly, mixed initial condition  $\hat{\mathbf{x}}_{t-1|t-1}^{(0i)}$  and covariance  $\mathbf{P}_{t-1|t-1}^{(0i)}$  are calculated by applying mixing probabilities  $\mu_{t-1|t-1}^{(i|j)}$  to the input. Secondly, the filter is matched to different models and run in parallel to calculate the model-conditioned estimates  $\hat{\mathbf{x}}_{t|t}^{(j)}$ , covariances  $\mathbf{P}_{t|t}^{(j)}$  and likelihood functions  $\Lambda_t^{(j)}$  by using the mixed initial condition and observations  $\mathbf{z}_t$ . The mixing probabilities  $\mu_t^{(j)}$  are then updated by combining the likelihood functions  $\Lambda_t^{(j)}$ . The estimate and covariance are updated as a mixture of the model-conditioned estimates and covariance [7].

The modular structure of the IMM estimator makes it possible to be applied to either linear or non-linear model by setting up KF or EKF in the corresponding behaviour modes. In practice, IMM has been widely applied to the target tracking systems, such as the Global Positioning System (GPS) [68], air traffic management system [165] and human motions tracking [87].

### 2.2.5 Hidden Markov models and expectation maximisation

Given a time sequence  $\{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ , the joint distribution is

$$p(\mathbf{x}_1, \dots, \mathbf{x}_T) = \prod_{t=1}^T p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}) \quad (2.65)$$

Assuming that the conditional distribution,  $p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1})$ , is dependent on the most recent observation, a first-order Markov chain is constructed as the diagram in Figure 2.7a and the joint distribution for the data sequence is

$$p(\mathbf{x}_1, \dots, \mathbf{x}_T) = p(\mathbf{x}_1) \prod_{t=2}^T p(\mathbf{x}_t | \mathbf{x}_{t-1}) \quad (2.66)$$

According to (2.65) and (2.66), the conditional distribution of  $\mathbf{x}_t$  given all previous observations is

$$p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}) = p(\mathbf{x}_t | \mathbf{x}_{t-1}) \quad (2.67)$$

which indicates that the predicted distribution of  $\mathbf{x}_t$  depends only on  $\mathbf{x}_{t-1}$ .

In order to not be restricted to the Markov assumption of the model orders, a latent variable  $\mathbf{x}_t$  is introduced to control the corresponding observation  $\mathbf{z}_t$ . Meanwhile,  $\mathbf{x}_t$  is assumed to be a first-order Markov chain. The joint distribution for the observations and latent variables is given by

$$p(\mathbf{z}_1, \dots, \mathbf{z}_T, \mathbf{x}_1, \dots, \mathbf{x}_T) = p(\mathbf{x}_1) \left[ \prod_{t=2}^T p(\mathbf{x}_t | \mathbf{x}_{t-1}) \right] \prod_{t=1}^T p(\mathbf{z}_t | \mathbf{x}_t) \quad (2.68)$$

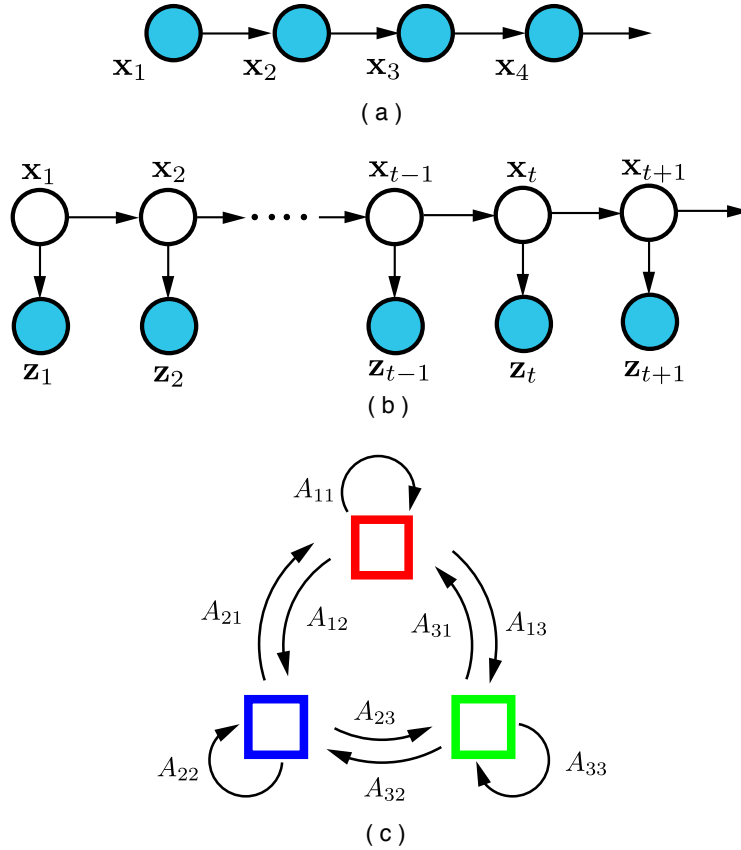
which indicates that any two observations  $\mathbf{z}_{t_1}$  and  $\mathbf{z}_{t_2}$  are connected through latent variables (see Figure 2.7b). Therefore, the prediction of  $\mathbf{z}_{t+1}$ , depends on all the previous observations. The model with discrete latent variables in a Markov chain is a HMM.

With respect to a HMM with binary latent variable of  $K$  discrete states, the conditional distribution  $p(\mathbf{x}_t | \mathbf{x}_{t-1})$  corresponds to a transition matrix  $\mathbf{A} \in \mathbb{R}^{K \times K}$ , where the elements represent the transition probabilities from the state at time  $t-1$  to the state at  $t$ :

$$A_{jk} = p(x_{t,k} = 1 | x_{t-1,j} = 1) \quad (2.69)$$

where  $j$  and  $k$  are the indices of row and column of matrix  $\mathbf{A}$  and  $x_{t,k}$  denotes the state  $k$  of latent variable at  $t$ . The transition probabilities satisfy

$$\sum_{k=1}^K A_{jk} = 1 \quad 0 \leq A_{jk} \leq 1. \quad (2.70)$$



**Figure 2.7: Diagram of HMM [17].** (a) A first-order Markov chain, where  $\{x_1, x_2, \dots, x_T\}$  is a set of observations. (b) A model of sequential data, of which each observation  $z_t$  is conditioned on the corresponding latent variables  $x_t$ . At the same time, the latent variables form a Markov chain. (c) Transition diagram of a HMM with three states, i.e.,  $K = 3$ . The red, green and blue squares represent three different states of latent variables and the black arrows indicate the transition matrix, where  $A_{ik}$  indicates the transition probability of latent variables changing from state  $i$  to  $k$  from moment  $t - 1$  to  $t$ .

and can be depicted in a transition diagram, where the states of latent variables are represented as nodes, shown in Figure 2.7c.

The conditional distribution is then given by

$$p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{A}) = \prod_{k=1}^K \prod_{j=1}^K A_{jk} \quad (2.71)$$

When  $t = 1$ , a marginal distribution is

$$p(\mathbf{x}_1 | \boldsymbol{\pi}) = \prod_{k=1}^K \pi_k \quad (2.72)$$

where  $\pi_k = p(x_{1,k})$  and  $\sum_{k=1}^K \pi_k = 1$ . With respect to the probability model, emission probabilities are defined as  $p(\mathbf{z}_t | \mathbf{x}_t, \boldsymbol{\theta}_p)$ , where  $\boldsymbol{\theta}_p$  is the set of parameters of the distributions. Additionally, the joint probability for the observations and latent variables is

$$p(\mathcal{Z}, \mathcal{X} | \boldsymbol{\theta}) = p(\mathbf{x}_1 | \boldsymbol{\pi}) \left[ \prod_{t=2}^T p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{A}) \right] \prod_{m=1}^T p(\mathbf{z}_m | \mathbf{x}_m, \boldsymbol{\theta}_p) \quad (2.73)$$

where  $\mathcal{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_T\}$  is the set of observations,  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$  is the set of latent variables and  $\boldsymbol{\theta} = \{K, \mathbf{A}, \boldsymbol{\theta}_p\}$  is the set of parameters of the HMM [17, 101].

The parameters involved in the HMM are commonly identified using the expectation maximisation (EM) algorithm, which identifies the parameters by maximising a likelihood function. The EM algorithm is demonstrated in Section 4.3.1 and the general steps are summarised as follows [16, 17, 105]:

- Initialise the model parameters,  $\boldsymbol{\theta}^{old}$  randomly or according to the prior knowledge about the system.
- Calculate the posterior distribution,  $p(\mathcal{X} | \mathcal{Z}, \boldsymbol{\theta}^{old})$ , using the current parameters, so as to evaluate the expectation of the log likelihood of complete-data,  $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old})$  (E-step).
- Calculate the new parameters,  $\boldsymbol{\theta}^{new}$ , by maximising the function  $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old})$  (M-step).
- If the log likelihood is not converged, the parameters are updated by setting  $\boldsymbol{\theta}^{old} \leftarrow \boldsymbol{\theta}^{new}$  and the algorithm goes back to E-step.

In summary, HMMs model the sequence data with hidden states that are assumed to be a Markov process. They have been extensively used in temporal data analysis, such as speech recognition, precipitation modelling [113] and biological sequence analysis [175].

### 2.2.6 Smoothing for discrete-time state-space models

State smoothing applies data interval  $\mathcal{Z}^j$  up to future moment  $j$  to estimate states at  $j$  and it is defined by,

$$\hat{\mathbf{x}}_{t|j} = E[\mathbf{x}_t | \mathcal{Z}^j], \quad j > t \quad (2.74)$$

According to the relationship between  $j$  and  $t$ , the state smoothing is classified into the following types [7, 23]:

- Fixed-interval smoothing: The measurement data interval of fixed length is applied to estimate the interior states (i.e.,  $j = T, t = 1, 2, \dots T$ ).
- Fixed-point smoothing: It is an estimation of the state at fixed moments by adding the measurement ahead iteratively (i.e.,  $j = t + 1, t + 2, \dots$  with fixed value of  $j$ ).
- Fixed-lag smoothing: In this case, the state is estimated along with the time by using the measurement interval with the fixed time lead  $\tau$  (i.e.,  $t = 0, 1, 2, \dots$  and  $j = t + \tau$ ).

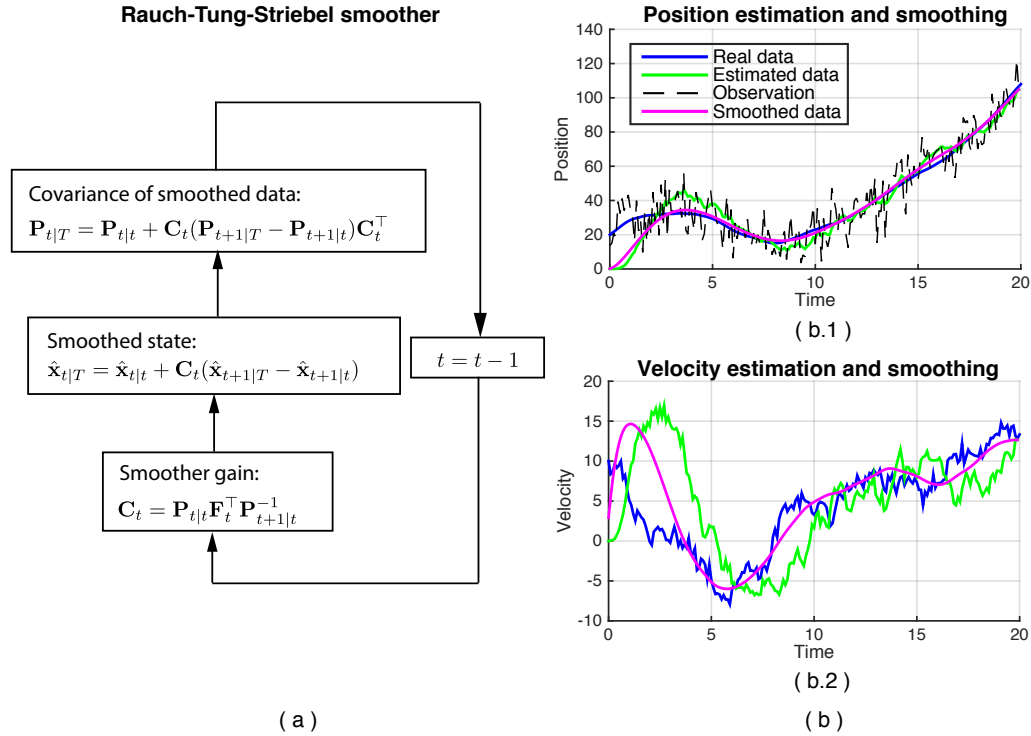
Fixed-interval smoothing is commonly applied along with the corresponding filter for the processing of noisy measurements off-line. For example, the Rauch-Tung-Striebel (RTS) smoother can work backwards after the KF estimator for linear Gaussian dynamic models. The extended RTS smoother cooperates with EKF for non-linear models [133]. A RTS smoother consists of a calculation of smoothing gain, states smoothing and the update of the smoothed state's covariance (see Figure 2.8a). It is initialised by the last estimated states after the KF estimation and iterates backward to the first moment states. A cooperative application of forward running KF and backward recursion RTS is demonstrated on a DWNA model, which was introduced in Section 2.2.3 and the results are shown in Figure 2.8b. It is seen that the smoother significantly smoothes the errors introduced by the random initialisation. Explicitly, the smoothed position is seen to more quickly, less noisily and accurately follow the real situations, as does the velocity.

## 2.3 Methods for motility and morphology analysis

### 2.3.1 Overview of object tracking

Object tracking is the process to regenerate the spatial and temporal trajectory of an object and it is widely applied in vehicle navigation, visual surveillance and the study of unfamiliar objects' behaviours. Numerous tracking approaches have been proposed and two issues, in general, are addressed during the tracking process:

- A suitable object representation for tracking.
- Tracking methods selection according to the object representation.



**Figure 2.8: The demonstration of a RTS smoother.** (a) The diagram of a RTS smoother; (b) State estimation and smoothing of a DWNA model by KF and a RTS smoother where the position and velocity estimation and smoothing are given in subfigures (b.1) and (b.2), respectively.

The object representation is deemed to capture the object's spatial features, which is the basis of temporal analysis. The representation is required to be easily detected and efficiently represented. According to the object's geometric properties, the commonly used representations are of the following three types [173]:

- In the scenarios of tracking of small-sized objects or a large number of objects, the objects are generally represented by points. For example, the localisation of a specific kind of protein is often tracked to analyse its functions [21]. Kadiramanathan et al. studied the neutrophil chemotaxis patterns in vivo through the estimation of the multiple cell tracks using a cell centroids tracking approach [73].
- Simple geometric shapes such as ellipses and rectangles are used to represent rigid objects. For example, a rectangular tracking window was applied to handle visual tracking and recognition problems [177].
- Contours, a representation of object boundaries, are used for non-rigid ob-

jects tracking; for instance, human motion tracking from the dance videos or the tracking of cell divisions, where the target object presents irregular shapes [34, 147].

The object representations mentioned above correspond to different tracking methods [69]:

- Points tracking problems can be described as the frame-to-frame correspondence identification of points that represent objects. In respect of the assumptions of point correspondence, tracking methods are separated into deterministic and statistical methods. Deterministic methods define a correspondence cost using different motion constraints, such as the assumptions of limited location and velocity changes across consecutive frames [65]. Statistical methods, on the other hand, solve object correspondence by employing the measurement of object state with the consideration of model uncertainty; for example, performing the tracking of a single object state, which is assumed to be Gaussian distributed, of a linear system using a KF [69].
- Kernel tracking, which refers to the problem of tracking objects represented by geometric shapes, is performed by describing the motion of the object frame-to-frame through parametric geometric transformations, such as translation, rotation and affine transformation. Tracking can be carried out based on either single or multiview appearance models. For example, the mean-shift tracking algorithm estimates the object translocation through maximising the similarity of colour histograms of predefined object regions between consecutive frames [35]. In addition, Yu et al. proposed an approach that uses mean-shift tracking of the weighted spatial histogram of the background combining multi-scale models of objects to solve the problem of scale-changed object tracking [176].
- Silhouette-tracking algorithms describe object shape accurately in every frame using object edges, colour histograms or contours. These tracking methods are classified into two categories: (i) shape matching, and (ii) contour tracking. Shape matching identifies an object by maximising the similarity between the current frame and the constructed shape model. Huttenlocher et al. first proposed shape recognition from rigid motions using edge-based representation and Hausdorff distance between the object model and the images. The Hausdorff distance is a correlation surface, where the minimum is the location of the object [62]. Unlike shape matching, contour tracking uses



an energy function to evolve object contour from the previous frame [12]. For example, Zimmer et al. proposed an energy function that integrated the repulsive forces and 'topological operators' into a classical 'snake' model to perform tracking of contacting cells and cell divisions [185, 186].

In summary, with regard to the particular representation of objects, tracking methods are divided into three categories: points tracking, kernel tracking and silhouette tracking. Approaches in points tracking segment the objects from the frames first and link up the detected point features across the frames, separately [172]. Whereas, the kernel and silhouette tracking approaches perform object detection and correspondence establishment jointly during the estimation of object statement from the previous frame. Compared to the tracking of the object regions, point tracking is more straightforward in object detection but sacrifices the appearance characteristics of the objects. Kernel tracking keeps the information about object appearance using geometric shape representation, but which is only applicable to rigid object tracking. With respect to silhouette tracking approaches, object appearance, including rigid, non-rigid shapes and even topological changes, are mostly tracked. However, the complexity of shape descriptions leads to a restriction on the number of tracked objects [172, 186].

The object-tracking approaches have been widely applied to different fields, such as the video-based traffic monitoring, human-computer interaction and ultrastructural investigation under the electron microscope. A challenge in object tracking is to extend the traditional approaches to satisfy the different tracking requirements of a particular scenerio. One popular solution is to involve additional sources of object information, such as prior knowledge and contextual information. Jo et al. introduced a synergistic framework for integrated vehicle and lane tracking modules, in which both tracking performances were improved by utilising contextual information [68]. The road tracking was improved by using the vehicle tracking results to reduce the errors, brought by the vehicles on the road, during road localisation. At the same time, the lane models were able to add geometric constraints of vehicle models that improved vehicle detection. Papenmeier et al. proposed an approach that uses the surface features to recover the target objects lost, caused by spatiotemporal discontinuities in the multiple-object tracking [114].

In terms of microscopy image sequences of breast cancer cells, the deformable cell shapes are of interest to track, in addition to cell trajectories and correspondence over frames. Therefore, point tracking integrated with shape features is a potential direction for a tracking solution.

### 2.3.2 Image segmentation

Segmentation, which subdivides images into constituent regions, has been widely used to support the development of medical imaging, object detections, traffic monitoring and machine vision. In general, image segmentation is deemed to be an essential starting step to ensure the subsequent process and analysis is carried out successfully [119, 141]. According to the features of identified constituent region, image segmentation can be approached through pixel-based, edge-based or region-based methods [136].

#### Intensity thresholding

Intensity thresholding is the most straightforward pixel-based segmentation approach that segments the object regions using the image intensity, given by

$$I_b(i_x, i_y) = \begin{cases} 1 & \text{if } I(i_x, i_y) > h^* \\ 0 & \text{otherwise} \end{cases} \quad (2.75)$$

where  $I(i_x, i_y)$  denotes the image intensity at coordinate  $(i_x, i_y)$ .  $I_b$  is the resultant binary image from the grey-scale image,  $I$ , with threshold  $h^*$  [141].

For an automated image segmentation, threshold,  $h$ , needs to be calculated automatically [107]. Ostu's method [112] was proposed for automated calculation of  $h$  by maximising the inter-class variances. The method is performed as follows [2, 107, 154]:

- Assuming that  $h \in [1, H]$  is the grey level and  $N_{pix}$  is the number of pixels of the image, the probability of pixels occurring at grey level  $h$  is

$$p(h) = \frac{N_{pix}^{(h)}}{N_{pix}} \quad (2.76)$$

where  $N_{pix}^{(h)}$  denotes the number of pixels occurring at level  $h$ .

- The image pixels are separated by the grey level  $h$  into two classes:  $C_1$  and  $C_2$ . The grey levels from 1 to  $h$  are included in class  $C_1$  and other levels from  $h + 1$  to  $H$  are included in class  $C_2$ . The class probabilities are calculated by

$$p_1(h) = \sum_{i=1}^h p(i) \quad p_2(h) = \sum_{i=h+1}^H p(i) \quad (2.77)$$

where  $p_1(h)$  and  $p_2(h)$  have the relation:

$$p_1(h) + p_2(h) = 1 \quad (2.78)$$

- With respect to each class, the mean value is calculated by

$$\mu_1(h) = \frac{\sum_{i=1}^h p(i)i}{p_1(h)} \quad \mu_2(h) = \frac{\sum_{i=h+1}^H p(i)i}{p_2(h)} \quad (2.79)$$

and the mean of the entire data set is

$$\mu_H = \sum_{i=1}^H p(i)i \quad (2.80)$$

The variance of each class is

$$\sigma_1^2(h) = (\mu_1(h) - \mu_H)^2 \quad \sigma_2^2(h) = (\mu_2(h) - \mu_H)^2 \quad (2.81)$$

The inter-class covariance,  $\sigma_{1,2}(h)$ , is calculated, as follows

$$\sigma_{1,2}(h) = p_1(h)\sigma_1^2(h) + p_2(h)\sigma_2^2(h) \quad (2.82)$$

- Start with  $h = 1$ , perform the above steps iteratively until  $h = H$ . The threshold is then set to  $h^*$ , where  $\sigma_{1,2}(h)$  reaches the maximum.

$$h^* = \arg \max_{1 \leq h \leq H} \sigma_{1,2}(h) \quad (2.83)$$

The common drawbacks of thresholding segmentation is its sensitivity to illumination variations and the disregard of connections between pixels. Any uneven illumination can cause contrast differences across different parts in the same image. Lack of consideration of the neighbourhood of pixels leads to the segmented objects to be discontinuous. For complex image segmentation problems, more post-processing or other approaches are then needed to obtain more accurate segmentation.

### Canny edge detector

With respect to edge-based segmentation, three general criteria are proposed as follows [103]:

- For detection, the real edges should be detected as much as possible while

faulty detections, where the no non-edge points are detected as edge, are minimised.

- The location differences between the detected and real edges are minimised.
- A real edge should correspond to only one direction.

Considering these above criteria, the Canny edge detector [28] is one of the most reliable detection approaches and it is performed by the following steps [66, 141]:

- The original image is smoothed by a Gaussian filter to reduce the effects of image noise.
- The intensity gradients of the image in the horizontal,  $\Delta I_x(i_x, i_y)$ , and vertical directions,  $\Delta I_y(i_x, i_y)$ , are calculated to measure the edge, where the strength,  $\Delta I(i_x, i_y)$ , and direction,  $\phi(i_x, i_y)$ , are given,

$$\Delta I(i_x, i_y) = \sqrt{\Delta I_x(i_x, i_y)^2 + \Delta I_y(i_x, i_y)^2} \quad (2.84)$$

$$\phi(i_x, i_y) = \tan^{-1} \frac{\Delta I_y(i_x, i_y)}{\Delta I_x(i_x, i_y)} \quad (2.85)$$

- Four angles are defined as:  $0^\circ$ ,  $90^\circ$ ,  $45^\circ$  and  $135^\circ$ . The edge directions are digitised to their closest angle.
- The edges extracted from the image intensity gradient may be still blurred; non-maximum suppression, therefore, is performed. At every pixel, the pixel value is set to zero, if its strength is smaller than the strength of neighbours at both sides of the gradient direction.
- After the non-maximum suppression, image edges are finally clarified by hysteresis. Two distinct thresholds  $h_H$  and  $h_L$  ( $h_H > h_L$ ) are selected empirically. If the edge strength,  $\Delta I(i_x, i_y)$ , is smaller than  $h_L$ , the pixel is rejected from image edges. If  $\Delta I(i_x, i_y) > h_H$ , the pixel is marked as on the strong edges. Furthermore, if  $h_L < \Delta I(i_x, i_y) < h_H$ , the pixel is identified exclusive of the edges unless the pixel connects to the strong edges.

The downside of Canny edge detectors is the application of Gaussian filter, which can smooth the noise, as well as the edges. The information lost from edges may result in faulty detection. In the last step, the empirically selected thresholds need to be adjusted with respect to different images. Hence, more operations and segmentation approaches are proposed later for further improvement [157].

### Watershed algorithm

Region-based segmentation directly returns an object region, which is a group of pixels corresponding to the same object. The segmentation has the following formulations [106]:

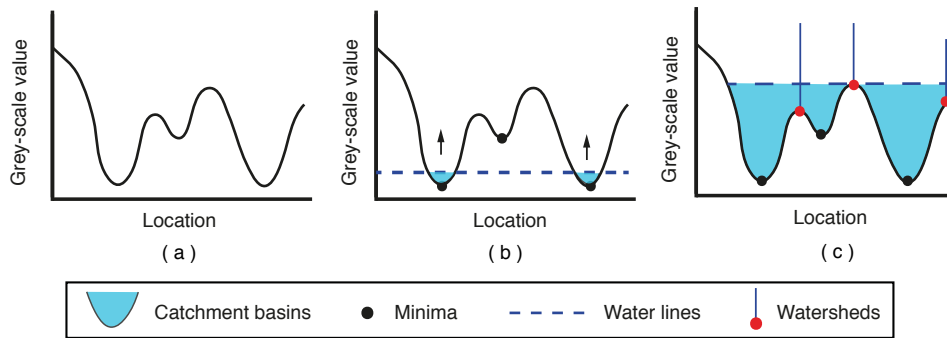
- Every pixel in the image should be assigned to a region.
- In the same region, the pixels should have a common property.
- Different regions have their own properties that differ from others.
- A region must not connect to any other regions.

Beucher and Lantuejoul proposed the watershed algorithm, which has been commonly used to process segmentation of touching objects [15]. A grey-scale image is regarded as 'Landscape' with grey-scale levels as altitude. A 1-D landscape is shown in Figure 2.9a. The watershed segmentation is carried out by the following steps:

- Assuming that water gradually floods into the landscape, the water firstly enters the local minima and generates a water line (see Figure 2.9b).
- The water line starts rising and is disconnected by the landscape. The valleys that hold water (see Figure 2.9b) are called catchment basins, which in an image, are the groups of pixels that correspond to the same region.
- When the water line reaches the local maximum, the water can flow into adjacent catchment basins. These local maxima are defined as watersheds. The detected pixels at watersheds are adjacent to more than one basins, thus they are object boundaries (see Figure 2.9c).

The watershed segmentation separates an image into three different types of data: (i) the pixels of the local minima give the locations of the segmented objects, (ii) the catchment basins are the regions of the objects, (ii) the detected watersheds are the object boundaries. Practically, different algorithms have been developed to detect watersheds, according to the specific images [141, 156, 183].

Regarding the segmentation of time-lapse image sequences of breast cancer cells, an automatic segmentation process is necessary. Due to the scenario of cell connectedness even overlap, the watershed algorithm is considered [141].



**Figure 2.9: The basic idea of 1-D watershed segmentation [141].** (a) A grey-scale image is represented by a landscape where the amplitude is the grey-scale level at corresponding image location. (b) Object areas are detected by increasing the water line iteratively. (c) The detected catchment basins and watersheds which are the object areas and boundaries, respectively.

### 2.3.3 Active contour models for shape representation

Active contour models, also known as deformable contour model or ‘snakes’, were proposed to assist computer image analysis of the deformable shapes by combining image information with geometry, physics and optimal approximation principles. Geometry is employed for the description of object shapes, physics is applied to define the constraints on shape transform over space and time and approximation theory is applied to fit the model to the image data. Compared to traditional image processing, active contour models are able to achieve a more accurate delineation of object outlines by incorporating prior knowledge into image information. Furthermore, deformable energy terms in the active contour models enable the description of deformable shapes [76, 99, 100].

In the past decades, active contours have been widely used for the image segmentation, shape representation and object tracking. Different image informations, such as image intensity [67], colours [94] and different texture features [102], were included into the construction of image fitting energy functions so as to perform image segmentations according to the specific image types [76].

#### Geometric active contours

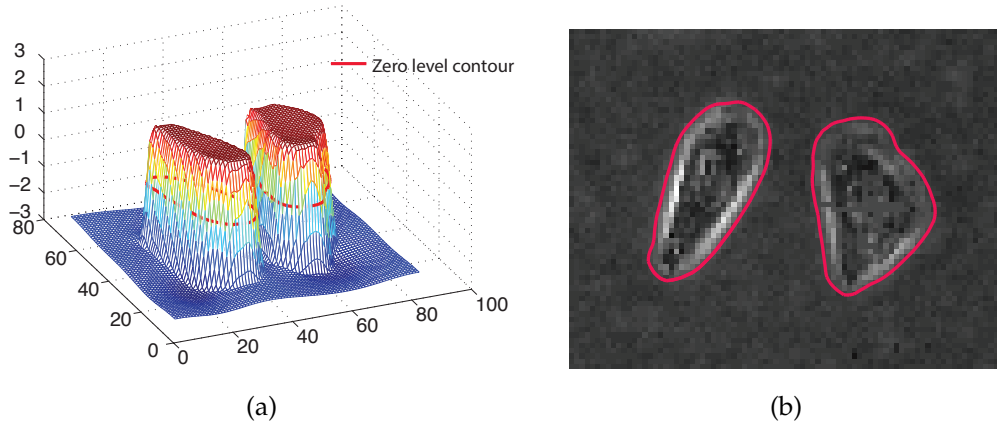
According to their mathematical derivations, active contours are divided into geometric active contours and parametric active contours [169, 184]. Geometric active contours are derived from geometric evolution and the contours are calculated as the zero level set of a 2-D scalar function of image,  $I(i_x, i_y)$ , and time period,

$t \in [0, T]$ , given by

$$\mathcal{U}_t = \{(i_x, i_y) \in \Omega | \varphi(i_x, i_y, t) = 0\} \quad (2.86)$$

where  $\mathcal{Y}_t$  are object contours at  $t$  and  $\varphi(i_x, i_y, t)$  is a time-dependent level set function defined on a domain,  $\Omega$  [89].

Figure 2.10 shows an example of cell contour segmentation using geometric active contour. The evolution of the level set function is shown in Figure 2.10a and the corresponding segmentation results are shown in Figure 2.10b.

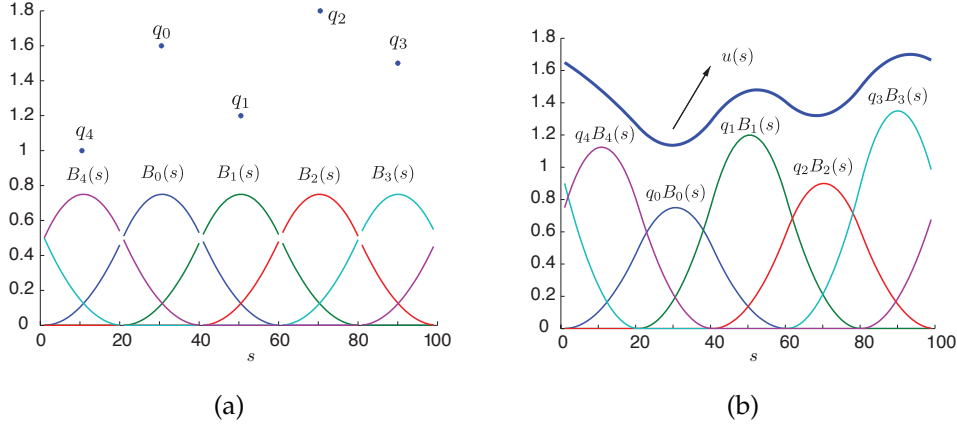


**Figure 2.10: An example of cell contour representation by geometric active contour [90].** (a) Level set function after 210 iterations and the related zero level contours, which are represented by red contours. (b) Overlay of zero level contours on the raw image, where the contours describe the outline of cells effectively.

The calculated contours can be regarded as a time-dependent surface where the scalar function  $\varphi(i_x, i_y, t)$  equals zero permanently. It is hence known as an implicit representation and enables to describe contour topological changes over time, such as division and merging of contours. Therefore, geometric active contours have been widely applied to describe the scenarios with frequent topological changes [92, 169]; for example, Zimmer et al. achieved the tracking of contacting cells and cell divisions by involving repulsive forces and ‘topological operators’ into the scalar function [186].

### Parametric active contours

The parametric active contour model is derived by minimising the energy functions to balance the external constraint forces and internal forces of a curve [186]. The active contour model, as an explicit representation, represents contours in the form of parametric functions, which is convenient for the application of signal processing techniques.



**Figure 2.11: An example of B-spline contours [19].** (a) Quadratic B-spline basis functions and control points in the same splines; (b) B-spline contours constructed by the sum of basis functions weighted by control points.

A B-spline parametric contour was proposed as a typical parametric contour and is constructed as the weighted sum of B-spline basis functions, given by,

$$u(s) = \sum_{l=0}^{L-1} B_l^d(s) \mathbf{q}_l = B(s) \mathbf{Q} \quad (2.87)$$

where  $s$  is the parameter of constructed curves  $u(s)$ ,  $B_l^d(s)$  is the  $l$ -th basis function with parameter  $s$ , which is a polynomial of order  $d - 1$ , and  $\mathbf{q}_l$  is the control points, calculated by fitting to image data. A schematic of B-splines with quadratic basis functions is given in Figure 2.11. Due to the computational simplicity and ease of manipulation, B-spline active contours have been applied to image processing in various areas, such as traffic monitoring, robot grasping and facial animation [19].

### 2.3.4 Morphology description techniques

Shape features can be measured by different parameters. The morphology parameters, including shape perimeters, areas, minor and major axes, can be directly obtained from images or segmented object regions. Parameters, such as circularity, eccentricity and concavity, require a calculation using the basic parameters [41].

Furthermore, various shape descriptors were proposed to address specific shape feature changes, and at the same time being invariant with other transformations. The shape descriptors can be used for solving the problems associated with shape matching and shape recognition based on the similarity of shape descriptors [54].



Curvature scale-space (CSS) descriptor is a scale space representation and is utilised to describe the local concavity or convexity of two dimensional shapes. The curvature scale-space (CSS) descriptor is derived from the definition of curvature. Assume that a two dimensional parametric curve  $u(s) = (x(s), y(s))$  is recursively smoothed by a Gaussian filter with mean value  $s$  and standard deviation  $\sigma$ , given by

$$g(s, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{s^2}{2\sigma^2}\right) \quad (2.88)$$

the evolved curve is  $u_\sigma(s) = (x_f(s, \sigma), y_f(s, \sigma))$  and calculated by

$$x_f(s, \sigma) = x(s) * g(s, \sigma) \quad (2.89)$$

$$y_f(s, \sigma) = y(s) * g(s, \sigma) \quad (2.90)$$

where  $*$  denotes convolution operator. The curvature of the evolved curve is then calculated by

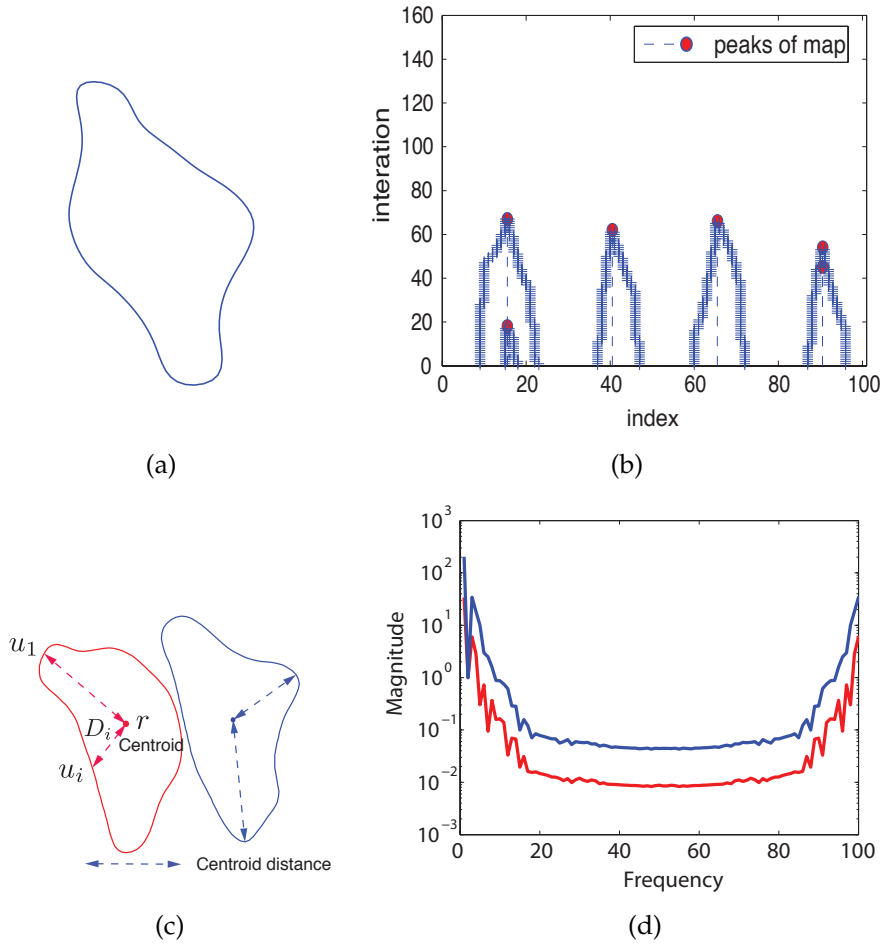
$$\kappa_f(s, \sigma) = \frac{\dot{x}_f(s, \sigma)\dot{y}_f(s, \sigma) - \dot{x}_f(s, \sigma)\ddot{y}_f(s, \sigma)}{(\dot{x}_f^2(s, \sigma) + \dot{y}_f^2(s, \sigma))^{3/2}} \quad (2.91)$$

Figure 2.12b shows an example of a CSS map of the cell contour given in Figure 2.12a. In Figure 2.12b, the horizontal coordinate indicates the index of points on the shape outline and vertical coordinate is the evolution times. The CSS map is constructed by the zero crossing points of the curvature over the evolution (i.e.,  $\kappa_f(s, \sigma) = 0$ ). The indices of the peaks are the locations of four deep concave or convex areas on the original contour, which is consistent with the actual situation in Figure 2.12a.

Fourier descriptor (FD) is another commonly applied shape descriptor. It was successfully used in the analysis of cell shape populations combined with principal component analysis (PCA) [121]. As shown in Figure 2.12c, shape centroid distance  $D_i$  is selected as the signature. According to the Fourier transformation

$$a_i = \frac{1}{N} \sum_{i=0}^{N-1} D_i \exp\left(-j\frac{2\pi ni}{N}\right), \quad n = 0, 1, \dots, N-1 \quad (2.92)$$

where  $N$  is the number of points on the boundary. Normalisation is carried out afterwards to keep the descriptors invariant to shape shifting, scale changing and shape rotation [179, 187]. Figure 2.12d is the FDs of the shapes given in Figure 2.12c based on the morphology signature of centroid distance. The different shapes give different patterns in the frequency domain, especially their magnitudes, which are commonly selected as criteria to measure the similarity between two shapes.

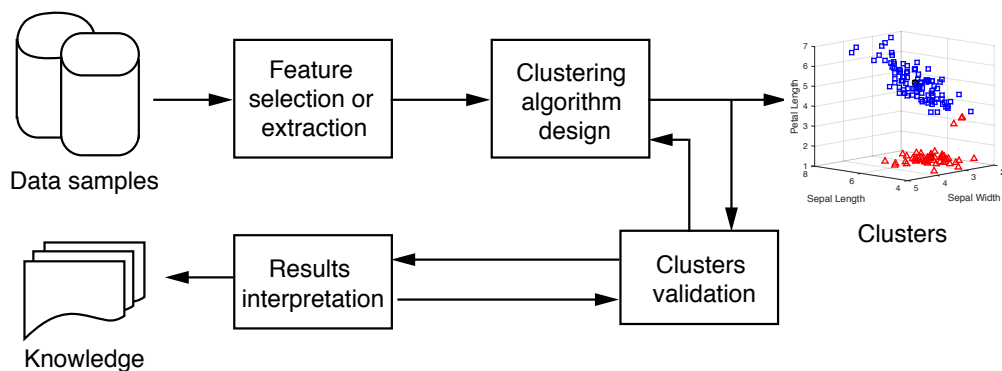


**Figure 2.12: Examples of cell shape descriptors.** (a) The original shape; (b) CSS map of subfigure (a); (c) Two different shapes; (d) The FDs of shapes shown in subfigure (c). The red and blue descriptors correspond to the red and blue shapes in subfigure (c), respectively.

Compared to CSS descriptors, FDs require smaller computational complexity and easier normalisation. CSS descriptors capture the global curvature features while FDs keep global and local features at different frequencies. Therefore, FDs are more suitable for the problems involving a high number of objects. However, FDs have the drawback of signature selection, which needs to be determined with respect to the actual situations [179].

### 2.3.5 Cluster analysis

Classification analysis plays a vital role in exploring unfamiliar phenomena. With a large amount of collected data, researchers attempt to group the data into different categories based on specific features. These categories can then be tested to



**Figure 2.13: Procedure for cluster analysis.** Cluster analysis is separated into four steps: feature selection, clustering algorithm design, clusters validation and results. The analysis system is performed as a feedback pathway, where these four steps corporation with each other to give an insight into the experimental data [170].

match the well-known data, therefore helping to understand the original phenomena [170]. Classification systems are divided into supervised and unsupervised systems [74]. Supervised classification is performed to map the input data to finite labelled classes, whereas, unsupervised classification does not have labelled data. Unsupervised classification, also known as clustering, aims to separate the unlabelled input data into discrete groups automatically [32].

Figure 2.13 shows the procedure for cluster analysis, which consists of feature selection or extraction, clustering algorithm design, clusters validation and interpretation of results [170].

First, feature selection or extraction is the process to describe the original data by its distinctive features, which can be selected from the given features or be transformed from the data. An elegant feature should be interpreted and obtained easily, insensitive to noise in the data and able to simplify the studied process. Crucially, the features used should present distinct patterns for different clusters.

Second, the clustering algorithm should be designed according to the extracted features. A measure of the features is required to ensure that feature similarity and difference can be evaluated. Moreover, a clustering criterion needs to be constructed to optimise the clusters. According to the cluster structures, clustering algorithms are separated into hierarchical clustering and partitional clustering. Hierarchical clustering performs clustering by connecting objects and gives a tree-like structure of each cluster. Partitional clustering directly divides the data into a finite number of clusters without hierarchical relations between objects. Different clustering algorithms are introduced in Section 2.3.6.

Then, validation for the clustering results is carried out because using different approaches to the same data usually leads to different results. External evaluation is commonly used to compare the clustering results from different methods. The external validation is based on a standard, such as the labelled classes and benchmarks, which are generated by experts and defined in advance, indicating the prior knowledge about the data [44].

Further analysis of the cluster results in combination with the original data is carried out afterwards to assist in the study of the original objectives.

Cluster analysis has been widely applied to solve practical problems. For example, Rennard et al. identified five subgroups of chronic obstructive pulmonary disease by cluster analysis of clinical features [125]. Spasova et al. analysed the genetic remoteness of Bulgarian and Macedonian cotton lines using hierarchical cluster analysis [144]. Similarly, hierarchical cluster analysis was applied to study the socio-economic status of international immigrants in Chile [25]. Accordingly, cluster analysis can deal with data from a wide variety of disciplines, ranging from engineering, computer science, chemical science to social science, life science and economics [170]. Practically, there does not exist a clustering algorithm that can deal with all the problems. Specific algorithms need to be selected or designed according to the particular purpose of the study and features of the data.

### 2.3.6 Clustering algorithms

As mentioned above in Section 2.3.5, clustering algorithms can be roughly categorised into hierarchical and partitional clustering, which are mathematically described as follows [56]:

- Assuming the input data set

$$\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N\} \quad (2.93)$$

where  $\mathbf{x}_n = [x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(M)}]^\top \in \mathbb{R}^M$  is the vector of observed features.

The partitional clustering is aimed to divide  $\mathcal{X}$  into  $K$  partitions:

$$\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_K\}, \quad K \leq N \quad (2.94)$$

The partitions attempt to satisfy the following:

- $\mathcal{C}_k \neq \emptyset, k = 1, \dots, K$ , where  $\emptyset$  denotes an empty set.
- $\cup_{k=1}^K \mathcal{C}_k = \mathcal{X}$
- $\mathcal{C}_{k_1} \cap \mathcal{C}_{k_2} = \emptyset, k_1, k_2 = 1, \dots, K$  and  $k_1 \neq k_2$

**Algorithm 2.1 Hierarchical clustering: Agglomerative approach****Input:**Data set:  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N\}$ Distance function:  $\text{dist}(\mathcal{C}_{g_1}, \mathcal{C}_{g_2})$  // Distance between  $\mathcal{C}_{g_1}$  and  $\mathcal{C}_{g_2}$ **Output:**Clustering results:  $\mathcal{C}$ **For**  $n \leftarrow 1$  to  $N$ Set  $\mathcal{C}_n = \{\mathbf{x}_n\}$ **End For**Set  $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_N\}$ **While**  $\mathcal{C}.\text{length} > 1$  **do**Set  $(\mathcal{C}_{\min 1}, \mathcal{C}_{\min 2}) = \text{minimum dist}(\mathcal{C}_{g_1}, \mathcal{C}_{g_2})$  for all  $\mathcal{C}_{g_1}, \mathcal{C}_{g_2}$  in  $\mathcal{C}$ Remove  $\mathcal{C}_{\min 1}$  and  $\mathcal{C}_{\min 2}$  from  $\mathcal{C}$ Add  $\{\mathcal{C}_{\min 1}, \mathcal{C}_{\min 2}\}$  to  $\mathcal{C}$ **End While**

- The hierarchical clustering is aimed at organising the observation set,  $\mathcal{X}$ , into tree-like structured subsets:  $\mathcal{D} = \{\mathcal{D}_1, \dots, \mathcal{D}_G\}$  where  $G \leq N$ . Hence,  $\mathcal{C}_{k_1} \in \mathcal{D}_{g_1}$ ,  $\mathcal{C}_{k_2} \in \mathcal{D}_{g_2}$  and  $g_1 > g_2$ , where  $g_1, g_2, k_1, k_2 = 1, 2, \dots, G$ , indicate the relation of  $\mathcal{C}_{k_2} \in \mathcal{C}_{k_1}$  or  $\mathcal{C}_{k_2} \cap \mathcal{C}_{k_1} = \emptyset$ .

**Hierarchical clustering**

In hierarchical clustering, algorithms, according to the direction of data processing, are divided into division and agglomerative approaches.

The general agglomerative approach is given in Algorithm 2.1 [131]. The clustering starts with initialising each input vector,  $\mathbf{x}_n$ , as a single cluster,  $\mathcal{C}_n$ . Subsequently, the clusters of the minimum distance are merged as one until only one cluster is left in  $\mathcal{C}$ .

In the applications, different distance functions are employed in different algorithms. The most straightforward method is single linkage clustering, where the distance between two clusters is calculated as the minimum distance between the elements in each cluster [170]:

$$\text{dist}(\mathcal{C}_{g_1}, \mathcal{C}_{g_2}) = \min\{\|\mathbf{c}_i - \mathbf{c}_j\| : \mathbf{c}_i \in \mathcal{C}_{g_1}, \mathbf{c}_j \in \mathcal{C}_{g_2}\} \quad (2.95)$$

where  $\|\cdot\|$  denotes the Euclidean distance and  $\mathbf{c}_i$  and  $\mathbf{c}_j$  are the elements in cluster  $\mathcal{C}_{g_1}$  and  $\mathcal{C}_{g_2}$ , respectively. The complete lineage method, on the other hand, uses the maximum distance between the elements of each cluster,

$$\text{dist}(\mathcal{C}_{g_1}, \mathcal{C}_{g_2}) = \max\{\|\mathbf{c}_i - \mathbf{c}_j\| : \mathbf{c}_i \in \mathcal{C}_{g_1}, \mathbf{c}_j \in \mathcal{C}_{g_2}\} \quad (2.96)$$

The other commonly used lineage method is average linkage, and the distance between clusters is given by [170]

$$\text{dist}(\mathcal{C}_{g_1}, \mathcal{C}_{g_2}) = \frac{1}{N_{g_1} N_{g_2}} \sum_{i=1}^{N_{g_1}} \sum_{j=1}^{N_{g_2}} \|\mathbf{c}_i - \mathbf{c}_j\| \quad (2.97)$$

where  $\mathbf{c}_i \in \mathcal{C}_{g_1}$ ,  $\mathbf{c}_j \in \mathcal{C}_{g_2}$ ,  $N_{g_1}$  and  $N_{g_2}$  are the size of the cluster  $\mathcal{C}_{g_1}$  and  $\mathcal{C}_{g_2}$ , respectively.

The division method, performed in the reverse direction than the agglomerative method, first groups all the observations,  $\mathbf{x}_n$ , as a unit cluster, then recursively divides the cluster until that each cluster only has one observation.

When hierarchical clustering is carried out, the observations will not be checked after being assigned to a cluster. The algorithm is not able to correct for previously misaligned elements, which causes hierarchical clustering to be sensitive to noises and outlier observations, which is a challenge for large-scale data set clustering.

### Partitional clustering

Partitional clustering assigns the observations,  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N\}$ , into  $K$  clusters,  $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_K\}$  according to data-intrinsic distributions. Many clustering algorithms have been generated by assuming specific data distributions. Due to the complexity of realistic data, a universal algorithm and objectively ‘correct’ clustering has not been achieved. Hence, such algorithms are empirically selected for the specific practical problems [64].

K-means, first published in [97], is the simplest partitional clustering algorithm and is still widely used. The K-means algorithm performs clustering by assigning objects to the nearest cluster based on the centroid distance. The algorithm is precisely explained in Algorithm 2.2 and is summarised in the following steps:

- Clusters are either initialised randomly by setting cluster centres,  $\mathcal{M}$ , randomly or initialised using prior knowledge.
- The observations are assigned to the cluster for which the centre has the smallest distance to the observations.
- Cluster centres are recalculated using the mean of observations assigned to the cluster.
- The previous two steps are iteratively performed until the cluster centres stop changing.

---

**Algorithm 2.2 Partitional clustering: K-means algorithm**

---

**Input:**Data set:  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N\}$ Number of clusters:  $K$ Maximum of iterations:  $\text{maxIter}$ **Output:**Cluster labels of  $\mathcal{X}$ :  $\text{labelX} = \{\text{label}(\mathbf{x}_n) \mid n = 1, 2, \dots, N\}$ Cluster centres:  $\mathcal{M} = \{\mu_1, \dots, \mu_K\}$ **For**  $k \leftarrow 1$  to  $K$     Set  $\mu_k = \mathbf{x}_n \in \mathcal{X}$      // Initialise cluster centres by random observations.**End For****For**  $n \leftarrow 1$  to  $N$     Set  $\text{label}(\mathbf{x}_n) \leftarrow \arg \min_k \text{dist}(\mathbf{x}_n, \mu_k), k \in \{1, \dots, K\}$ 

// Initialise the label set by the cluster of minimum distance.

**End For**Set  $\text{iter} \leftarrow 0$ Set  $\text{updated} \leftarrow 1$ **While**  $\text{iter} \leq \text{maxIter}$  &  $\text{updated} = 1$  **do**    **For**  $k \leftarrow 1$  to  $K$          $\text{updateCluster}(\mu_k)$     **End For**    Set  $\text{updated} \leftarrow 0$     **For**  $n \leftarrow 1$  to  $N$         Set  $\text{minD} \leftarrow \arg \min_k \text{dist}(\mathbf{x}_n, \mu_k), k \in \{1, \dots, N\}$         **If**  $\text{minD} \neq \text{label}(\mathbf{x}_n)$             Set  $\text{label}(\mathbf{x}_n) \leftarrow \text{minD}$             Set  $\text{updated} \leftarrow 1$         **End If**    **End For**Set  $\text{iter} \leftarrow \text{iter} + 1$ **End While**

---

In practice, the K-means algorithm is easy to implement. The development of parallel techniques makes the K-means clustering capable of dealing with large-scale data sets [81]. However, the K-means algorithm has the drawback of clustering initialisation, including the identification of cluster numbers and locations. An inappropriate initialisation can significantly increase the computational complexity [170]. The K-means algorithm attempts to assign the observations into same sized partitions, which leads to the clustering results converging to a local optimum instead of a global optimum, especially for the unevenly distributed data. Besides, the strong links to the data means that K-means algorithm is sensitive to noise and outliers.

Distribution-based clustering is another widely used algorithm type, apart from the centroid-based algorithm. In different clusters, the observations are assumed to be generated from different probability distributions, where the prior probability of cluster  $\mathcal{C}_k$  is  $p(\mathcal{C}_k)$  and the corresponding conditional probability density is  $p(\mathbf{x} | \mathcal{C}_k, \boldsymbol{\theta}_k)$  with unknown parameter vector,  $\boldsymbol{\theta}_k$ . The probability density of mixture models,  $p(\mathbf{x} | \boldsymbol{\theta})$ , is

$$p(\mathbf{x} | \boldsymbol{\theta}) = \sum_{k=1}^K p(\mathbf{x} | \mathcal{C}_k, \boldsymbol{\theta}_k) p(\mathcal{C}_k) \quad (2.98)$$

where  $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K]$  and the component distributions, in theory, can be of any type. A mixture of multivariate Gaussian distributions is a prominent structure, integrated with the EM algorithm for parameters estimation through maximising the likelihood function (see Section 4.2.2).

A drawback of the EM algorithm is its sensitivity to the parameter initialisation. The random initialisation may lead to be converged to a local optimum, consequently leading to wrong estimations of the cluster parameters [170].

### Sequential data clustering

Sequential data are generated from scenarios having relevant spatial or temporal structure such as DNA sequences, video sequences and stock markets [95]. Sequential data clustering uses unsupervised learning algorithms to find the statistically significant potential patterns of the sequence data. The clustering algorithms are categorised into: sequence similarity clustering, indirect sequence clustering and statistical sequence clustering [170].

The sequence similarity clustering method first measures the similarity between different data sequences, then groups similar sequences using clustering algorithms, which can be either hierarchical or partitional algorithms. Sequence



clustering analysis can efficiently remove redundant information from the original data set, so as to simplify the computation of later analyses. At the same time, some fundamental questions are addressed: for example, clustering analysis of metagenomics can help identify protein families [93].

Indirect sequence clustering first extracts features from the sequences. All the sequences are then mapped to the extracted feature space. Afterwards, the classical clustering algorithms can be employed to perform cluster analysis in the feature space. Indirect sequence clustering can efficiently reduce computational complexity by sacrificing some information from the original sequences. Thus, the feature extraction is essential for keeping enough information for clustering [108].

Statistical sequence clustering is generally used for numerical data sequence clustering, while the indirect sequence and sequence similarity clustering are more commonly employed for alphabet sequences analysis [170]. Statistical sequence clustering describes the dynamics of each group of sequences through statistical models. The most commonly used model is HMM, combined with EM for parameter identification (see Section 4.2.2). According to the particular data, different models can also be used. For example, Xiong and Yeung proposed an algorithm for time series clustering using mixtures of autoregressive moving average (ARMA) model and Markov chain [168]. This statistical model-based clustering can intuitively describe the dynamics of different data types and is capable of processing data sequences of different lengths. However, in order to better capture data characteristics, selection of an appropriate model is necessary [170].

In this section, the commonly used cluster algorithms were reviewed. In practice, the algorithms should be chosen according to the specific problems. A large number of clustering algorithms have been proposed in the literature. However, some problems still remain challenging, such as large-scale data and high-dimensional data clustering.

## 2.4 Summary

In this chapter, the basic techniques that are required to construct a framework for breast cancer cell behaviour analysis has been presented. The methods of image processing, active contour models, discrete-time kinematic models and state estimation and smoothing provide the basis of an automated cell boundary tracking system, which is proposed in Chapter 3, and applied to breast cancer control cells. The shape descriptors combined with clustering algorithms can be applied to characterise cell morphology states, which is covered in Chapter 4. The morphology

analysis, carried out after the cell tracking, is applied to the tracking results of cell boundaries.

## Chapter 3

# High-throughput Tracking of Cancer Cells Using Parametric Active Contour and Kalman Smoothing

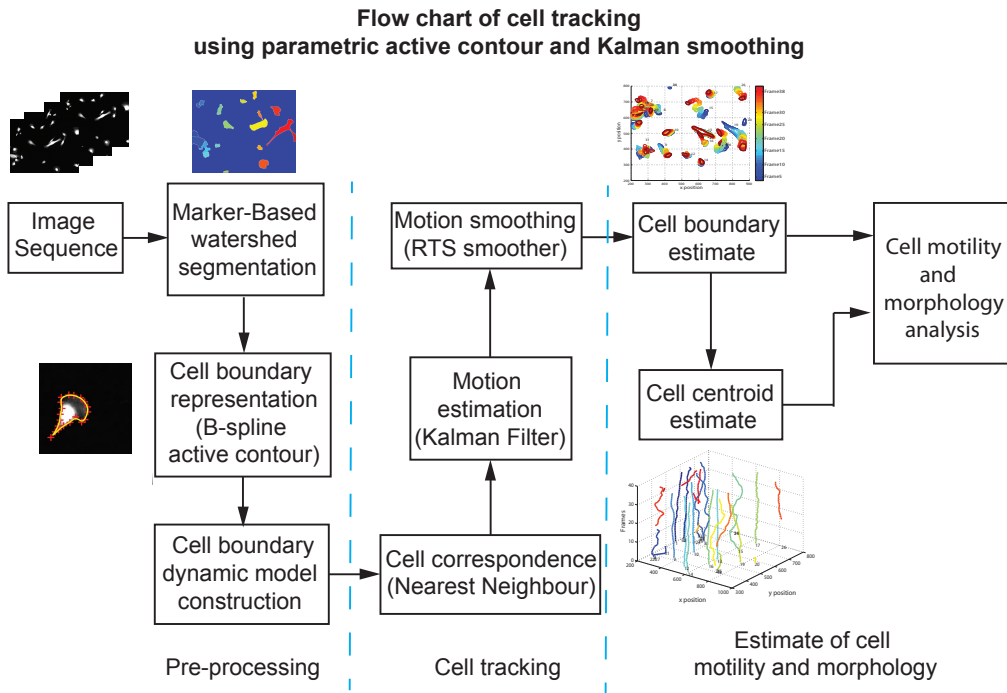
High-throughput cell tracking is essential for investigating functions and dynamic characteristics of multiple cells. The large amounts of information, generated after tracking, can then be used for analysing various aspects of cell behaviours, such as the study of cell cycle, the gene expression and the drug discovery [182].

Traditional manual tracking and analysis of high-throughput data is restrictive because of the time consuming and large requirement of human resources. There is thus a need for automatic cell tracking to support the life science researchers. The basis for most cell tracking methods is the particle tracking techniques. For example, Li et al. proposed a tracking algorithm by combining IMM filter with maximum-likelihood matching to accomplish high-throughput cell centroid tracking [92]. Reyes-Aldasoro et al. proposed a tracking algorithm based on a 'keyhole' model, which predicted the probable cell landing position between connective frames and achieved the centroid tracking of fluorescent labelled red blood cells migrating in microvessels of tumour [126]. Additionally, active contours derived from evolving deformable shapes have been used to describe cell shapes. For example, level-set functions derived from different shape constraints were commonly used to segment the overlapped cells from the microscopy images [185].

The particles tracking techniques capture the spatio-temporal trajectories of cells, but completely ignore the information about the deformable cell shapes. The

active contours, in turn, capture cell shape deformation but ignore the cell spatio-temporal information. This results in the corresponding applications being limited to the study of either cell motility or extract morphology features. In cancer cells primarily, and in other type of cells, it has been observed that cell motility and cell morphology characteristics have high correlation [48].

In this chapter, a tracking framework is proposed that can simultaneously track the cell boundaries, their spatio-temporal dynamics and cell centroid trajectories. Unlike in previous cell tracking literature, this framework uses B-spline active contours for cell boundary description. This linear dimensional parametric description enables an efficient tracking of cell boundaries by a lower dimensional dynamic model, additionally links the cell motility and morphology naturally. This enables the possibility of carrying out integrated cell motility and morphology analysis.



**Figure 3.1: Overview of the cell tracking framework.**

The structure of the proposed tracking framework is shown in Figure 3.1 and is summarised as follows:

- The marker-controlled watershed cell segmentation, a non-parametric segmentation algorithm, is used first to extract the cell boundary features from the image sequences to enable the tracking.

- The B-spline active contours are employed to describe the extracted cell boundaries efficiently parametrised through control points.
- The dynamics of cell shapes are represented by a second-order kinematic model of control points whose positions and velocities are estimated by Kalman filtering and smoothing.
- The parametric active contour representation and the estimated position and velocity time sequences of the control points together are provided to a module of quantitative analysis of cell motility features.

### 3.1 Automatic cell boundary segmentation using marker-controlled watershed algorithm

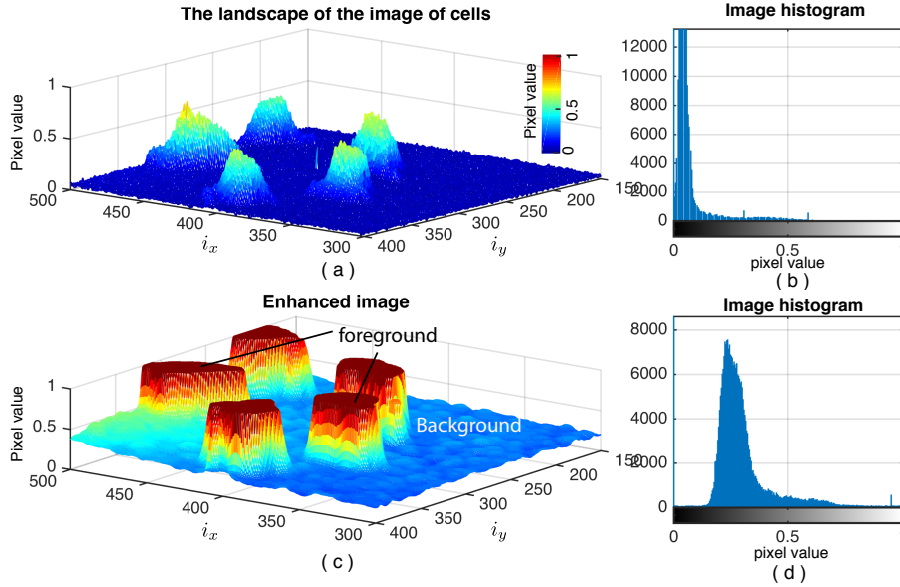
The watershed algorithm, being a non-parametric image segmentation method, operates based on pixel intensities [14]. In the applications, the main advantage of watershed algorithm is the capability to separate closely located objects, such as in delineating touching cells. Images can be envisaged as a landscape (see Figure 3.2a), where  $(i_x, i_y)$  is the coordinate of pixel with value  $I(i_x, i_y)$ . In a grey-scale image, pixel values are in the range  $[0, 1]$ , corresponding to the colour varying from black to white.

In the conventional watershed algorithm, ‘water lines’ are initialised at the areas with local minimum intensity and are then increased until it meets others at the ‘watersheds’. Segmented cell contours are essentially the watersheds. The main drawback of the watershed algorithm is the over-segmentation, which is resulted in the image noise, diversity of cell shapes and large cell quantity. In order to avoid the over-segmentation, the marker-controlled watershed algorithm is applied to perform the automatic cell segmentation. The segmentation is designed as follows [24]:

- The photobleach induces significant change in the image contrast across different frames in the same image sequence. The pixel value of original image,  $I_{original}(i_x, i_y)$ , is enhanced first according to the formula:

$$I_{enhanced}(i_x, i_y) = (I_{original}(i_x, i_y) - \epsilon_a) \left( \frac{1}{\epsilon_b - \epsilon_a} \right) \quad (3.1)$$

where  $I_{enhanced}$  is the enhanced image,  $[\epsilon_a, \epsilon_b]$  is set to be  $[0, 0.1]$  and it is the range of the pixel values, according to the image pixel histogram (see Figure 3.2b), that need to be emphasised. The effect of contrast stretching is shown in Figure 3.2c and Figure 3.2d.



**Figure 3.2: An example of landscape view of a grey-scale image.** (a) Landscape view of a grey scale image; (b) Pixel histogram of image in subfigure (a); (c) Landscape view of the contrast stretched image corresponding to subfigure (a); (d) The pixel histogram of subfigure (c). In subfigures (a) and (c) the pixel value is mapped to a RGB colour scale to aid visualisation. In the grey-scale images, pixel values from 0 to 1 correspond to the colour varying from black to white. In the histograms, subfigures (b) and (d),  $x$ -axis is the pixel values and  $y$ -axis is the number of pixel values occurred in the image.

- The background marker is the collection of the pixels not belonging to any cells. They are obtained by applying the intensity thresholding:

$$I_{bm}(i_x, i_y) = \begin{cases} 1 & \text{if } I_{enhanced}(i_x, i_y) > \epsilon_o \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

where,  $\epsilon_o$  is the threshold selected by Otsu's method, calculated by minimising the interclass variance [112].

- The foreground markers, as shown in Figure 3.2c, are used to locate the cells. They are obtained by applying morphological operators followed by intensity thresholding.
- The above two different markers are set to be minus infinity, where the 'water lines' are initialized. The cell regions then can be segmented by performing the watershed segmentation.

- Cell boundaries can be obtained from the segmented cell areas. Meanwhile, intensity thresholding is applied one more time to each of the segmented cell area to separate the cell main area and the shadow area, which is normally composed of cell protrusion.

### 3.2 Cell boundary representation by B-spline active contours

The observed boundary of a single cell is modelled by,

$$\mathbf{z} = \mathbf{u} + \boldsymbol{\omega} \quad (3.3)$$

where  $\mathbf{z} = \begin{bmatrix} \mathbf{z}^{(x)} \\ \mathbf{z}^{(y)} \end{bmatrix}$ , and  $\mathbf{z}^{(x)}, \mathbf{z}^{(y)} \in \mathbb{R}^{N_z}$  are the vector of  $x$  and  $y$  coordinates of cell boundary obtained from cell segmentation, and  $N_z$  is the number of points on cell boundary.  $\mathbf{u}$  is the B-spline active contour representation and  $\boldsymbol{\omega}$  is the noise contributing to the model fit errors. The B-spline contours are constructed as:

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}^{(x)} \\ \mathbf{u}^{(y)} \end{bmatrix} = \begin{bmatrix} \sum_{l=0}^{L-1} B_l(s) q_l^{(x)} \\ \sum_{l=0}^{L-1} B_l(s) q_l^{(y)} \end{bmatrix} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{q}^{(x)} \\ \mathbf{q}^{(y)} \end{bmatrix} \quad (3.4)$$

where,

$$\mathbf{u}^{(x)} = \begin{bmatrix} u_1^{(x)} & \cdots & u_{N_z}^{(x)} \end{bmatrix}^\top \quad (3.5)$$

$$\mathbf{u}^{(y)} = \begin{bmatrix} u_1^{(y)} & \cdots & u_{N_z}^{(y)} \end{bmatrix}^\top \quad (3.6)$$

$$\mathbf{B} = \begin{bmatrix} B_0(s) & B_1(s) & \cdots & B_{L-1}(s) \end{bmatrix} \quad (3.7)$$

$$\mathbf{q}^{(x)} = \begin{bmatrix} q_0^{(x)} & \cdots & q_{L-1}^{(x)} \end{bmatrix}^\top \quad (3.8)$$

$$\mathbf{q}^{(y)} = \begin{bmatrix} q_0^{(y)} & \cdots & q_{L-1}^{(y)} \end{bmatrix}^\top \quad (3.9)$$

$$(3.10)$$

$\mathbf{u}^{(x)}, \mathbf{u}^{(y)} \in \mathbb{R}^{N_z}$ , and  $(u_{n_z}^{(x)}, u_{n_z}^{(y)})$  is the coordinate of the point on the B-spline contour, where  $n_z$  is the point index ( $n_z \in [1, N_z]$ ). The contour  $\mathbf{u}$  consists of  $L$  spans ( $L < N_z$ ).  $B_l(s)$  is the basis function for span  $l \in [0, L-1]$ , which is a function of a parameter  $s \in [l, l+1]$ , and  $(q_l^{(x)}, q_l^{(y)})$  are the Cartesian coordinates of the corresponding control point. The basis function matrix,  $\mathbf{B} \in \mathbb{R}^{N_z \times L}$ , is able

to be calculated off-line and the weights vector,  $\mathbf{q} = \begin{bmatrix} \mathbf{q}^{(x)} \\ \mathbf{q}^{(y)} \end{bmatrix} \in \mathbb{R}^{2L}$  is calculated by LS method. Due to the linear-parameter representation of B-spline contours and the dimension transform from  $\mathbf{u} \in \mathbb{R}^{2N_z}$  to  $\mathbf{q} \in \mathbb{R}^{2L}$ , the control points vector,  $\mathbf{q}$ , is later applied as the states to be tracked in the cell boundary tracking system.

### 3.2.1 The construction of B-spline active contours

The periodic basis function is used because of the wrap-around feature associated with the closed curve representation, which can be used to represent cell boundary. The basis function  $B_l(s)$  in span  $l \in [0, L)$  is

$$B_l(s) = \begin{bmatrix} s^0 & s^1 & \dots & s^{d-1} \end{bmatrix} \mathbf{B}_l \mathbf{G}_l \quad (3.11)$$

which is a vector of polynomials of parameter  $s$  whose highest order is  $d - 1$ . The basis function is calculated by the multiplication of a span matrix,  $\mathbf{B}_l$ , and the corresponding placement matrix,  $\mathbf{G}_l$ .

An index  $b_l$  is firstly calculated to indicate the first span affected by  $\mathbf{B}_l$ ,

$$b_l = \begin{cases} l - d + 2 & \text{if } l > d - 3 \\ l - d + L + 2 & \text{otherwise} \end{cases} \quad (3.12)$$

$B_{b_l+i-1,l}^d(s)$  is used to denote the  $i$ -th basis polynomials for span  $l$  of order  $d$  and it is recursively calculated from ground instance  $d = 1$ .

$$B_{n,l}^1(s) = \begin{cases} 1 & \text{if } n \leq s < n + 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.13)$$

where,  $n = b_l + i - 1$  and  $i = 1 \dots d$ . The higher order polynomials are then calculated recursively using

$$B_{n,l}^d(s) = \frac{(s + l - n)B_{n,l}^{d-1}(s) + (n + d - s - l)B_{n+1,l}^{d-1}(s)}{d - 1} \quad (3.14)$$

$\mathbf{B}_l$  is then structured with the coefficients of the basis polynomial  $B_{n,l}^d(s)$  on the  $i$ -th column and the constant terms on the first row.  $\mathbf{G}_l$  is the related placement matrix and computed using

$$\mathbf{G}_{l(i,j)} = \begin{cases} 1 & \text{if } i - b_l = j \\ 0 & \text{otherwise} \end{cases} \quad (3.15)$$



where  $i$  and  $j$  are the indices of row and column, respectively.

The estimation of the control points can be found using LS applied to (3.4),

$$\hat{\mathbf{q}} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{z} \quad (3.16)$$

where the design matrix  $\Phi$  is constructed by the vector of basis functions,

$$\Phi = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \quad (3.17)$$

Therefore, the estimated B-spline contour,  $\hat{\mathbf{z}}$ , is calculated by,

$$\hat{\mathbf{z}} = \Phi \hat{\mathbf{q}} \quad (3.18)$$

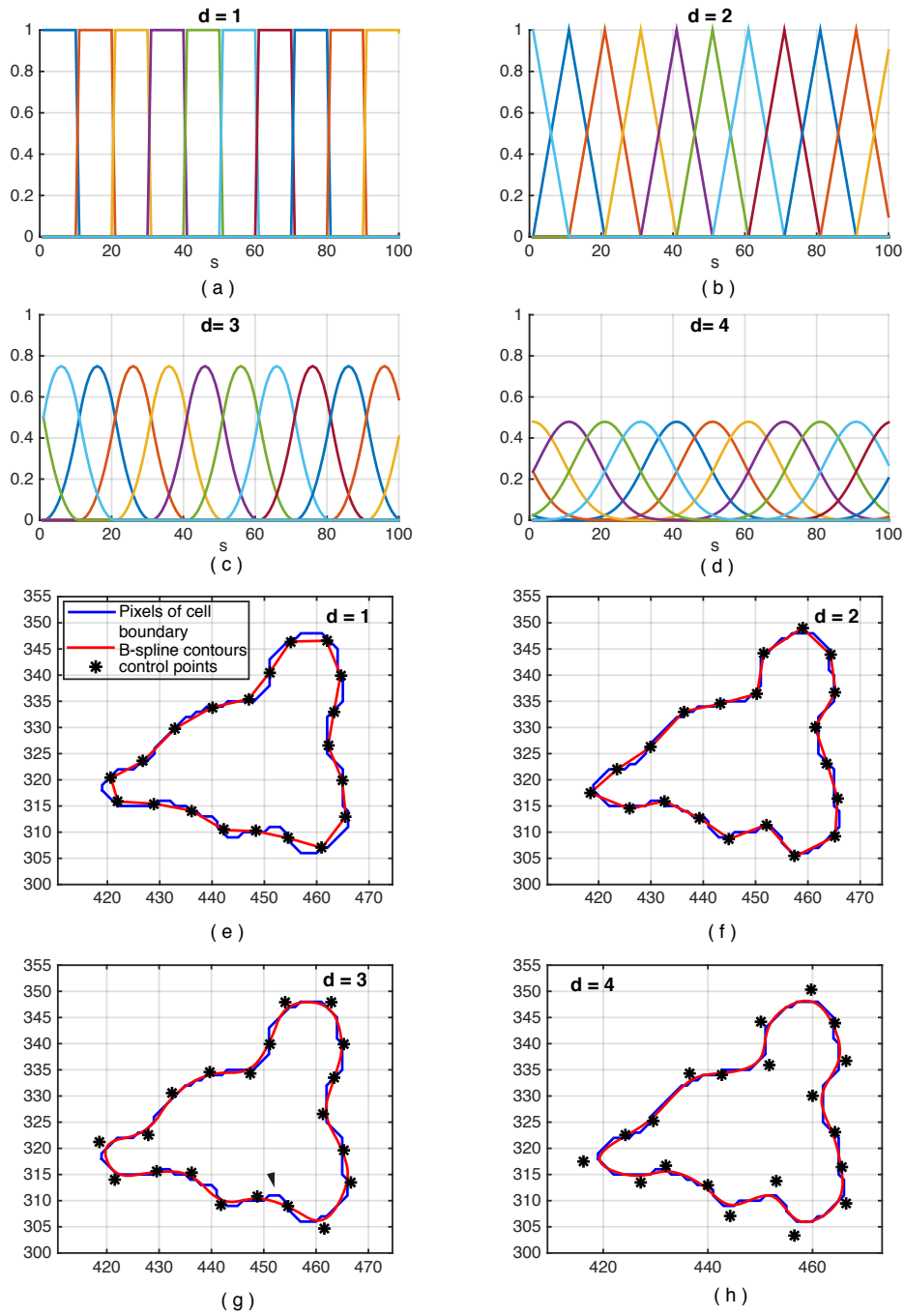
Control points can be calculated separately, here it is only used for initialisation and take advantage of dynamic model representation and state estimation to track them across frames.

### 3.2.2 The parameter selection of B-spline active contours

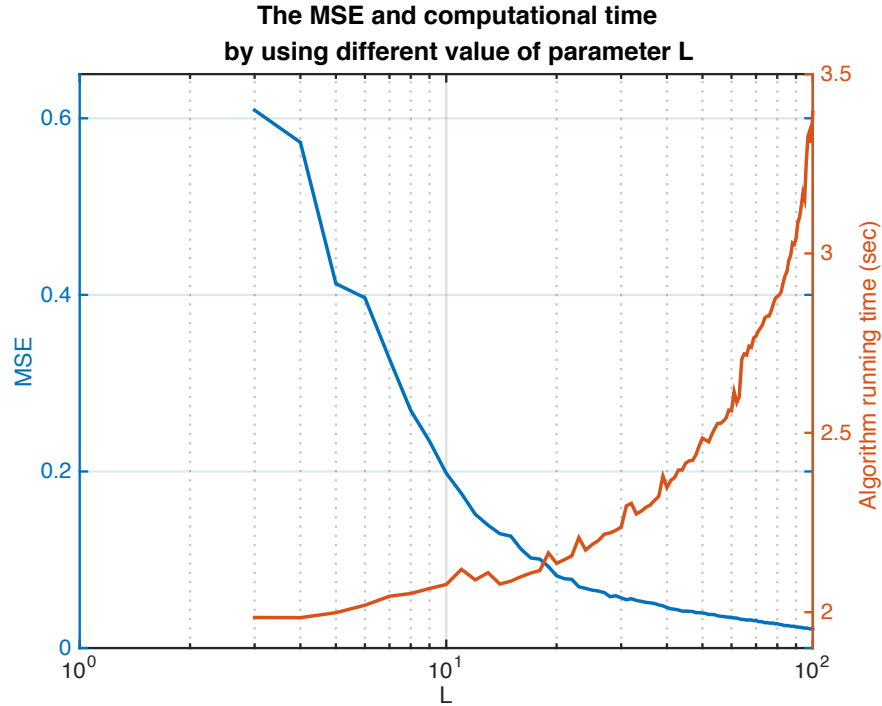
Two parameters are involved in constructing B-splines, which are: (i) the degree of the spline; (ii) the number of spans involved in the contour. Generally, the spline degree is fixed to be a small number in order to reduce the computational complexity. The model fitting accuracy can be adjusted by changing the number of spans.

The basis functions with spline orders,  $d = 1, 2, 3, 4$ , are used to fit an identical cell boundary with a choice  $L = 20$ . (see Figure 3.3). When  $d = 1$ , the constant basis functions generate a binary spline, given in Figure 3.3a. When  $d$  increases from 2 to 4, the constructed contours, which correspond to the linear ( $d = 2$ ), the quadratic ( $d = 3$ ) and the cubic ( $d = 4$ ) functions, are piece-wise, slope and curvature continuity respectively (see Figure 3.3b-d). The curvature continuity is essential to ensure the contours are smooth, therefore  $d$  is set to be 4 in order to describe the cell boundary, as well as boundary velocity, smoothly.

With the fixed degree  $d = 4$ ,  $L$  is set to different values in the range [3, 100] for boundary estimation with B-spline applied to the same image sequence of migrating cancer cells. The mean squared error (MSE) calculated to measure the



**Figure 3.3:** An example of fitting the same cell boundary using B-spline basis functions of different spline orders. The number of span is set to be 20. Subfigures (a)–(d), respectively, show the periodical basis functions of spline order,  $d = 1, 2, 3$ , and 4, over the spline  $s \in [0, 100]$ . Subfigures (e)–(h) present the fitting results of an identical cell boundary using the B-spline basis functions in subfigures (a)–(d), respectively.



**Figure 3.4:** The MSE and algorithm running time of B-spline estimation with different number of spans. Cubic basis functions were used for the estimation of cell boundaries of the image sequence given in Table 3.1 and number of span is set to  $L \in [3, 100]$ . The MSE and algorithm running time are the average situation over 50 times experiments.

estimation errors and the formula is:

$$MSE = \frac{1}{\sum_{i=1}^{N_c} N_z^{(i)}} \sum_{i=1}^{N_c} \sum_{n=1}^{N_z^{(i)}} \left( z_n^{(i)} - \hat{z}_n^{(i)} \right)^2 \quad (3.19)$$

where,  $z_n^{(i)}$  is the point  $n$  on the boundary of cell ( $i$ ), which is obtained from image segmentation, and  $\hat{z}_n^{(i)}$  is the corresponding point obtained from the estimated B-spline contour.  $N_c$  is the number of cells used and  $N_z^{(i)}$  is the number of pixels on the boundary of cell ( $i$ ). The same process was performed 50 times and the average algorithm running time and average MSE are shown in Figure 3.4. It is shown that with increasing span, the MSE decreases while the running time increases. To achieve a compromise,  $L$  is set to 20. Information on this experiment are provided in Table 3.1.

**Table 3.1: Experimental information for identification of the number of spans on B-spline contours.** The experiment was performed with cubic basis functions ( $d = 4$ ).

<b>Experimental data</b>	<ul style="list-style-type: none"> <li>• Cell line: MDA-MB-231 breast cancer cells</li> <li>• Label: J21-F1 (labelled with CellTracker Orange)</li> <li>• Cell culture: Cells are cultured on tissue culture plastic</li> </ul>
<b>Image sequences</b>	<ul style="list-style-type: none"> <li>• Image Type: Time-lapse grey-scaled image sequence</li> <li>• Frames: 38 Frames</li> <li>• Time intervals: 15 minutes between two frames</li> <li>• Cell quantity: 451 cells showed in the video</li> </ul>
<b>Computer and software</b>	<ul style="list-style-type: none"> <li>• Processor: Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz 3.10 GHz</li> <li>• RAM: 4GB</li> <li>• System: Windows 7 Professional (64-bit)</li> <li>• Software: MATLAB R2015a</li> </ul>

### 3.3 Cell boundary tracking with Kalman filter estimation and smoothing

In this section, a discrete second order spatio-temporal model is designed to describe the dynamics of cells, including the tracking of cell boundaries and centroids. The tracking consists of two parts, which are the kinematic model construction and the tracking with the KF estimation and smoothing.

#### 3.3.1 A discrete white noise acceleration model of cell migration

It is assumed that cells are migrating with constant velocity and subjected to random disturbances from the sampling moment  $t\Delta T$  to  $(t + 1)\Delta T$ . The periodical basis functions  $\mathbf{B}$  in (3.7) are of the highest degree 3 so that the second order derivative of the cell boundaries are continuous, which is indicative of a second-order kinematic model. A DWNA kinematic model for a single cell boundary is constructed with the following structure:

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{\Gamma}\boldsymbol{\omega}_t \quad (3.20)$$

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \quad (3.21)$$

where the state at frame  $t$ ,  $\mathbf{x}_t$ , consists of locations and velocities of control points,

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{q}_t^{(x)} \\ \mathbf{q}_t^{(y)} \\ \Delta \mathbf{q}_t^{(x)} \\ \Delta \mathbf{q}_t^{(y)} \end{bmatrix} \quad (3.22)$$

where  $\Delta \mathbf{q}_t = \begin{bmatrix} \Delta \mathbf{q}_t^{(x)} \\ \Delta \mathbf{q}_t^{(y)} \end{bmatrix}$  is the velocities of control points at frame  $t$ . The related transition matrix is

$$\mathbf{F} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \Delta T \cdot \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \Delta T \cdot \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (3.23)$$

and

$$\mathbf{\Gamma} = \begin{bmatrix} \frac{\Delta T^2}{2} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \frac{\Delta T^2}{2} \mathbf{I} \\ \Delta T \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \Delta T \mathbf{I} \end{bmatrix} \quad (3.24)$$

where,  $\mathbf{0}$  and  $\mathbf{I}$  are the zero and identity matrix of size  $\mathbb{R}^{L \times L}$ ,  $\Delta T$  is the sampling time. The segmented cell boundary is the model observation.

$$\mathbf{z}_t = \begin{bmatrix} \mathbf{z}_t^{(x)} \\ \mathbf{z}_t^{(y)} \end{bmatrix} \quad (3.25)$$

The structure of B-splines induces the measurement matrix to be in the form:

$$\mathbf{H} = \begin{bmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.26)$$

where  $\mathbf{B}$  is the basis function. In the state-space model,  $\boldsymbol{\omega}_t$  and  $\mathbf{v}_t$  are the noise presenting in the state estimation and measurement, respectively, and are assumed to be random variables from:

$$\begin{bmatrix} \boldsymbol{\omega}_t \\ \mathbf{v}_t \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right) \quad (3.27)$$

where  $\mathbf{Q}$  and  $\mathbf{R}$  are the related covariances.  $\mathbf{R}$  is set to be the identity matrix

of size  $\mathbb{R}^{2 \times L}$  and  $\mathbf{Q}$  is calculated by

$$\mathbf{Q} = \mathbb{E}(\mathbf{\Gamma} \boldsymbol{\omega}_k \boldsymbol{\omega}_k^\top \mathbf{\Gamma}^\top) = \mathbf{\Gamma} \sigma_\omega^2 \mathbf{\Gamma}^\top \quad (3.28)$$

where  $\mathbb{E}(\cdot)$  is the expectation function, the argument  $\sigma_\omega$  is selected according to the maximum magnitude of acceleration,  $a_M$ . For object tracking, practically,  $\sigma_v$  is chosen to be in the range  $[0.5a_M, a_M]$  [7].

### 3.3.2 Cell tracking with Kalman estimation and smoothing

The cell tracking is carried out by using Kalman filtering and the RTS smoothing with a population of cells. The tracking, demonstrated in Algorithm 3.1, consists of cell correspondence identification, cell tracks estimation and cell tracks smoothing. KF, an optimal Bayesian method, is applied forwards to estimate the tracks of the cell boundaries. The smoothing is performed backwards on the estimated tracks to improve the initial tracking results.

By following the Algorithm 3.1, the tracking process is performed as follows:

- In the function *Tracking\_initialisation(Image[1])*, tracking states are initialised by the LS estimated control points of cell contours in the first frame, given in (3.29), and the velocity states are initialised as zero vectors of length  $L$ . The covariances of measurement are initialised as identity matrices of size  $\mathbb{R}^{L \times L}$  and the estimation covariance is given in (3.28) with  $\sigma_v = 1$ .

$$\hat{\mathbf{x}}_{0|0} = \left[ \hat{\mathbf{q}}_0^{(x)} \quad \hat{\mathbf{q}}_0^{(y)} \quad \mathbf{0} \quad \mathbf{0} \right]^\top \quad (3.29)$$

- In the function *Predict(\cdot)*, the prediction of cell states at moment  $t + 1$ ,  $\hat{\mathbf{x}}_{t+1|t}$ , and the estimation covariance,  $\mathbf{P}_{t+1|t}$ , are calculated:

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F} \hat{\mathbf{x}}_{t|t} \quad (3.30)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F} \mathbf{P}_{t|t} \mathbf{F}^\top + \mathbf{Q} \quad (3.31)$$

- In the function *Cell\_correspondence(\cdot)*, the correspondence includes the frame to frame cell correspondence and the identical sequence of cell boundary.

Frame to frame cell correspondence is identified by the nearest-neighbour (NN) method [130]. The NN method is both simple and effective for scenarios such as this where cell speed is relatively slow compared to the frame rate [38]. In general, the NN method estimates the cell correspondence over frames based on the similarity of features with cell centroid location, which

---

**Algorithm 3.1 Cell Tracking Algorithm**

---

**Input:**Image sequence after segmentation: Image[k],  $k = 1, \dots, \text{numberOfFrames}$ **Output:**

Smoothed states of cell boundary: smoothedState[k][iC]

Cell correspondence in different frames: liveCells[k]

*// Initialisation of cell tracking using cells in the first frame*

Set ( liveCells[1], estimatedState[1] ) = Tracking\_initialisation( Image[1] )

*// Cell tracks estimation***For**  $k \leftarrow 1$  to numberOfFrames - 1  **For** iC  $\leftarrow$  liveCells[k]    Set ( predictedState, predictedCovariance ) = ...  
        ... Predict( cellState[k][iC], kinematicModel )    Set ( correspondCell, correspondBoundary ) = ...  
        ... Cell\_correspondence( predictedState, cellCentroid[k+1], Image[k+1] )    **If** correspondCell ==  $\emptyset$       Set liveCells[k+1] = liveCells[k] - iC     *// Remove the dead tracks*    **Else**      Set ( predictBoundary, measurementCovariance ) = ...  
          ... Boundary\_prediction( predictedState, kinematicModel )      Set ( estimatedState[iF+1][iC], estimatedCovariance ) = ...  
          ... Update( correspondBoundary, predictBoundary, ...  
          ... kinematicModel, predictedCovariance )    **End If**  **End For***// Check new tracks*

Set newCells = not Assigned cell at Frame k

**If** newCells  $\neq \emptyset$   Set ( liveCells[k+1], estimatedState[k+1][newCells] ) = ...  
    ... Tracking\_initialisation( Image[k+1], newCells )**End If****End For***// Cell tracks smoothing*

Set smoothedState[numberOfFrames] = estimatedState[numberOfFrames]

**For**  $k \leftarrow$  numberOfFrames - 1 to 1 in steps of -1  **For** iC  $\leftarrow$  liveCells[k]

Set smoothedState[k][iC] = State\_smooth( smoothedState[k+1][iC] )

**End For****End For**

---

can be directly calculated from the control points of B-spline contours, given in:

$$\mathbf{r}_t = \begin{bmatrix} r_t^{(x)} \\ r_t^{(y)} \end{bmatrix} = \begin{bmatrix} \langle \mathbf{1}, \mathbf{q}_t^{(x)} \rangle \\ \langle \mathbf{1}, \mathbf{q}_t^{(y)} \rangle \end{bmatrix} \quad (3.32)$$

where  $\langle, \rangle$  represents the inner product and  $\mathbf{1} \in \mathbb{R}^L$ . Here, the cell centroid distance,  $D(\mathbf{r}_t)$ , is calculated between the estimated centroid at frame  $t + 1$  given  $t$  and cell centroid at frame  $t + 1$ , is used as the feature distance, given by:

$$D(\mathbf{r}_t) = \sqrt{(r_{t+1}^{(x)} - \hat{r}_{t+1|t}^{(x)})^2 + (r_{t+1}^{(y)} - \hat{r}_{t+1|t}^{(y)})^2} \quad (3.33)$$

The cell in validated measurements, which has the minimum centroid distance is designed as the corresponding cell as explained by

$$\mathbf{r}_t^* = \arg \min_{\mathbf{r}_t \in \mathcal{R}_t} D(\mathbf{r}_t) \quad (3.34)$$

where  $\mathcal{R}_t$ , expressed in (3.35), is a set of validated measurements of which the feature distance is below the threshold  $\epsilon_d$ .

$$\mathcal{R}_t = \{\mathbf{r}_{t+1} : D(\mathbf{r}_t) \leq \epsilon_d\} \quad (3.35)$$

On the other hand, the cells at frame  $t$  with empty  $\mathcal{R}_t$  are considered to either be dead or have moved outside the frame. At frame  $t + 1$ , if any cell is not linked to cells in the previous frame, they are assumed to be the start of new tracks.

After the identification of frame to frame cell correspondence, cell states correspondence is determined by maintaining the starting point of constructed B-spline contours using the Bresenham line algorithm [22]. Firstly, the cell contour at frame  $t$  is shifted to the centroid location where the corresponding cell is at frame  $t + 1$ . Figure 3.5a demonstrates the shift of B-spline contours, of which the start point is labelled by a yellow asterisk and calculated by

$$\begin{aligned} \begin{bmatrix} u_{t,1}^{(x)} \\ u_{t,1}^{(y)} \end{bmatrix} &= \begin{bmatrix} B_1(s) & \mathbf{0} \\ \mathbf{0} & B_1(s) \end{bmatrix} \begin{bmatrix} q_{t,1}^{(x)} \\ q_{t,1}^{(y)} \end{bmatrix} \\ &= \begin{bmatrix} s^0 & s & \dots & s^{d-1} & s^0 & s & \dots & s^{d-1} \end{bmatrix} \dots \\ &\dots \begin{bmatrix} \mathbf{B}_1 \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_1 \mathbf{G}_1 \end{bmatrix} \begin{bmatrix} q_{t,1}^{(x)} \\ q_{t,1}^{(y)} \end{bmatrix} \end{aligned} \quad (3.36)$$



where  $(u_{t,1}^{(x)}, u_{t,1}^{(y)})$  is the Cartesian coordinate of start point,  $\mathbf{B}_1$  and  $\mathbf{G}_1$  are the span matrix and placement matrix of the basis function at first span,  $B_1(s)$ . Due to the calculation of the starting point, the parameter  $s$  set to 0 and  $(q_{t,1}^{(x)}, q_{t,1}^{(y)})$  is the first point of the estimated control points. The tangent of contour is calculated by

$$\begin{bmatrix} u_{t,1}^{(x')} \\ u_{t,1}^{(y')} \end{bmatrix} = \begin{bmatrix} 0 & 1 & \dots & (d-1)s^{d-2} & 0 & 1 & \dots & (d-1)s^{d-2} \end{bmatrix} \dots \\ \dots \begin{bmatrix} \mathbf{B}_1 \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_1 \mathbf{G}_1 \end{bmatrix} \begin{bmatrix} q_{t,1}^{(x)} \\ q_{t,1}^{(y)} \end{bmatrix} \quad (3.37)$$

so that the normal vector is

$$\begin{bmatrix} u_{t,1}^{(x^\perp)} \\ u_{t,1}^{(y^\perp)} \end{bmatrix} = \begin{bmatrix} -u_{t,1}^{(y')} \\ u_{t,1}^{(x')} \end{bmatrix} \quad (3.38)$$

The normal line is then generated from the shifted start point on both positive and negative directions using the Bresenham line algorithm:

$$U^\perp(\eta) = \begin{bmatrix} u_{t,1}^{(x)} \\ u_{t,1}^{(y)} \end{bmatrix} + \Delta r \pm \eta \cdot \frac{u_{t,1}^{(y^\perp)}}{u_{t,1}^{(x^\perp)}} \quad (3.39)$$

where  $\eta$  is the length of the normal line and increases until the two joint points with segmented boundary at frame  $t+1$  are found, as illustrated in Figure 3.5b. The first joint point is shifted to the start of the segmented boundary at frame  $t+1$ , which is considered as the measurement at next frame,  $\mathbf{z}_{t+1}$ .

- In the function *Boundary\_prediction*( $\cdot$ ), the prediction of cell boundaries (3.40) and the related measurement covariance (3.41) are calculated:

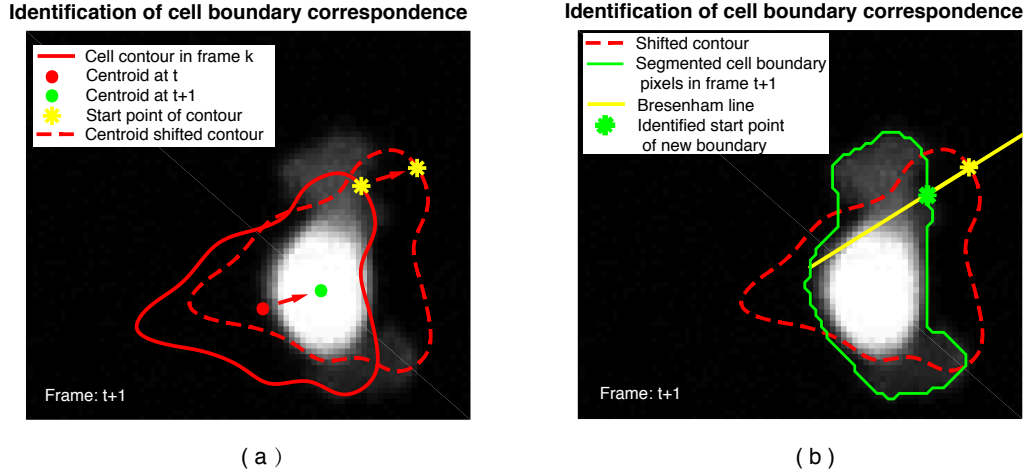
$$\hat{\mathbf{z}}_{t+1|t} = \mathbf{H} \hat{\mathbf{x}}_{t|t} \quad (3.40)$$

$$\mathbf{S}_{t+1} = \mathbf{H} \mathbf{P}_{t+1|t} \mathbf{H}^\top + \mathcal{R} \quad (3.41)$$

- In the function *Update*( $\cdot$ ), the prediction of tracking states and the estimation covariance are amended through the Kalman gain as follows[160]:

$$\mathbf{K}_{t+1} = \mathbf{P}_{t+1|t} \mathbf{H}^\top \mathbf{S}_{t+1}^{-1} \quad (3.42)$$

$$\hat{\mathbf{x}}_{t+1|t+1} = \hat{\mathbf{x}}_{t+1|t} + \mathbf{K}_{t+1} (\mathbf{z}_{t+1} - \hat{\mathbf{z}}_{t+1|t}) \quad (3.43)$$



**Figure 3.5: Illustration of cell contour correspondence.** (a) Shift the constructed B-spline contours at frame  $t$  to the location where cell centroid at frame  $t + 1$  is. The red solid boundary is the B-spline contour, constructed according to the estimated control points at frame  $t$ . Red point is the centroid of the red contour calculated from (3.32) and is shifted to the location of green point, which is the centroid of the green boundary shown in subfigure (b). The B-spline contour is, therefore, shifted to the dotted red outline. The yellow asterisk is the start point of the B-spline contour. (b) Identify the start point of segmented cell boundary at frame  $t + 1$ . The normal line (yellow line) starts extending from the yellow asterisk on both positive and negative directions until that two joint points are identified between the yellow line and the green boundary. The joint point, which is first identified, is considered as the start location of green boundary.

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|k+1} - \mathbf{K}_{t+1} \mathbf{S}_{t+1} \mathbf{K}_{t+1}^\top \quad (3.44)$$

- In the function of  $State\_smooth(\cdot)$ , a fixed-interval smoother is applied afterwards in the backward direction of the time-line to minimise the errors that are introduced in the tracking system because of the random initialisation [124]. The smoother is initialised by the cell states estimation at the last frame  $T$ . A smoother gain is calculated in (3.45) and is applied for the smoothing states in (3.46) and the smoothing covariance given in (3.47).

$$\mathbf{C}_t = \mathbf{P}_{k|k} \mathbf{F}_t^\top \mathbf{P}_{t+1|t}^{-1} \quad (3.45)$$

$$\hat{\mathbf{x}}_{t|T} = \hat{\mathbf{x}}_{t|t} + \mathbf{C}_t (\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}) \quad (3.46)$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{C}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{C}_t^\top \quad (3.47)$$

### 3.4 Cell motility analysis

In the tracking scheme, the coordinates and velocities of cell centroid locations and velocities can be calculated through the inner product of control points and vector of ones.

$$\begin{bmatrix} r_t^{(x)} \\ r_t^{(y)} \\ \Delta r_t^{(x)} \\ \Delta r_t^{(y)} \end{bmatrix} = \begin{bmatrix} \langle \mathbf{1}, \hat{\mathbf{q}}_{t|t}^{(x)} \rangle \\ \langle \mathbf{1}, \hat{\mathbf{q}}_{t|t}^{(y)} \rangle \\ \langle \mathbf{1}, \Delta \hat{\mathbf{q}}_{t|t}^{(x)} \rangle \\ \langle \mathbf{1}, \Delta \hat{\mathbf{q}}_{t|t}^{(y)} \rangle \end{bmatrix} \quad (3.48)$$

where  $r_t = (r_t^{(x)}, r_t^{(y)})$  is the Cartesian coordinate of cell centroid at frame  $t$  and  $\Delta r_t = (\Delta r_t^{(x)}, \Delta r_t^{(y)})$  is the corresponding velocity.  $\mathbf{1} \in \mathbb{R}^L$  is a vector of ones, and  $\langle \cdot, \cdot \rangle$  is the inner product for spline functions [19], and it is calculated by

$$\langle \mathbf{1}, \hat{\mathbf{q}} \rangle = \mathbf{1}^\top \left( \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{G}_l^\top \mathbf{B}_l^\top \mathcal{P} \mathbf{B}_l \mathbf{G}_l \right) \hat{\mathbf{q}} \quad (3.49)$$

where  $L$  is the number of spans,  $\mathbf{B}_l$  and  $\mathbf{G}_l$  are the span and placement matrix of basis function at span  $l$ .  $\mathcal{P}$  is

$$\mathcal{P} = \int_0^1 \begin{bmatrix} 1 \\ \vdots \\ s^{d-1} \end{bmatrix} \begin{bmatrix} 1 & \dots & s^{d-1} \end{bmatrix} ds \quad (3.50)$$

Discretely,  $\mathcal{P}$  is constructed as a 'Hilbert' matrix [19], of which the coefficients are

$$\mathcal{P}_{ij} = \frac{1}{i+j-1}. \quad (3.51)$$

In addition, the definition of B-spline active contours can be used to calculate cell boundaries through the estimated states of control points shown as follows:

$$\begin{bmatrix} \mathbf{u}_t^{(x)} \\ \mathbf{u}_t^{(y)} \\ \Delta \mathbf{u}_t^{(x)} \\ \Delta \mathbf{u}_t^{(y)} \end{bmatrix} = \begin{bmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B} \end{bmatrix} \hat{\mathbf{x}}_{t|t} \quad (3.52)$$

where  $\mathbf{u}_t = (\mathbf{u}_t^{(x)}, \Delta \mathbf{u}_t^{(y)})$  is the Cartesian coordinate of cell contour at frame  $t$ .  $\Delta \mathbf{u}_t = (\Delta \mathbf{u}_t^{(x)}, \Delta \mathbf{u}_t^{(y)})$  is the corresponding velocity.

In this section, cell motility analysis is carried out based on cell centroids and

cell boundary motility features separately.

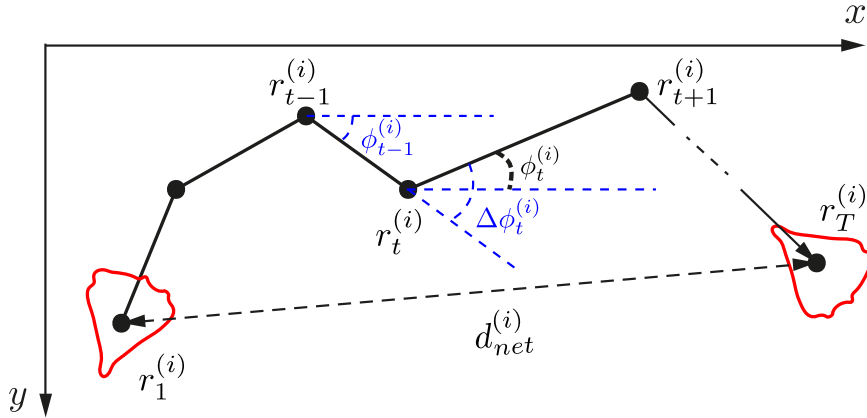
### 3.4.1 Cell centroid motility features

Cell migration length, confinement ratio and migration angle change can be calculated from cell centroid tracks for the quantitative analysis of cell global behaviours.

The centroid trajectory of cell ( $i$ ) is shown in Figure 3.6 [100]. The cell migration length is defined by,

$$d_{tot}^{(i)} = \sum_{t=1}^{T-1} \|r_t^{(i)}, r_{t+1}^{(i)}\| \quad (3.53)$$

where,  $\|\cdot\|$  denotes the Euclidean distance,  $T$  is the number of frames included in the trajectory.



**Figure 3.6: An example of cell centroid trajectory.** The trajectory starts from  $t = 1$  to  $t = T$ .  $r_t^{(i)}, t \in [1, T]$  are cell centroid locations.  $\phi_t^{(i)}$  is the migration angle at frame  $t$  and  $\Delta\phi_t^{(i)}$  is the migration angle change at frame  $t$ .

The confinement ratio is defined as the ratio of cell migration displacement,  $d_{net}^{(i)}$ , to the travelled distance,  $d_{tot}^{(i)}$ , given by

$$C_r^{(i)} = \frac{d_{net}^{(i)}}{d_{tot}^{(i)}} = \frac{\|r_1^{(i)}, r_T^{(i)}\|}{\sum_{t=1}^{T-1} \|r_t^{(i)}, r_{t+1}^{(i)}\|} \quad (3.54)$$

The confinement ratio is a measurement of the straightness of cells' migration. When  $C_r^{(i)} = 1$ ,  $d_{net}^{(i)} = d_{tot}^{(i)}$ . This tells that the cell travels in a straight line. Oppo-

sitely,  $C_r^{(i)} = 0$  indicates that  $d_{net}^{(i)} \ll d_{tot}^{(i)}$ . The cell migration is restrained in a small area so that the cell moves around in the area with small displacement, but large distance [61].

The cell directional change,  $\Delta\phi_t$ , is a motility feature that is associated with cell orientation and is calculated via [100],

$$\Delta\phi_t = \phi_t - \phi_{t-1} \quad (3.55)$$

where  $\phi_t$  is cell migration angle at frame  $t$  and is computed by

$$\phi_t = \arctan \left( \frac{r_{t+1}^{(y)} - r_t^{(y)}}{r_{t+1}^{(x)} - r_t^{(x)}} \right) \quad (3.56)$$

### 3.4.2 Cell boundary motility features

The B-spline active contour representation of cell boundaries allows the measurement of cell boundary local motility in the same framework with the measurement of cell centroid motility. The orientation and protrusion distribution around cell boundary are quantified in this section using cell boundary location and velocity calculated in (3.48).

The algorithm that measures cell boundary orientation is illustrated in Figure 3.7a. For cell ( $i$ ) at frame  $t$ , the tracking results of cell boundary location,  $\mathbf{u}_t^{(i)}$ , and velocity,  $\Delta\mathbf{u}_t^{(i)}$ , are in the form:

$$\mathbf{u}_t^{(i)} = \left[ u_{t,1}^{(i)} \quad u_{t,2}^{(i)} \quad \cdots \quad u_{t,M}^{(i)} \right]^\top \quad (3.57)$$

$$\Delta\mathbf{u}_t^{(i)} = \left[ \Delta u_{t,1}^{(i)} \quad \Delta u_{t,2}^{(i)} \quad \cdots \quad \Delta u_{t,M}^{(i)} \right]^\top \quad (3.58)$$

where  $M$  is the number of points used to construct the cell boundary and  $u_{t,m}^{(i)}$  is  $m$ -th point on the boundary with velocity  $\Delta u_{t,m}^{(i)}$  ( $m \in [1, M]$ ). The radius vector of cell ( $i$ ) at frame  $t$ ,  $\boldsymbol{\rho}_t^{(i)}$ , is first calculated using,

$$\boldsymbol{\rho}_t^{(i)} = \mathbf{u}_t^{(i)} - r_t^{(i)} \quad (3.59)$$

Next, the angles,  $\boldsymbol{\delta}_t^{(i)}$ , between the radius vectors and corresponding velocity angles are calculated in order to identify the point at the front,  $\bar{u}_t^{(i)}$ , referred to as the front point, where the radius vector represents the most likely direction for cell movement in the next image frame.

$$\boldsymbol{\delta}_t^{(i)} = \left[ \delta_{t,1}^{(i)} \quad \delta_{t,2}^{(i)} \quad \cdots \quad \delta_{t,M}^{(i)} \right] = \angle(\boldsymbol{\rho}_t^{(i)} - \Delta\mathbf{u}_t^{(i)}) \quad (3.60)$$

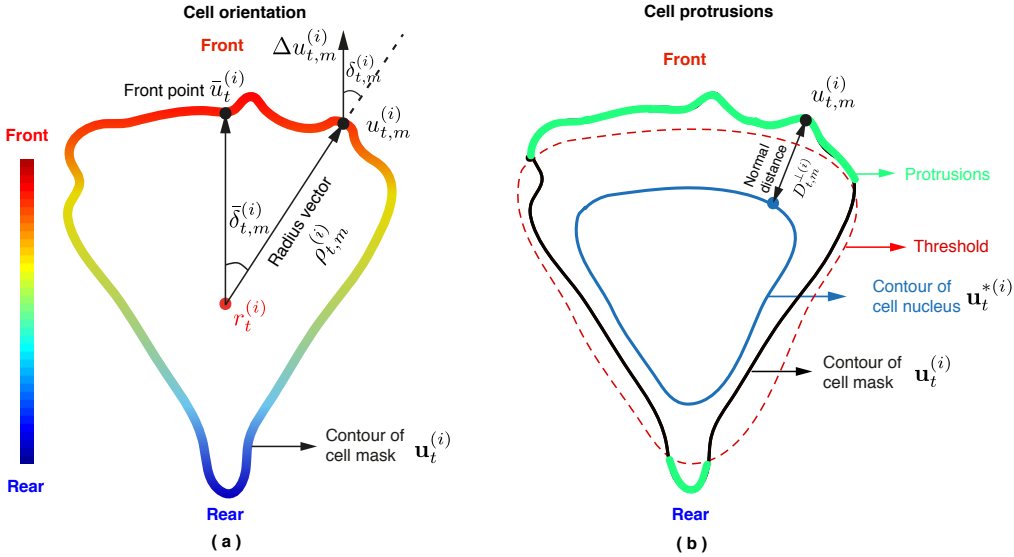
The front point, is then detected via a search to find the minimum  $\delta_t^{(i)}$ ,

$$\bar{u}_t^{(i)} = \arg \min_{u_t^{(i)} \in \mathbf{u}_t^{(i)}} \delta_t^{(i)} \quad (3.61)$$

The corresponding radius vector  $\bar{\rho}_t^{(i)}$  at the same point is regarded as the reference location. Finally, the cell orientation is measured by the angle difference to a reference, illustrated in

$$\bar{\delta}_{t,m}^{(i)} = \angle(\rho_{t,m}^{(i)} - \bar{\rho}_t^{(i)}) \quad m \in [0, M] \quad (3.62)$$

A smaller  $\bar{\rho}_{t,m}^{(i)}$  indicates that the point is closer to the cell migration front, whereas a larger value indicates that the point is closer to the side opposite to cell migration direction, referred to as cell rear.



**Figure 3.7: Illustration of cell orientation and protrusion identification.** (a) Orientation identification of cell  $(i)$  at frame  $t$ . Cell boundary orientation is represented by a RGB colour map. With the increase of  $\bar{\delta}_{t,m}^{(i)}$ , cell boundary starts from the front (red) to the rear (blue). (b) Protrusions identification of cell  $(i)$  at frame  $t$ . The protrusions, labelled by green lines, are identified where the normal distance,  $D_{t,m}^{\perp}$ , is larger then the protrusion threshold.

Under the florescence microscopy, protrusions are generated around the cell boundary and are represented as light grey areas due to their cytoplasm thin structure. In the segmentation step, the cell body is separated into the mask and the nucleus area. Within the tracking process, these two boundary types are tracked

separately. The protrusion locations are then identified via the normal distance between these two contours. For each point along the cell mask contour, the closest distance between it and the corresponding nucleus contour is calculated as the normal distance,  $D_{t,m}^{\perp(i)}$ , and the location of protrusions are identified by going through every point on the mask contour. If  $D_{t,m}^{\perp(i)}$  is larger than a threshold,  $\epsilon_{protrusion}$ ,  $u_{t,m}^{(i)}$  is the protrusion points. The protrusion identification of cell ( $i$ ) at frame  $t$  is illustrated in Figure 3.7b.

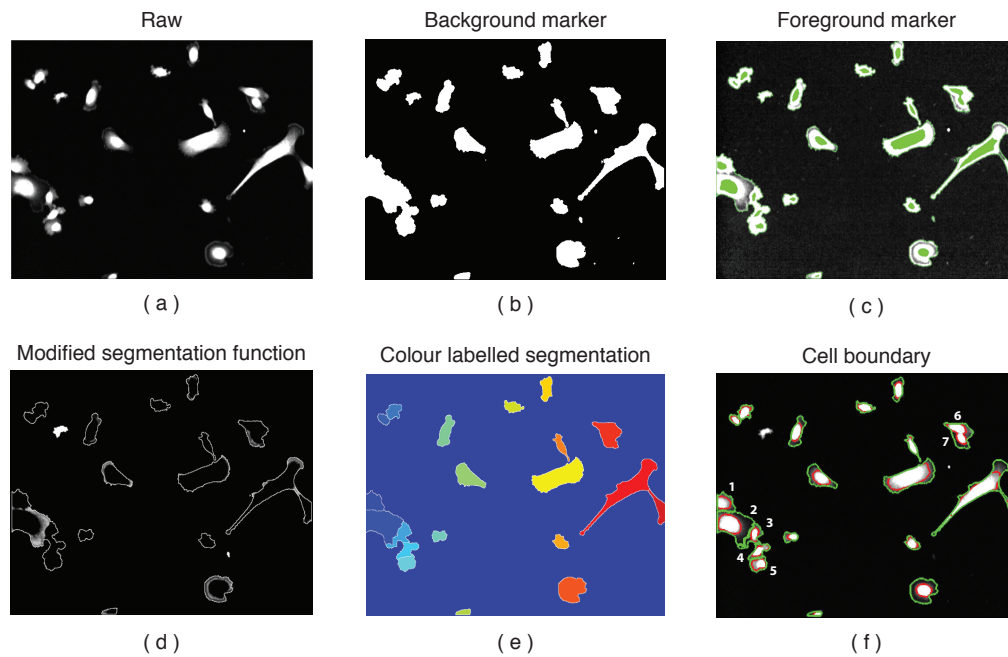
### 3.5 Breast cancer migration

The aforementioned algorithms are applied to a set of MDA-MB-231 breast cancer cells, described in Table 3.1. This section will present the results of segmentation, tracking and cell motility analysis.

#### 3.5.1 Cell segmentation results

The marker-controlled watershed segmentation is demonstrated in Figure 3.8. Figure 3.8a gives a grey-scale frame extracted from the original image sequence. By following the procedures outlined in Section 3.1, after the frame contrast enhancement, the background marker is obtained using global intensity thresholding and set to black area, shown in Figure 3.8b. The cell foreground markers, which indicates the location of cells, are coloured green in Figure 3.8c. In Figure 3.8d, both marker types are set to be minus infinity so that the water lines of the watershed algorithm are initialised from these areas. After the watershed algorithm is carried out, the segmented areas are marked with different colours, shown in Figure 3.8e. For each segmented cell area, intensity thresholding algorithm is used to segment the cell nucleus, surrounded by the red contours, and entire cell masks, surrounded by the green contours (see Figure 3.8f). These boundary segmentations demonstrate that the marker-controlled watershed algorithm can identify cells preventing combining the cells that are closely contacting (i.e., cells: 1– 5), but the algorithm cannot segment the cells, whose nucleus are overlapped (i.e., cells: 6 and 7).

Since we used the Otsu's method for automated thresholding, the segmentation algorithm can be applied to frames with different pixel intensities. Figure 3.9 illustrates the effectiveness of the segmentation algorithm when applied to images of different contrast. The subfigures on the left column are the raw images and those on the right column correspond to the respective segmentation results. Figure 3.9a and Figure 3.9b demonstrate that the segmentation algorithm is capable of segmenting the images, where cells have different shape types to those shown



**Figure 3.8: Image segmentation process.** (a) The grey-scale raw frame. (b) The identified background marker (the black area). (c) The foreground markers (green areas). (d) The modified segmentation function, where the background and foreground markers are set to minus infinity (i.e. black colour). (e) The colour labelled segmentation; each colour indicates a separate segmented area. (f) The segmented cell boundaries; the green and the red contours represent the boundaries of cell masks and nucleus, respectively.



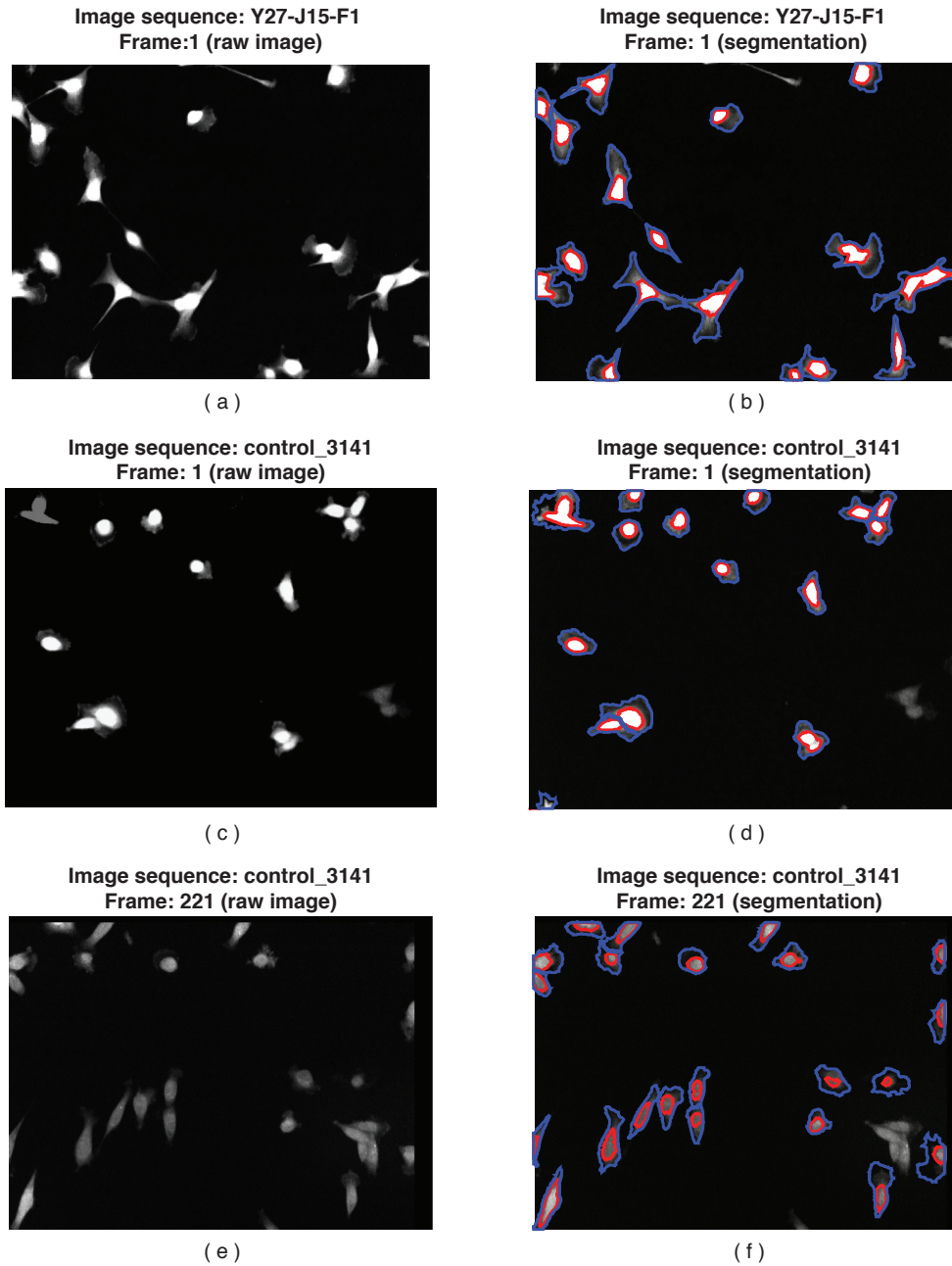
in Figure 3.8a. Additionally, frame 1 and frame 221 belonging to the same image sequence are shown in Figure 3.9c and Figure 3.9e, respectively. However, the image quality is reduced significantly from frame 1 to frame 221. This phenomenon is because the fluorochrome molecules lose the fluorescence due to the exposure to light, known as photobleaching [143]. The associated segmentation results are shown in Figure 3.9d and Figure 3.9f.

### 3.5.2 Cell tracking results

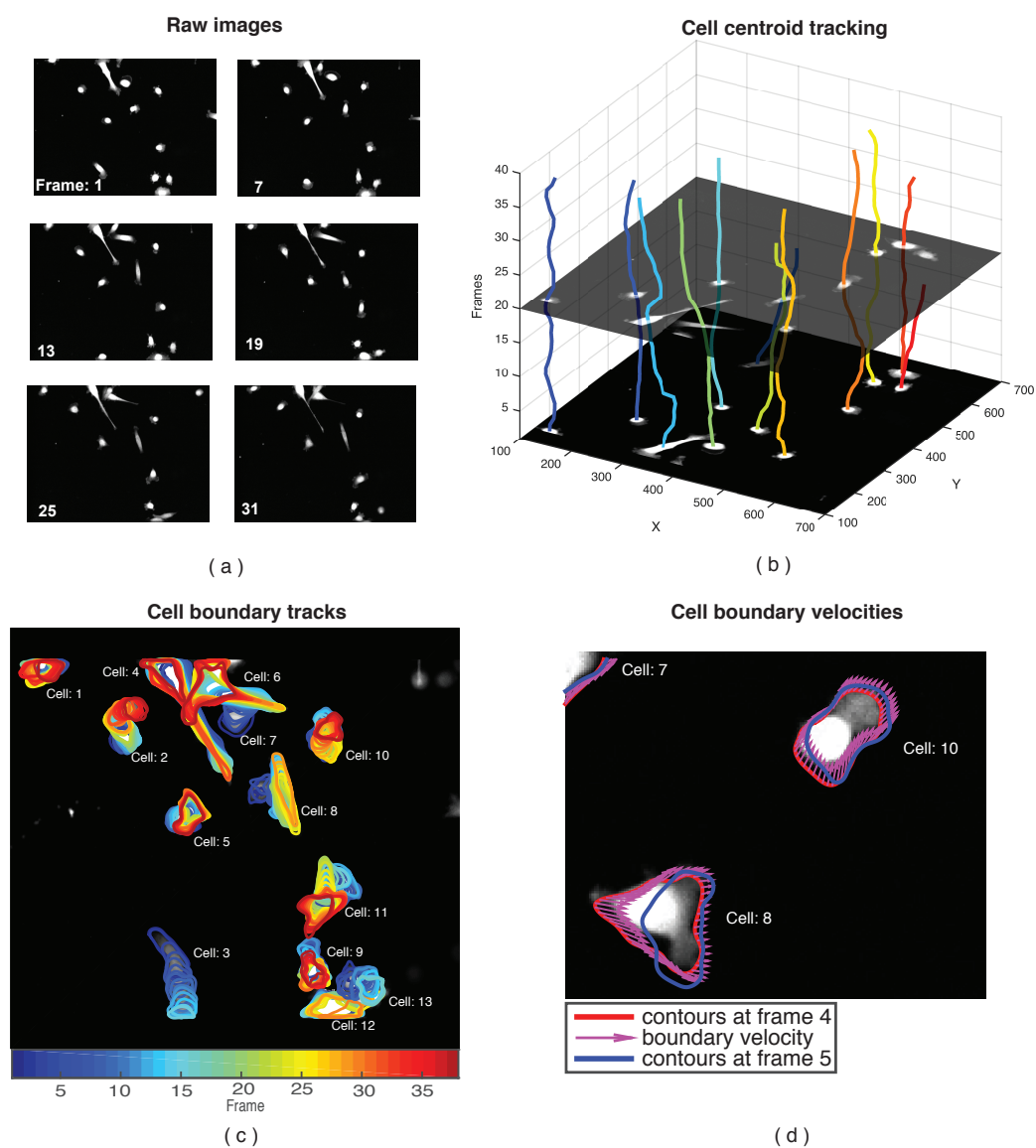
Section 3.2 and Section 3.3 described how B-spline active contours are used to represent the segmented cell boundaries and a DWNA kinematic model is then used to track the corresponding control points using the Kalman filter and smoother in order to describe the cell boundary dynamics. Invalid trajectories, such as those from disturbances, are removed by only considering trajectories that are longer than 5 frames (75 minutes).

Figure 3.10 presents the tracking results of the control data set 'J21-F1'. Six frames extracted from the raw image sequence are provided in Figure 3.10a. These images show that cell travelling distances between two adjacent frames are much small compared to the average size of cell shapes (i.e., radius of cell shape) and the cells show shape variety during migration. Cell centroid trajectories and boundary trajectories are shown in Figure 3.10b and Figure 3.10c, respectively. 10 of 12 cell centroid tracks starting from the first frame separately go through the same cell at frame 20. One cell track (cell: 3) stops earlier because the cell travelled out of the view. Another track (cell: 13) stops due to segmentation errors where the cell body was recognised as part of cell 12 at frame 16. In addition, Figure 3.10d shows an example of tracking cell boundary velocities.

The tracking errors are detected from visual inspection. Tracking inaccuracies arise from two possible scenarios: i) wrong trajectory caused by the limitation of cell correspondence identification using NN algorithm; ii) other errors that are generated in the cell segmentation step. The former occurs when tracking cells which cross paths can lead to the two cell trajectories being confused. In the latter case, cells with significant overlap can be incorrectly interpreted as a single cell in the segmentation step. For statistical analysis after tracking, such as the speed histogram, migration angle and cell orientation identification, tracking errors only affect the data of single cell for one frame and it is not capable of altering the overall statistical trends. However, such mistakes can affect the analysis that focus on single cell and is relevant to time. For example, the confinement ratio may change substantially with incorrect trajectory.



**Figure 3.9:** Examples of the image segmentation algorithm working on the images with cell shape variety and different image quality. (a) The cell shape diversity has been increased compared to Figure 3.8a. (c) and (d) are frame 1 and frame 221, extracted from the same image sequence. (b), (d) and (f) show the corresponding segmentation results of subfigures (a), (c) and (e), respectively.



**Figure 3.10: Tracking of cell trajectories and velocities.** (a) The raw images of frames 1, 7, 13, 19, 25, 31 of cell line 'J21-F1'. (b) Cell centroid trajectories; different cell tracks are indexed by lines of different colours. (c) Cell boundary trajectories alongside overlapping the raw images; colour is used as time index; these cells migrated from the blue to the red contours. (d) The boundary velocities of cell 8 and cell 10 at frame 4.

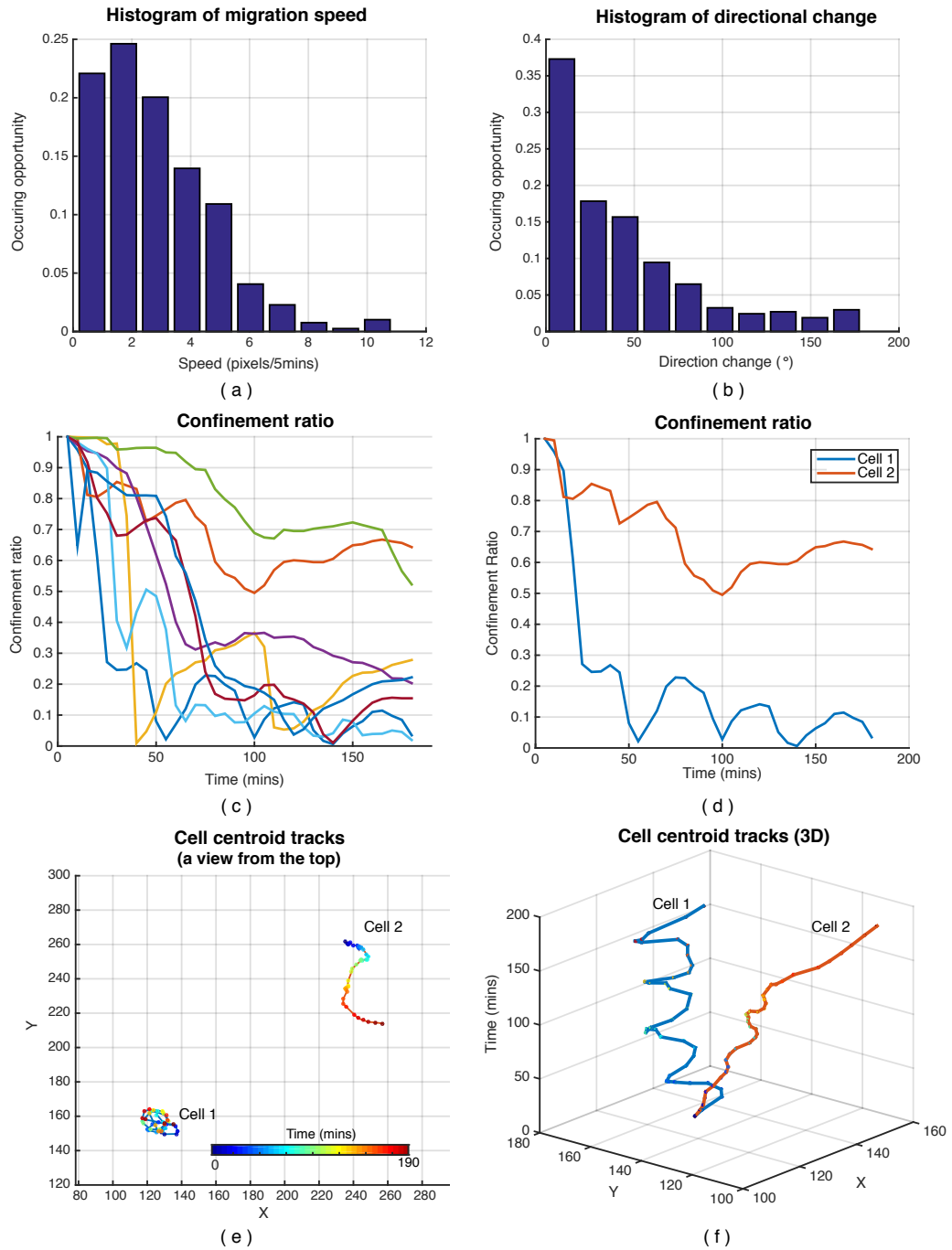
### 3.5.3 Statistical analysis of cell motility

The results of centroid motility analysis is presented in Figure 3.11. The histogram of migration speed is shown in Figure 3.11a and it indicates that those cells cannot move faster than 10 pixels in 5 minutes (i.e., time interval between two frames) and over 80% cells migrated slower than 4 pixels in 5 minutes. Figure 3.11b is the histogram of cell directional changes and shows that cells had the highest occurrence of maintaining the migration direction (i.e., the directional changes were smaller than  $15^\circ$ ) when moving to the next frame. Moreover, about 90% of cells migrated with a directional change smaller than  $90^\circ$ , which indicates that 5 minutes was not long enough for these cells to move in the opposite direction. Figure 3.11c shows the confinement ratios, calculated from (3.54), of cells, that presented throughout the duration of the video. Figure 3.11d shows the comparison of two cells with significantly different confinement ratios, where cell 1 had larger confinement ratios than cell 2. The corresponding cell trajectories of a bird's-eye view (see Figure 3.11e) and a 3D view (see Figure 3.11f) indicate that cell 1 migration was confined to a small area while cell 2 maintained its direction.

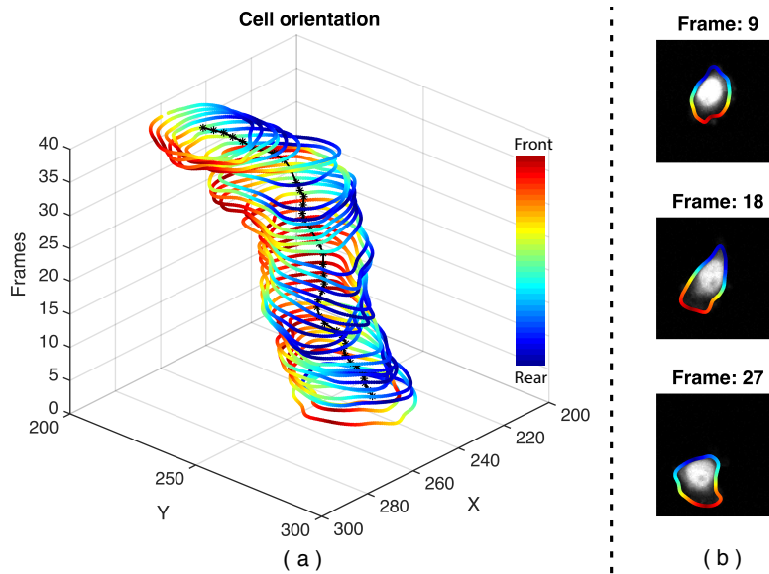
Figure 3.12 provides an example of the orientation identification (method is seen in Section 3.4.2). The cell orientation throughout the video are shown in Figure 3.12a and the identified front areas are consistent with the cell centroid migration directions. Three selected frames are extracted along with the original grey-scale images are shown in Figure 3.12b and illustrates that this algorithm of orientation identification is capable of being identical to the visual inspection.

The effect of protrusion identification is illustrated in Figure 3.13 (method is provided in Section 3.4.2). The protrusion identification of a frame randomly extracted from the image sequence J21-F1 and the protrusions of a cell at different frames are shown in Figure 3.13a and Figure 3.13b, respectively. The identified protrusion locations connect to the light grey cell areas exhibiting large and broad shapes, which are generally characterised as protrusions [127]. This indicates that the protrusion identification is capable of showing consistency with observations.

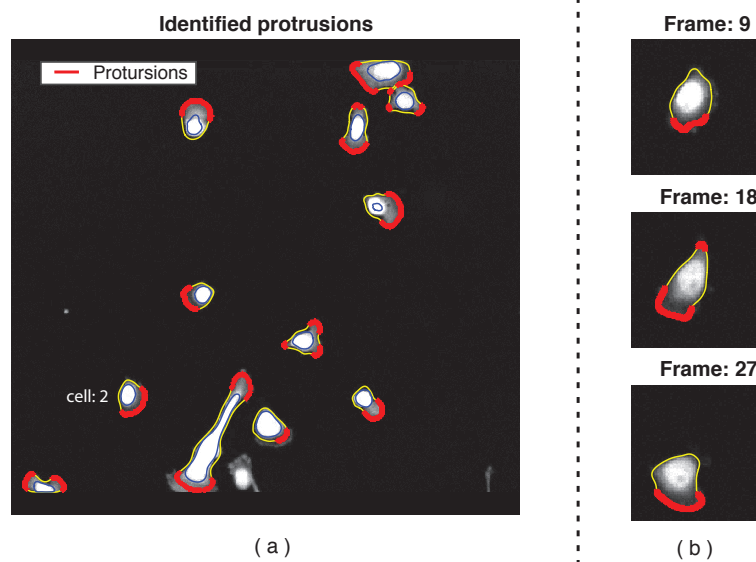
After the protrusions are located on the cell boundary, the length of protrusive boundary are normalised against the entire length of cell boundary and are categorised according to cell orientations. The results are expressed as a polar histogram, shown in Figure 3.14. It is a mirror image relative to the polar axis that the histogram in the range  $[0^\circ, -180^\circ]$  is mirrored of the part in the range  $[0^\circ, 180^\circ]$ . The advantage of using the symmetric polar histogram is the similarity between the shape and closed cell contours, which is convenient for understanding the behaviours of cells. Additionally, the background colour is indicative of the orientation angle amplitude. The polar histogram in Figure 3.13 gives an average



**Figure 3.11: Analysis of cell centroid motility.** (a) Speed histogram of cell centroid migration. (b) Histogram of cell centroid migration directional change. (c) Confinement ratios of cells, which existed over 175 minutes (i.e., 35 frames). Each line represents the confinement ratios over time of a single cell. (d) The confinement ratios of two cells with significantly different values. (e) and (f) are the bird's-eye view and the 3D view, respectively, of the cells with the confinement ratios shown in subfigure (d).

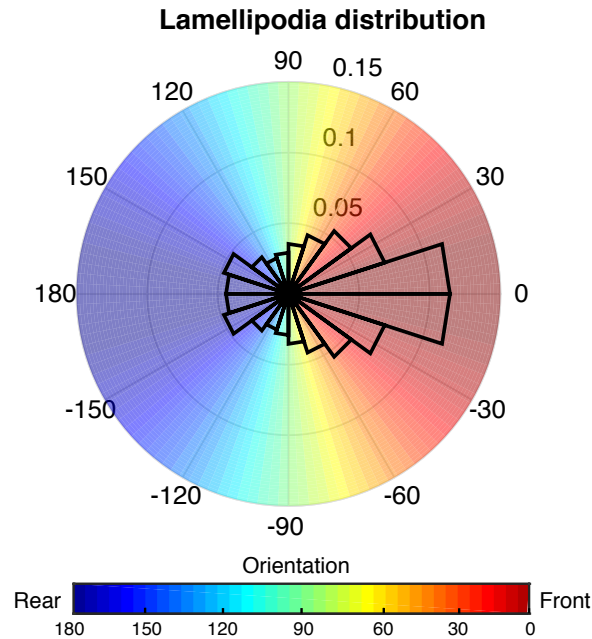


**Figure 3.12: An example of cell orientation identification.** (a) The identified cell orientation throughout cell migration video; colours are used as the index of orientation that cell front and rear areas are marked by red and blue, respectively, and the cell centroid trajectory is linked by black line. (b) Frame: 9, 18 and 27 are extracted from subfigure (a) along with the grey-scale cell image.



**Figure 3.13: Example of cell protrusion identification.** (a) The identified protrusion locations at the first frame of the image sequence of cell line J21-F1; the tracked cell contours are shown in yellow and the allocated protrusions are in red. (b) The protrusion locations of cell 2 at frames: 9, 18 and 27.

lamellipodia distributions of cells in cell line: J31-F1. Explicitly, a greater number of protrusions are located in the front area compared to other orientations and more protrusions are located in the rear area compared to the both sides of the cells.



**Figure 3.14: Polar histogram of protrusion distributions associated with cell orientation (J21-F1).** The polar histogram in  $[0^\circ, -180^\circ]$  is mirrored of  $[0^\circ, 180^\circ]$  and cell orientation is indexed by the colours the same with Figure 3.12 that red area represents the front and when the colour close to blue the area is close to cell rear.

### 3.6 Conclusion

In this chapter, a novel tracking framework was presented alongside a statistical analyse of cell motility features. The system was demonstrated by applying it to a grey-scale time-lapse image sequence of breast cancer cells migrating.

The tracking framework consists of cell segmentation, cell boundary representation and cell dynamic tracking modules. The marker-controlled watershed segmentation was used for automated cell segmentation. This algorithm can segment the closely contacting cells and overcome over-segmentation, which is a common problem with the watershed algorithms. B-spline active contours were then applied to represent the segmented cell boundaries as a linear model with lower dimensional parameters (control points). Accordingly, a DWNA kinematic model

was constructed to track the locations and velocities of the control points. Cell correspondence was identified using the NN algorithm. KF estimation and smoothing were performed individually from forward and backward time lines to achieve the multiple cells tracking.

Cell centroid and boundary dynamics can be calculated from the tracked control points. The centroid dynamics were computationally analysed by the speed histogram, the histogram of migration directional changes and the confinement ratios. Cell boundary motility was quantified by cell protrusion distributions associated with cell orientations and was presented as a normalised polar histogram.

The proposed tracking framework was applied to the breast cancer cell line J21-F1. The cell segmentation, centroid and boundary tracks, confinement ratio analysis, protrusion identification and cell orientations showed the consistency with the visual inspection. The statistical analysis of cell protrusions associated with orientations indicated that, for the tracked cells, more protrusions distributed in the front of the cell. This analysis provided a way to quantify cell motility, therefore, to compare the motility of different cells or cell lines, which is essential for the drug testing and the study of unknown cells.

In summary, the presented tracking framework can simultaneously track the cell boundary and centroid dynamics. The tracking results provided a basis for further morphology and motility analysis, which is the focus of Chapter 4.



## Chapter 4

# A Computational Module for Cell Morphology Analysis

Cell morphology is affected by both intracellular and extracellular environment changes. The relevant parameters to cell morphology, such as cell area, shape size, boundary and shape polarisation, have been analysed with a view to help explore cell membrane structure [77], cell-to-cell interactions [140], morphology associated signalling networks [6, 138] and functional gene expression [6, 132]. More recently, discrete cell shape classes and the shape transitions have been observed [132, 174]. However, there has been little study linking cell motility behaviours to different shapes. This is important because when cells transition to a different state, cells generally change morphology, as well as motility characteristics. For example, when epithelial cells change to mesenchymal stem cells, the apical-basal polarised cell shapes morph into spindle-like shapes and the tight jointed cells become isolated and invasive [83]. There are currently no methods for computationally analysing how dynamics are linked to shapes. Therefore, in this chapter, a novel computational module for cell morphology analysis is proposed that for the first time extends analysis of shape to morphodynamics, i.e., analyses how the cell shape evolves over time and also how cell shapes are linked to motility.

The morphology analysis framework is built on the basis of boundary tracking results (see Chapter 3) and its schematic diagram is given in Figure 4.1. The analysis consists of cell shape description detailed in Section 4.1, shape clustering detailed in Section 4.2, and integrated motility and morphology analysis and morphodynamics analysis presented in Section 4.3. Section 4.4 demonstrates the application of the morphology analysis module to the datasets of wild breast cancer cells.

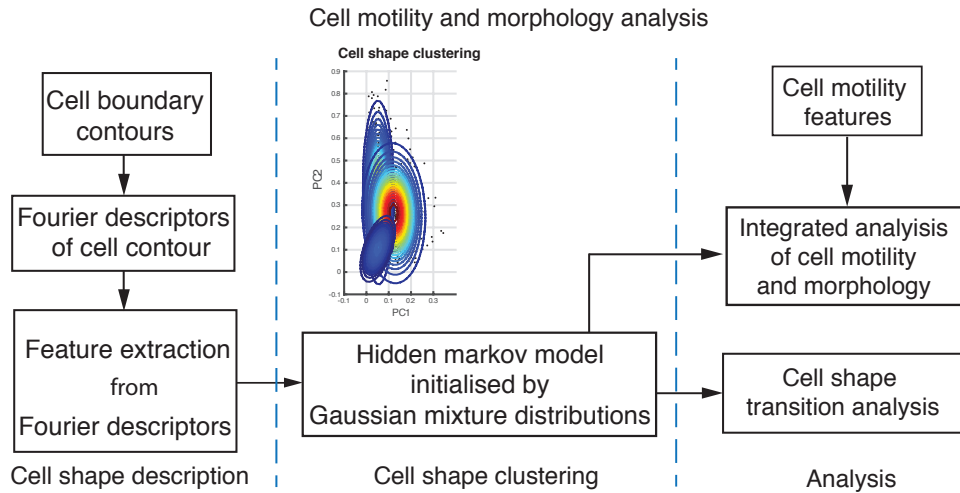


Figure 4.1: Flowchart of cell morphology analysis

## 4.1 Cell shape description

This session introduces the techniques used for cell shape description. With the estimated position of control points and cubic basis functions, cell boundary shapes can be constructed through the definition of B-spline active contours. FDs are applied to capture the shape features in frequency domain and has the following benefits [179].

- FDs have the corresponding physical explanation.
- FDs involve both local and global shape features since it covers components of different frequencies.
- FDs are easy normalised, namely invariance to shape transposition, scale and rotation, which simplifies the process of shape matching.
- FDs are more robust to noise and boundary variations because of its representation in the spectral domain.

The constructed FDs are the same length as the estimated cell contours after tracking. The features are still of a high dimension and will present difficulties in identifying different cell shapes consistently. Hence, a feature extraction using Fourier analysis, is applied to the FDs to perform the dimension reduction.

### 4.1.1 Fourier descriptors of cell contour

In what follows, the outline theory of the FDs are given. For a given

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_i \\ \vdots \\ u_M \end{bmatrix} = \begin{bmatrix} u_1^{(x)} + ju_1^{(y)} \\ \vdots \\ u_i^{(x)} + ju_i^{(y)} \\ \vdots \\ u_M^{(x)} + ju_M^{(y)} \end{bmatrix} \quad (4.1)$$

where  $(u_i^{(x)}, u_i^{(y)})$ ,  $i = 1, \dots, M$  is the coordinates of a cell contour (see Figure 4.2a), the Fourier transform is defined by

$$a_m = \frac{1}{M} \sum_{i=0}^{M-1} u_i \exp\left(-j\frac{2\pi im}{M}\right) \quad (4.2)$$

which has the following properties.

- The Fourier series is periodic of length  $M$ , i.e.,  $a_m = a_{m-cM}$ ,  $c \in \mathbb{Z}$
- When  $m = 0$ ,

$$a_0 = \frac{1}{M} \sum_{i=0}^{M-1} u_i \quad (4.3)$$

which is the constant component and represents the centroid location of cell boundary.

- For  $m = 1$ , the relevant frequency component represents the shape scale. It is assumed that  $u_i$  is on the boundary of a circle centred at the origin with radius  $R$ . If the circle is represented by

$$u_i = R \exp\left(2\pi j \frac{i}{M}\right) \quad (4.4)$$

then

$$\begin{aligned} a_1 &= \frac{1}{M} \sum_{i=0}^{M-1} u_i \exp\left(-j\frac{2\pi im}{M}\right) \\ &= \frac{1}{M} \sum_{i=0}^{M-1} R \exp\left(j\frac{2\pi i}{M} \exp\left(-j\frac{2\pi i}{M}\right)\right) \\ &= R \end{aligned} \quad (4.5)$$

- In terms of the other values of  $m$ ,  $a_m$  represent forces acting at different

locations on the circle. For  $m > 1$ ,  $a_m$  indicates push forces working on  $m - 1$  locations (4.6) simultaneously. For  $m < 0$ , it represents pull effects working on  $1 - m$  points placed on the circle. The point index,  $i$ , is given in (4.6).

$$i = \begin{cases} \frac{M}{2(m-1)} + k\frac{M}{m-1} + M\frac{\theta}{2\pi} & , k = 1, 2, \dots, m-1 & \text{if } m > 1 \\ \frac{M}{2(1-m)} + k\frac{M}{1-m} + M\frac{\theta}{2\pi} & , k = 1, 2, \dots, 1-m & \text{if } m < 0 \end{cases} \quad (4.6)$$

where  $\theta = \arctan\left(\frac{\text{Im}(a_m)}{\text{Re}(a_m)}\right)$  is the phase of the Fourier series,  $a_m$ .  $\text{Im}(a_m)$  is the imaginary part and  $\text{Re}(a_m)$  is the real part of  $a_m$ .

The FD are obtained by normalisation of the Fourier series (4.7). The location invariance is achieved by removing the constant component, the scale invariance is achieved by dividing the scale parameter  $a_1$  and the rotation invariance is performed by eliminating the phase components [121, 179, 187]. The resulting shape feature vector is given by,

$$\begin{aligned} \mathbf{f} &= \left[ f_0 \quad f_1 \quad f_2 \quad f_3 \quad \dots \quad f_{M-1} \right]^\top \\ &= \left[ 0 \quad 1 \quad \frac{|a_2|}{|a_1|} \quad \frac{|a_3|}{|a_1|} \quad \dots \quad \frac{|a_{M-1}|}{|a_1|} \right]^\top \end{aligned} \quad (4.7)$$

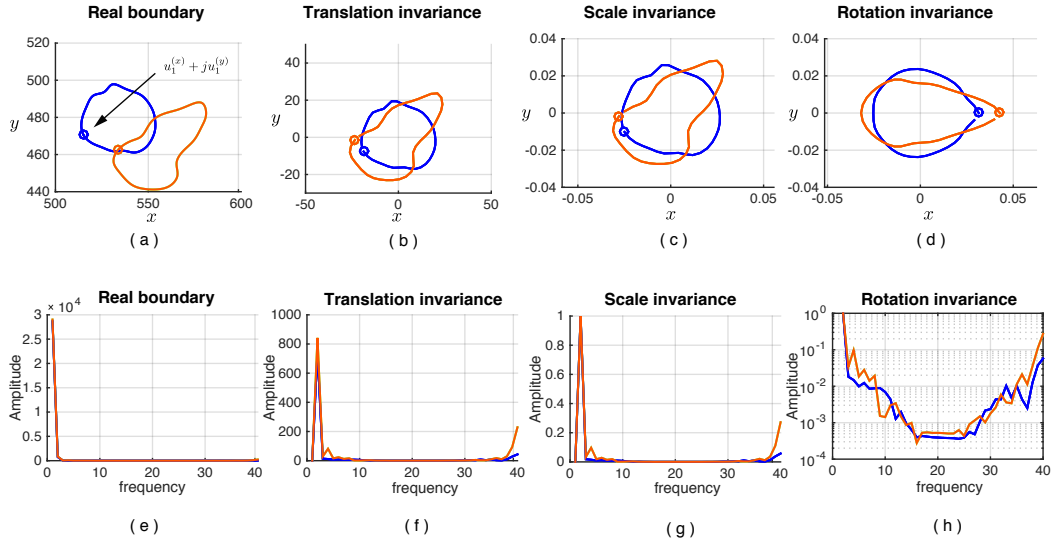
This normalisation process is demonstrated in Figure 4.2. The boundaries of two cells, which are located at different coordinates in raw image (see Figure 4.2a), correspond to two symmetric shapes located at origin with normalised scale (see Figure 4.2d). The rotation normalisation removes the phase information from the descriptor elements so that only the amplitude spectrum is considered in the obtained FD. The general shape features remain, which are sufficient to tell the differences between two general cell shape types, such as the circular and the teardrop shape.

#### 4.1.2 Feature extraction of cell boundary descriptors

The FD represents the features of the cell shape in the form of scale, translation and rotation invariance. However, the FD is a high-dimensional representation and a low-dimensional representation is required for use in clustering. In order to keep the essential shape information, the power spectrum of the Fourier descriptor is used to select the principal frequency components.

A data matrix  $\mathbf{D}^{(f)} \in \mathbb{R}^{M \times (N+1)}$  is first constructed as

$$\mathbf{D}^{(f)} = \left[ \mathbf{f}_{ref} \quad \mathbf{f}_1 \quad \mathbf{f}_2, \dots, \mathbf{f}_N \right] \quad (4.8)$$



**Figure 4.2: The normalisation of shape descriptors.** (a) The original cell boundaries, where the start points are marked by circles, and the Fourier transform is given in subfigure (e); (b) Normalisation of shape translation and the Fourier transform is given in subfigure (f); (c) Normalisation of shape scale and the Fourier transform is given in subfigure (g); (d) Normalisation of rotation and the Fourier transform is given in subfigure (h), which is the same with subfigure (g), but represented in logarithmic scale of the y-axis. Blue and orange are respectively employed to represent the boundaries of two different cell shapes and the corresponding Fourier transform. In order to observe the changes of Fourier transform from subfigures (e) – (h), 40 points are evenly sampled on each cell boundary.

where  $N$  is the number of cell shapes,  $\mathbf{f}_n$  is the FD of the  $n$ -th cell shape ( $n \in [1, N]$ ) and shown in Figure 4.3a.  $\mathbf{f}_{ref}$  is the FD of a unit circle, which is a reference for shapes,

$$\mathbf{f}_{ref} = \left[ 0 \ 1 \ 0 \ \dots \ 0 \ 0 \right]^\top \quad (4.9)$$

The reference shape is included in order to identify where this reference circle shape is placed in the reduced dimensional space.

The power spectrum of the FD is used to measure the frequency content, and the average power spectrum of the data base is shown in Figure 4.3b. The frequencies are arranged in descending order of the power. A simplified FD is constructed by only keeping the strong frequency components. The lost shape information is measured by the mean integrated squared error (MISE) between the FD retrieved shapes and the simplified FD retrieved shapes [13, 117]. Figure 4.3c shows the MISE curve from involving the strongest frequency component to the top 20 frequencies ( $M = 199$ ). Accordingly, the first three components are kept, and they

correspond to  $m = 1$ ,  $m = 197$  and  $m = 198$ , respectively.

In terms of the frequency component  $m = 1$ ,  $f_1$  constantly equals one in the normalised FDs. As such the frequency components  $f_{197}$ ,  $f_{198}$  are extracted as a 2D shape descriptor,

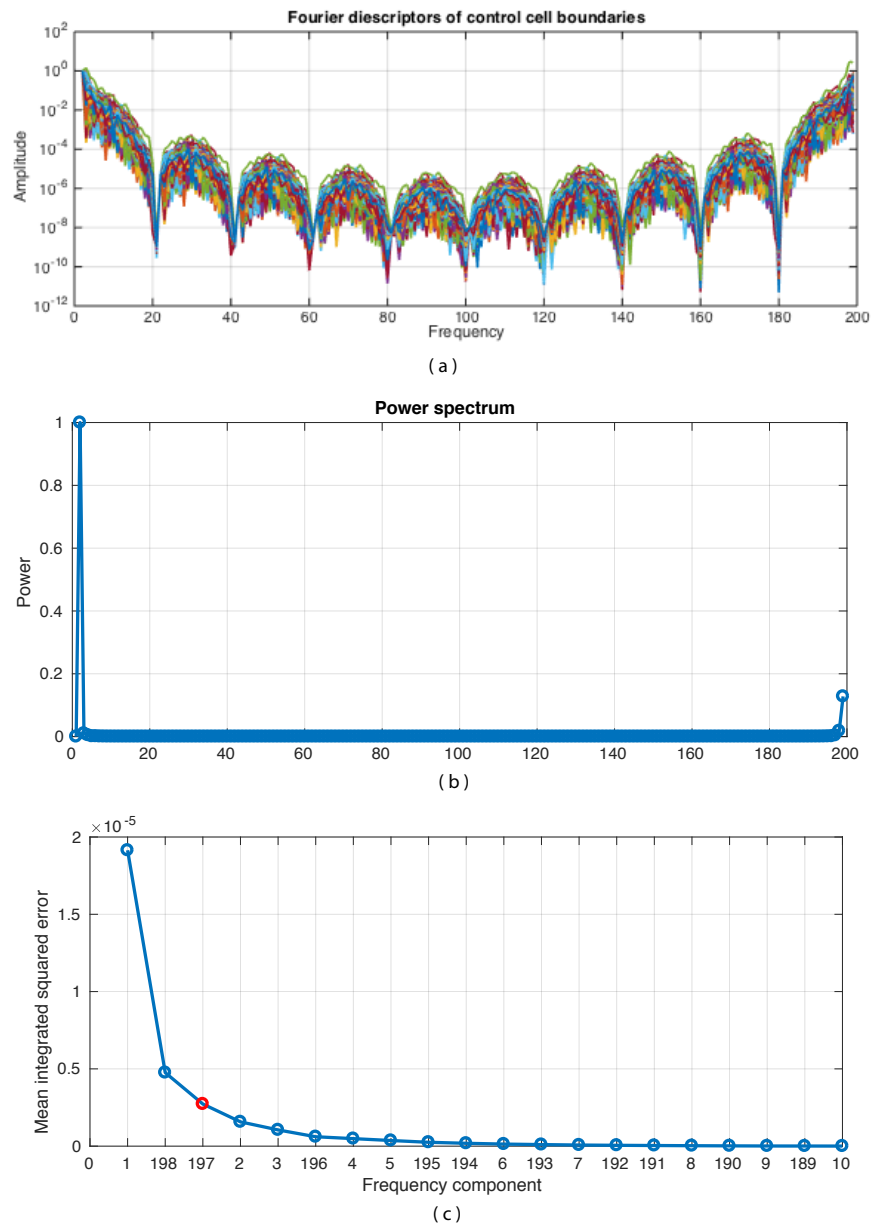
$$\mathbf{W} = \left[ \mathbf{w}_1 \quad \cdots \quad \mathbf{w}_N \right] = \left[ \begin{pmatrix} f_{197}^{(1)} \\ f_{198}^{(1)} \end{pmatrix} \quad \cdots \quad \begin{pmatrix} f_{197}^{(N)} \\ f_{198}^{(N)} \end{pmatrix} \right] \quad (4.10)$$

where  $f_m^{(n)}$  is the component at frequency  $m$  of FD of cell  $n$ .

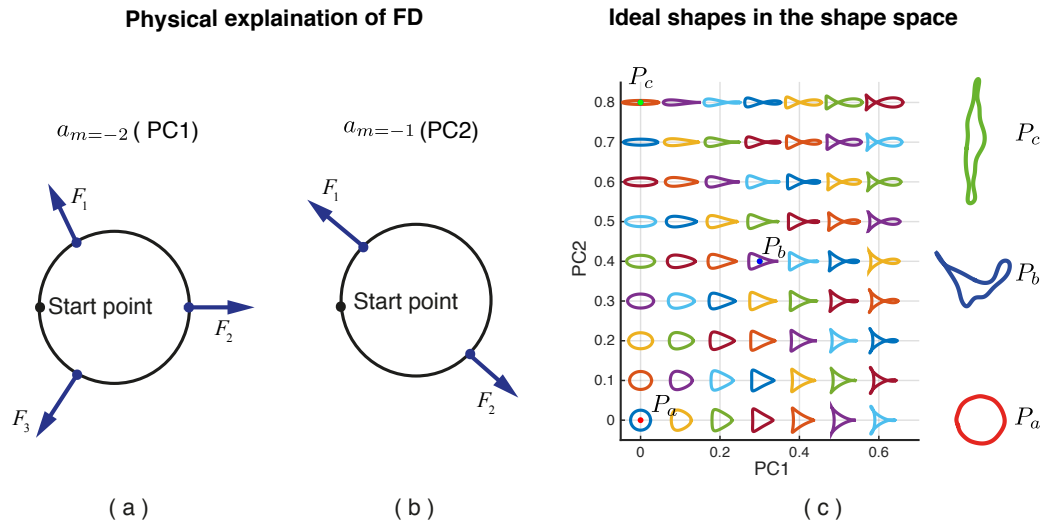
Regarding the physical explanation of FD (4.6), the value of frequency component  $f_{M-2}$  equals to  $f_{-2}$ , which indicates three pull forces separately working at locations of  $\frac{M}{6}$ ,  $\frac{M}{2}$  and  $\frac{5M}{6}$  on the unit circle (see Figure 4.4a). A bigger value of  $a_{M-2}$  represents stronger forces. Similarly,  $f_{M-1}$  generates two pull forces working on locations  $\frac{M}{4}$  and  $\frac{3M}{4}$  (see Figure 4.4b) on the circle. The new insight developed here for analysing cell shape is that the first two principal components (PCs) of FD describe the polarisation of the cell.

The two extracted principal shape features are relevant to the components at frequency  $m = -2$  and  $m = -1$ . A group of FD, combining these two frequency components over a range of amplitudes with zero values at other frequencies, are formulated via inverse Fourier Transform to reconstruct the expected cell shapes at different parts of the shape feature space. Figure 4.4c presents these spectrum of shapes and three template cell shapes from real cell images in the 2D feature space, defined as  $PC1 - PC2$ .

The shape spectrum illustrates that with the increase of  $PC1$  component, shapes polarise in three directions whereas bipolar shapes are associated with the increase of  $PC2$  component. In respect of the template shapes, the red point  $P_a$  is close to the origin and the corresponding shape is similar to a circle. The green point  $P_c$  is almost on the axis of  $PC2$  and corresponds to a bipolar shape. The blue point, on the other hand, has large value of  $PC1$  and a similar value for  $PC2$ . These represent a combination of the forces in Figure 4.4a and Figure 4.4b, and therefore the generated shape becomes a teardrop. The extracted real cell shapes are consistent with the shape inference and its associated principles based on forces. This indicates that the proposed 2D shape descriptors are able to capture the relevant geometric features of cell shapes.



**Figure 4.3: Feature extraction from the FDs.** (a) The FDs of entire data set. Each colour represents a FD of a single cell and 8894 cells are included. (b) The average power spectrum of the FD. (c) The MISE of FD that involves strong frequency components. For example, the red point in subfigure (c) indicates the measured descriptor contains the components at frequencies  $m = 197$ ,  $m = 198$  and  $m = 1$ .



**Figure 4.4: Cell shape description.** (a) The force effects generated by the components of  $a_{m=-2}$ ; (b) The forces generated by the components of  $a_{m=-1}$ ; (c) The ideal shapes in shape space  $PC1 - PC2$ . The descriptors of three practical shapes are labelled by red, blue and green spots and the real shapes are given in the right side.

## 4.2 Cell shape clustering

The shape of breast cancer cells are characterised by large variations. For example, the isolated migrating cancer cell can be in the elongated mesenchymal migration mode or in the rounded-shaped amoeboid migration mode [48]. Figure 4.5a shows the distribution of observed shapes of wild breast cancer cells in the shape space  $PC1 - PC2$ . The distribution of the cell shape features appear to be contained in a triangle region, with the vertices of the triangle corresponding to distinct cell shapes: round, teardrop and elongated. With these shapes associated with potential behavioural modes of the cells, charactering cell modes from the shape features becomes an important challenge.

This section presents the techniques used for clustering shape features. This is accomplished by representing the feature data distributions as Gaussian mixture models with three PCs, with its initialisation provided by the K-means clustering algorithm. The K-means algorithm itself is initialised with the vertices of the triangle region of the feature distribution.



### 4.2.1 Cell shape clustering using K-means

The distribution of cell shape descriptors  $\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_N \end{bmatrix}$  in  $PC1 - PC2$  were noted to be within a triangle region. The three vertices of this triangle were also associated with template cell shapes. Hence, the choice for the number of cell shape clusters are assigned to be  $K = 3$ . The simplest clustering method, K-means clustering, is first used.

The cluster centres are initialised at vertices of the data distribution triangle, which are  $\boldsymbol{\mu}_1 = [0, 0]$ ,  $\boldsymbol{\mu}_2 = [0, 0.8]$  and  $\boldsymbol{\mu}_3 = [0.2, 0.2]$ . The parameters are then identified by minimising the distortion function  $J$ ,

$$J = \sum_{n=1}^N \sum_{k=1}^K \zeta_{nk} \|\mathbf{w}_n - \boldsymbol{\mu}_k\|^2 \quad (4.11)$$

where,  $\zeta_{nk}$  is a binary cluster indicator variable and determined by

$$\zeta_{nk} = \begin{cases} 1 & \text{if } k = \arg \min_i \|\mathbf{w}_n - \boldsymbol{\mu}_i\|^2, \quad i = 1, 2, 3 \\ 0 & \text{otherwise} \end{cases} \quad (4.12)$$

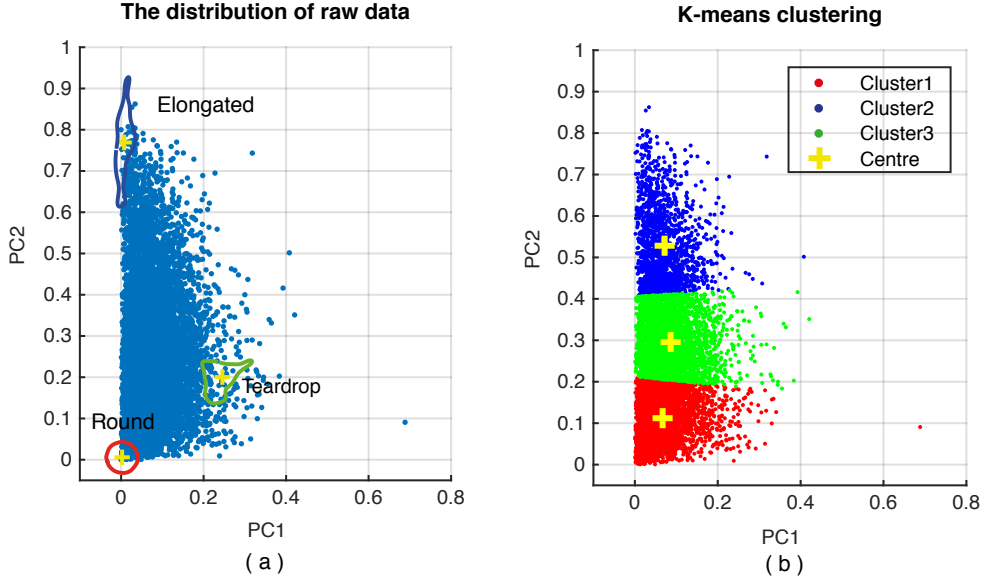
With the assigned indicators, the minimisation of distortion function  $J$  is achieved by setting its first derivative of  $\boldsymbol{\mu}_k$  to be zero which provides the expression of the derived mean values of clusters to be

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^N \zeta_{nk} \mathbf{w}_n}{\sum_{n=1}^N \zeta_{nk}} \quad k = 1, 2, \dots, K \quad (4.13)$$

These are actually the mean values of the observations assigned to the same cluster. The data points are then reassigned according to the distance to the updated cluster centres, the process repeated until the distortion function converges.

The results of K-means clustering are shown in Figure 4.5b, where the data points assigned to the same cluster are shown in the same colour and the identified cluster centres are  $\boldsymbol{\mu}_1 = \begin{bmatrix} 0.6660 & 0.1140 \end{bmatrix}$ ,  $\boldsymbol{\mu}_2 = \begin{bmatrix} 0.6960 & 0.5269 \end{bmatrix}$  and  $\boldsymbol{\mu}_3 = \begin{bmatrix} 0.0854 & 0.2956 \end{bmatrix}$ . The cluster centres identified do not conform to the analysis presented in Section 4.1.2. The cluster centres seem to suggest that there is no discriminatory information in  $PC1$ , whereas,  $PC1$  showed the importance of being associated with the teardrop cell shape and also do not seem to have an appropriate interpretation. Nevertheless, it appears that the boundaries between the different clusters may still be usefully associated with cell shape classes. However,

the cell shape may be better represented by a probabilistic distribution rather than in terms of cell shape classes.



**Figure 4.5: Cell shape clustering using K-means.** (a) Distributions of shape descriptors of wild breast cancer cells; Three practical shapes are projected to the corresponding descriptors in  $PC1 - PC2$ . The vertices of the triangle distributions correspond to distinct cell shapes: round, teardrop, elongated. (b) The K-means clustering results, where the cluster centres are labelled by yellow crosses and the points belonging to the same cluster is given by the same colour.

#### 4.2.2 Cell shape clustering using Gaussian mixture model

In Gaussian mixture modelling, the observations  $\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_N \end{bmatrix}$  are assumed to be a linear combination of Gaussian distributions, given by

$$p(\mathbf{w}_n) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{w}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (4.14)$$

where  $\mathcal{N}(\mathbf{w}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  are the Gaussian components with mean  $\boldsymbol{\mu}_k$  and variance  $\boldsymbol{\Sigma}_k$  in the form

$$\mathcal{N}(\mathbf{w}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{1/2}} \frac{1}{|\boldsymbol{\Sigma}_k|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{w}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{w}_n - \boldsymbol{\mu}_k) \right] \quad (4.15)$$

The mixing coefficient  $\pi_k$  is the probability of belonging to a specific component distribution associated with a cluster and is a latent variable that is required to

Table 4.1: Initialisation of Gaussian mixture model using K-means.

Clusters	Means	Covariances	Mixing coefficients
$k = 1$	$\begin{bmatrix} 0.0666 & 0.1140 \end{bmatrix}$	$\begin{bmatrix} 0.0025 & 0.0004 \\ 0.0004 & 0.0028 \end{bmatrix}$	0.4705
$k = 2$	$\begin{bmatrix} 0.0696 & 0.5269 \end{bmatrix}$	$\begin{bmatrix} 0.0021 & -0.0005 \\ -0.0005 & 0.0083 \end{bmatrix}$	0.1782
$k = 3$	$\begin{bmatrix} 0.0854 & 0.2956 \end{bmatrix}$	$\begin{bmatrix} 0.0029 & -0.0002 \\ -0.0002 & 0.0036 \end{bmatrix}$	0.3512

satisfy

$$0 \leq \pi_k \leq 1 \quad (4.16)$$

$$\sum_{k=1}^K \pi_k = 1 \quad (4.17)$$

The log likelihood function for a mixture Gaussian distribution model of (4.14), is expressed as

$$\ln p(\mathbf{w}_n | \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^N \ln \left[ \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{w}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right] \quad (4.18)$$

To identify the Gaussian mixtures model, the EM algorithm is used to find the solution that maximises the likelihood function. The EM algorithm is performed via steps of initialisation, expectation, maximisation and evaluation of the likelihood [17, 39].

- Initially, the model parameters are set up by K-means clustering, given in Section 4.2.1 (Table 4.1). The initial value of the log likelihood function is then calculated.
- In the expectation step, the conditional probability of  $z_k$  given  $\mathbf{w}_n$ , expressed as  $\gamma(z_{nk})$ , is calculated through the initial parameters and is regarded as the responsibility of cluster  $k$  for observations  $\mathbf{w}_n$ .

$$\begin{aligned}
\gamma(z_{nk}) &\equiv p(z_k = 1 \mid \mathbf{w}_n) = \frac{p(z_k = 1)p(\mathbf{w}_n \mid z_k = 1)}{\sum_{i=1}^K p(z_i = 1)p(\mathbf{w}_n \mid z_i = 1)} \\
&= \frac{\pi_k \mathcal{N}(\mathbf{w}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{i=1}^K \pi_i \mathcal{N}(\mathbf{w}_n \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)} \quad (4.19)
\end{aligned}$$

- In the step of maximisation, the parameter estimations are updated using the obtained responsibility. By setting the first derivative of log likelihood given in (4.18) to be zero, it is obtained that

$$-\sum_{n=1}^N \frac{\pi_k \mathcal{N}(\mathbf{w}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{i=1}^K \pi_i \mathcal{N}(\mathbf{w}_n \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)} \sum_{k=1}^K (\mathbf{w}_n - \boldsymbol{\mu}_k) = 0 \quad (4.20)$$

$$\Rightarrow \sum_{n=1}^N \gamma(z_{nk}) \sum_{k=1}^K (\mathbf{w}_n - \boldsymbol{\mu}_k) = 0 \quad (4.21)$$

$$\Rightarrow \hat{\boldsymbol{\mu}}_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{w}_n \quad (4.22)$$

where

$$N_k = \sum_{n=1}^N \gamma(z_{nk}) \quad (4.23)$$

The estimated variances are

$$\hat{\boldsymbol{\Sigma}}_k = \frac{1}{N_k} \gamma(z_{nk}) (\mathbf{w}_n - \hat{\boldsymbol{\mu}}_k)(\mathbf{w}_n - \hat{\boldsymbol{\mu}}_k)^\top \quad (4.24)$$

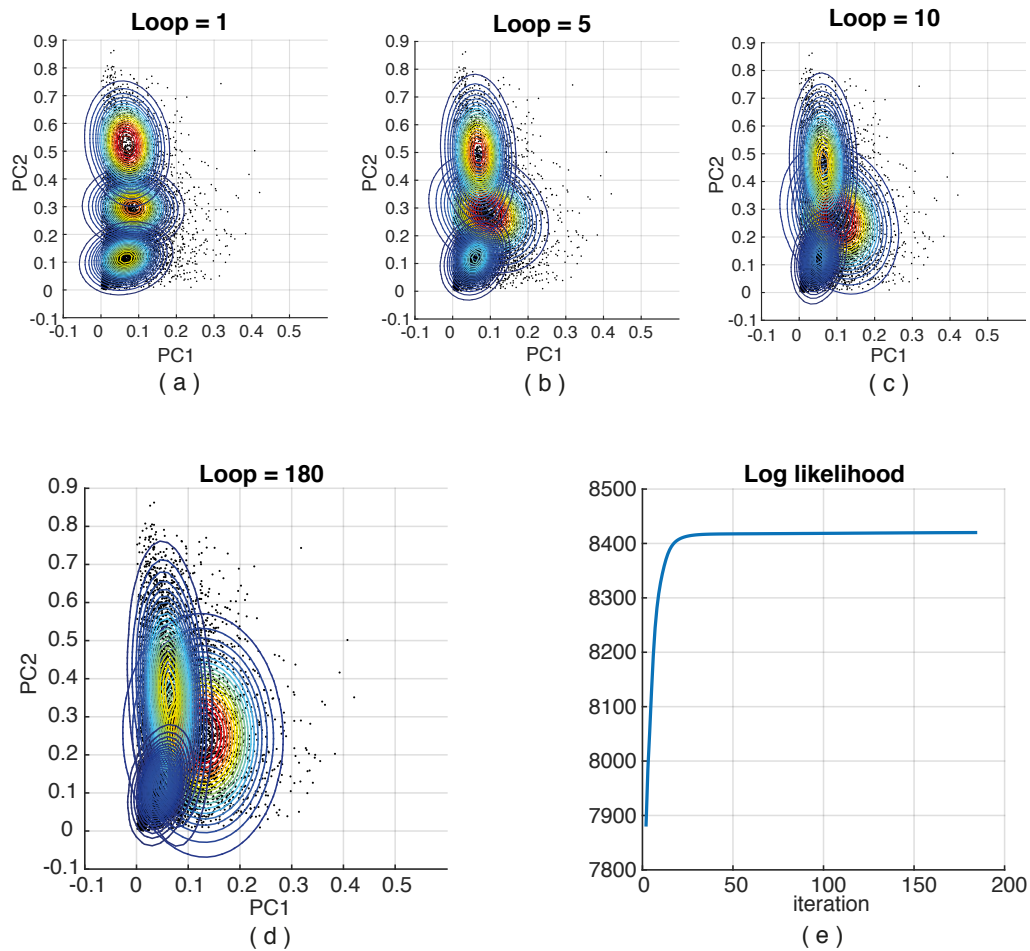
The mixing coefficients are estimated by a Lagrange multiplier and given in

$$\hat{\pi}_k = \frac{N_k}{N} \quad (4.25)$$

- Finally, the new log likelihood can be evaluated. The estimation and maximisation steps will be repeated until the log likelihood is converging and is tested by its change falling below a threshold.

The identification of a Gaussian mixture model, including the convergence of the mixture Gaussian distribution and the likelihood function, is illustrated in Figure 4.6 with the converged parameters given in Table 4.2. The coverage of the Gaussian distribution components in the  $PC1 - PC2$  space suggests a closer alignment to the template shapes, so that the three components are assumed to

correspond to the round shape ( $k = 1$ ), the elongated shape ( $k = 2$ ) and the teardrop shape ( $k = 3$ ).



**Figure 4.6: The identification of a Gaussian mixture model of cell shape distribution by the EM algorithm.** Subfigures (a)–(d) show the distribution of the Gaussian mixture model at the iteration 1, 5 and 10 and the final results; (e) Likelihood function over the iterations.

The results appear to support the notion that there are three fundamental cell shapes that can be found in the population of cells. The cell morphological analysis that uses FDs, followed by the feature extraction and the Gaussian mixture modelling can be used in an automatic analysis towards labelling each cell shape class to be round, elongated or teardrop.

Table 4.2: The parameters of converged Gaussian mixture model.

Clusters	Means	Covariances	Mixing coefficients
$k = 1$	$\begin{bmatrix} 0.0465 & 0.1205 \end{bmatrix}$	$\begin{bmatrix} 0.0006 & 0.0005 \\ 0.0005 & 0.0041 \end{bmatrix}$	0.3440
$k = 2$	$\begin{bmatrix} 0.0642 & 0.3604 \end{bmatrix}$	$\begin{bmatrix} 0.0011 & -0.0010 \\ -0.0010 & 0.0263 \end{bmatrix}$	0.4136
$k = 3$	$\begin{bmatrix} 0.1287 & 0.2510 \end{bmatrix}$	$\begin{bmatrix} 0.0039 & -0.0002 \\ -0.0002 & 0.0167 \end{bmatrix}$	0.2424

### 4.3 Cell morphodynamics analysis

The study of morphodynamics not only demands that we classify cell shapes but also quantify the dynamics of how the cell shapes evolve over time as well as to link them with cell motility. This problem is approached here with HMM modelling where cell shape classes are represented as states of the HMM.

#### 4.3.1 Description of cell shape evolution based on Gaussian mixture model

The evolution of cell shapes between adjacent frames can be modelled using the clustering results from the Gaussian mixture model. If the clustering results are assumed to be true, then the evolution of cell shapes can be thought of as a sequence of shape clusters which can be represented as a Markov model. This is a two stage method involving clustering and then using that result in modelling the dynamics of shapes.

Using the parameters of the Gaussian mixture model, cell shapes are classified into the shape class with the maximum conditional probability,  $\gamma(z_{nk})$ , defined in (4.19).

$$\chi_n = \arg \max_k \gamma(z_{nk}), \quad k = 1, 2, 3 \quad (4.26)$$

where  $n$  is the index of cell shape and  $\chi_n = 1, 2, 3$  represent the shape state of round, elongated and teardrop, respectively.

Subsequently, cell shapes are linked over time according to the cell correspondence calculated from the tracking process (Methods given in Chapter 3). Given  $\chi_t^{(i)}$  is the shape type of cell ( $i$ ) at moment  $t$ , there are nine possible shape evolu-

tions from  $\chi_{t-1}^{(i)}$  to  $\chi_t^{(i)}$ , which are:

$$\begin{array}{lll} \textit{Round} \rightarrow \textit{Round} & \textit{Round} \rightarrow \textit{Elongated} & \textit{Round} \rightarrow \textit{Teardrop} \\ \textit{Elongated} \rightarrow \textit{Round} & \textit{Elongated} \rightarrow \textit{Elongated} & \textit{Elongated} \rightarrow \textit{Teardrop} \\ \textit{Teardrop} \rightarrow \textit{Round} & \textit{Teardrop} \rightarrow \textit{Elongated} & \textit{Teardrop} \rightarrow \textit{Teardrop} \end{array} \quad (4.27)$$

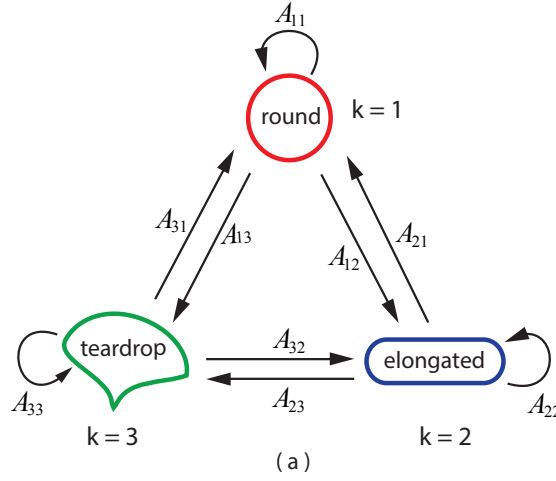
where  $\textit{Round} \rightarrow \textit{Elongated}$  represents that the cell at time  $t - 1$  is round shape and transitions to elongated shape in the next frame at time  $t$ . Using the cell track analysis, the number of all possible shape transitions can be counted and normalised on the basis of prior shape type, thereby a matrix  $\mathbf{A}_p \in \mathbb{R}^{3 \times 3}$  can be generated to statistically describe the behaviours of cell shape evolution in adjoining frames. Each element of  $\mathbf{A}_p$  corresponds to normalised (conditional on the prior state) frequency of the transition given in (4.27).

The two stage analysis is essentially an approximation of the morphodynamics model since the first stage of clustering assumes that there are no uncertainties in the classification process. Any clustering errors will then get passed on the Markov model stage which results in a model that is only an approximate representation. An improved modelling approach that integrates these two stages into one is offered by the use of HMM, as described in Section 4.3.2. As such, the two stage method must be viewed as a simplified approximation to the HMM.

### 4.3.2 Cell morphodynamics modelling using hidden Markov model

Let the sequence of cell shape features of a specific cell ( $i$ ) be denoted as  $\mathbf{w}_t^{(i)}$  over time  $t = 1, \dots, T_i$  where  $T_i$  is the number of samples for that specific cell track and  $i = 1, \dots, N_c$  is the index of cells. If it is assumed that the current shape feature sequence is dependent only on its most recent past observation, the shape change over time can then be modelled by a first-order Markov process. The cell shape class to which a shape feature pattern is to be associated with is unknown and represented as a latent variable  $\mathbf{z}$ . This latent variable can be one of the three discrete states: round, elongated and teardrop. The identified Gaussian mixture model of cell shape distributions (Table 4.2) can be treated as prior information to this stage of the modelling and therefore are employed as the initialisation of component densities. The time sequences of cell shape descriptors are then modelled by a HMM.

The HMM represents the conditional dependence in the time sequence through a set of parametrised conditional probability distribution. This conditional distri-



**Figure 4.7: Lattice diagram of cell shape transitions.** Cell shapes are represented as the states of the HMM and the transition probabilities between different shape states from  $t - 1$  to  $t$  are given by transition matrix, of which the element in row  $i$  and column  $k$  is  $A_{ik} \equiv p(z_{t,k} = 1 \mid z_{t-1,i} = 1)$ .

bution  $p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$  is then constructed as

$$p(\mathbf{z}_t \mid \mathbf{z}_{t-1}, \mathbf{A}) = \prod_{k=1}^K \prod_{i=1}^K A_{jk} \quad (4.28)$$

where  $\mathbf{z}_t$  is the shape states at time  $t \in [2, T]$  and  $T$  is the length of time sequence.  $\mathbf{A}$  is the transition matrix, which indicates the relations between latent states.  $A_{jk} \equiv p(z_{t,k} = 1 \mid z_{t-1,j} = 1)$  is the element in row  $j$  and column  $k$  of  $\mathbf{A}$  and represents the transition probabilities. These represent the parameters of the model and need to be determined from data.

The HMM representation of the cell shape evolution between the three classes of cell shapes is illustrated by a lattice diagram in Figure 4.7.

The parameters of the HMM are determined by likelihood maximisation, using the EM algorithm. Assuming that three Gaussian distributed classes are involved, the emission density of the HMM will be the corresponding component Gaussian distribution from amongst the mixture (4.14) with the distribution parameters  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\pi}$ . The transition matrix  $\mathbf{A}$  is initialised uniformly with different diagonal values in the range  $\rho \in [0.4, 0.98]$  to be consistent with the prior knowledge that the cell shape class has a higher probability to be unchanged at the next time



instant. With the off-diagonal transition probabilities set to be equal,  $\mathbf{A}$  becomes,

$$A_{jk} = \begin{cases} \rho & \text{if } j = k \\ \frac{1-\rho}{2} & \text{if } j \neq k \end{cases} \quad (4.29)$$

Next, for the shape feature sequence of cell ( $i$ ),  $\mathbf{W}^{(i)} = [\mathbf{w}_1^{(i)}, \dots, \mathbf{w}_{T_i}^{(i)}]$ , with the latent sequence  $\mathbf{z}^{(i)}$ , the joint posterior  $\zeta(\mathbf{z}_{t-1}^{(i)}, \mathbf{z}_t^{(i)})$  and the marginal posterior distribution  $\gamma(\mathbf{z}_t^{(i)})$  are evaluated using the forward-backward algorithm [17] at the Expectation stage of EM. Using Bayes' theorem,

$$\begin{aligned} \gamma(\mathbf{z}_t^{(i)}) &= p(\mathbf{z}_t^{(i)} | \mathbf{W}^{(i)}) \\ &= \frac{p(\mathbf{W}^{(i)} | \mathbf{z}_t^{(i)})p(\mathbf{z}_t^{(i)})}{p(\mathbf{W}^{(i)})} \end{aligned} \quad (4.30)$$

$$\begin{aligned} &= \frac{p(\mathbf{w}_1^{(i)}, \dots, \mathbf{w}_t^{(i)}, \mathbf{z}_t^{(i)})p(\mathbf{w}_{t+1}^{(i)}, \dots, \mathbf{w}_{T_i}^{(i)} | \mathbf{z}_t^{(i)})}{p(\mathbf{W}^{(i)})} \\ &= \frac{\alpha(\mathbf{z}_t^{(i)})\beta(\mathbf{z}_t^{(i)})}{p(\mathbf{W}^{(i)})} \end{aligned} \quad (4.31)$$

$$\begin{aligned} \zeta(\mathbf{z}_{t-1}^{(i)}, \mathbf{z}_t^{(i)}) &= p(\mathbf{z}_{t-1}^{(i)}, \mathbf{z}_t^{(i)} | \mathbf{W}^{(i)}) \\ &= \frac{p(\mathbf{W}^{(i)} | \mathbf{z}_{t-1}^{(i)}, \mathbf{z}_t^{(i)})p(\mathbf{z}_{t-1}^{(i)}, \mathbf{z}_t^{(i)})}{p(\mathbf{W}^{(i)})} \\ &= \frac{p(\mathbf{w}_1^{(i)}, \dots, \mathbf{w}_{t-1}^{(i)} | \mathbf{z}_{t-1}^{(i)})p(\mathbf{w}_{t-1}^{(i)} | \mathbf{z}_{t-1}^{(i)})p(\mathbf{w}_{t+1}^{(i)}, \dots, \mathbf{w}_{T_i}^{(i)} | \mathbf{z}_t^{(i)})}{p(\mathbf{W}^{(i)})} \dots \\ &\quad \dots \frac{p(\mathbf{z}_t^{(i)} | \mathbf{z}_{t-1}^{(i)})p(\mathbf{z}_{t-1}^{(i)})}{p(\mathbf{W}^{(i)})} \\ &= \frac{\alpha(\mathbf{z}_{t-1}^{(i)})p(\mathbf{w}_t^{(i)} | \mathbf{z}_t^{(i)})p(\mathbf{z}_t^{(i)} | \mathbf{z}_{t-1}^{(i)})\beta(\mathbf{z}_t^{(i)})}{p(\mathbf{W}^{(i)})} \end{aligned} \quad (4.32)$$

where,  $\alpha(\mathbf{z}_t^{(i)}) \equiv p(\mathbf{w}_1^{(i)}, \dots, \mathbf{w}_t^{(i)}, \mathbf{z}_t^{(i)})$  is the joint distribution of all the shape observations up to the moment  $t$  and the hidden states  $\mathbf{z}_t^{(i)}$ .  $\beta(\mathbf{z}_t^{(i)}) \equiv p(\mathbf{w}_{t+1}^{(i)}, \dots, \mathbf{w}_{T_i}^{(i)} | \mathbf{z}_t^{(i)})$  is the conditional probability of all following observations from  $t+1$  to the end of the sequence given the latent variable. By applying the sum and product rules, the joint probabilities are expressed by

$$\alpha(\mathbf{z}_t^{(i)}) = p(\mathbf{w}_t^{(i)} | \mathbf{z}_t^{(i)}) \sum_{\mathbf{z}_{t-1}^{(i)}} \alpha(\mathbf{z}_{t-1}^{(i)})p(\mathbf{z}_t^{(i)} | \mathbf{z}_{t-1}^{(i)}) \quad (4.33)$$

which is a part of forward recursion. The recursion is initialised by:

$$\alpha(\mathbf{z}_1^{(i)}) = p(\mathbf{z}_1^{(i)})p(\mathbf{w}_1^{(i)} | \mathbf{z}_1^{(i)}) \quad (4.34)$$

Similarly, a backward recursion is derived in terms of the conditional probabilities, given in (4.35), and is initialised by  $\beta(\mathbf{z}_{T_i}^{(i)}) = 1$ .

$$\beta(\mathbf{z}_t^{(i)}) = \sum_{\mathbf{z}_{t+1}^{(i)}} \beta(\mathbf{z}_{t+1}^{(i)})p(\mathbf{w}_{t+1}^{(i)} | \mathbf{z}_{t+1}^{(i)})p(\mathbf{z}_{t+1}^{(i)} | \mathbf{z}_t^{(i)}) \quad (4.35)$$

The function  $\alpha(\mathbf{z}_t^{(i)})$  is rescaled in (4.36) to avoid the issue that  $\alpha(\mathbf{z}_t^{(i)})$  becoming exponentially close to zero:

$$\hat{\alpha}(\mathbf{z}_t^{(i)}) = p(\mathbf{z}_t^{(i)} | \mathbf{w}_1^{(i)}, \dots, \mathbf{w}_t^{(i)}) = \frac{\alpha(\mathbf{z}_t^{(i)})}{\prod_{h=1}^t F_h^{(i)}} \quad (4.36)$$

where a scaling factor  $F_t^{(i)} = p(\mathbf{w}_t^{(i)} | \mathbf{w}_1^{(i)}, \dots, \mathbf{w}_{t-1}^{(i)})$  is employed. The forward recursion then becomes

$$F_t^{(i)} \hat{\alpha}(\mathbf{z}_t^{(i)}) = p(\mathbf{w}_t^{(i)} | \mathbf{z}_t^{(i)}) \sum_{\mathbf{z}_{t-1}^{(i)=1}}^K \hat{\alpha}(\mathbf{z}_{t-1}^{(i)})p(\mathbf{z}_t^{(i)} | \mathbf{z}_{t-1}^{(i)}) \quad (4.37)$$

The function  $\beta(\mathbf{z}_t^{(i)})$  is rescaled in (4.38) and the corresponding recursion equation is given in (4.39).

$$\hat{\beta}(\mathbf{z}_t^{(i)}) = \frac{\beta(\mathbf{z}_t^{(i)})}{\prod_{h=t+1}^{T_i} F_h^{(i)}} \quad (4.38)$$

$$F_{t+1}^{(i)} \hat{\beta}(\mathbf{z}_t^{(i)}) = \sum_{\mathbf{z}_{t+1}^{(i)=1}}^K \hat{\beta}(\mathbf{z}_{t+1}^{(i)})p(\mathbf{w}_{t+1}^{(i)} | \mathbf{z}_{t+1}^{(i)})p(\mathbf{z}_{t+1}^{(i)} | \mathbf{z}_t^{(i)}) \quad (4.39)$$

Consequently, the required joint and condition probabilities are expressed as

$$\gamma(\mathbf{z}_t^{(i)}) = \hat{\alpha}(\mathbf{z}_t^{(i)})\hat{\beta}(\mathbf{z}_t^{(i)}) \quad (4.40)$$

$$\zeta(\mathbf{z}_{t-1}^{(i)}, \mathbf{z}_t^{(i)}) = \frac{\hat{\alpha}(\mathbf{z}_{t-1}^{(i)})p(\mathbf{w}_t^{(i)} | \mathbf{z}_t^{(i)})p(\mathbf{z}_t^{(i)} | \mathbf{z}_{t-1}^{(i)})\hat{\beta}(\mathbf{z}_t^{(i)})}{F_t^{(i)}} \quad (4.41)$$

where, the scale factor  $F_t^{(i)}$  is cancelled out during the EM.

At the maximisation stage, the logarithm of the likelihood function is defined

as a function of the model parameters,

$$\begin{aligned}
Q(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) &= \sum_{\mathbf{Z}} p(\mathbf{Z} | \mathbf{W}, \boldsymbol{\theta}) \ln p(\mathbf{W}, \mathbf{Z} | \hat{\boldsymbol{\theta}}) \\
&= \sum_{i=1}^{N_c} \sum_{k=1}^K \gamma(z_{1k}^{(i)}) \ln \pi_k + \sum_{i=1}^{N_c} \sum_{t=2}^{T_i} \sum_{j=1}^K \sum_{k=1}^K \zeta(z_{t-1,j}^{(i)}, z_{tk}^{(i)}) \ln \hat{A}_{jk} + \\
&\quad \sum_{i=1}^{N_c} \sum_{t=1}^{T_i} \sum_{k=1}^K \gamma(z_{tk}^{(i)}) \ln p(\mathbf{w}_t^{(i)} | \boldsymbol{\mu}, \boldsymbol{\Sigma})
\end{aligned} \tag{4.42}$$

Parameters  $\hat{\boldsymbol{\theta}} = \{\hat{\pi}, \hat{\mathbf{A}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\}$  then can be estimated by maximising the log-likelihood function as follows:

$$\hat{\pi}_k = \frac{\sum_{i=1}^{N_c} \gamma(z_{1k}^{(i)})}{\sum_{i=1}^{N_c} \sum_{j=1}^K \gamma(z_{1j}^{(i)})} \tag{4.43}$$

$$\hat{A}_{jk} = \frac{\sum_{i=1}^{N_c} \sum_{t=2}^{T_i} \zeta(z_{t-1,j}^{(i)}, z_{tk}^{(i)})}{\sum_{i=1}^{N_c} \sum_{l=1}^K \sum_{t=2}^{T_i} \zeta(z_{t-1,j}^{(i)}, z_{tl}^{(i)})} \tag{4.44}$$

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^{N_c} \sum_{t=1}^{T_i} \gamma(z_{tk}^{(i)}) \mathbf{w}_t^{(i)}}{\sum_{i=1}^{N_c} \sum_{t=1}^{T_i} \gamma(z_{tk}^{(i)})} \tag{4.45}$$

$$\hat{\boldsymbol{\Sigma}}_k = \frac{\sum_{i=1}^{N_c} \sum_{t=1}^{T_i} \gamma(z_{tk}^{(i)}) (\mathbf{w}_t^{(i)} - \hat{\boldsymbol{\mu}}_k) (\mathbf{w}_t^{(i)} - \hat{\boldsymbol{\mu}}_k)^\top}{\sum_{i=1}^{N_c} \sum_{t=1}^{T_i} \gamma(z_{tk}^{(i)})} \tag{4.46}$$

The expectation and maximisation steps are then performed iteratively until the likelihood function  $p(\mathbf{W})$  converges to a local maximum. The  $p(\mathbf{W})$  is able to be found by the product rule and is given by (4.47) [16, 17, 52].

$$p(\mathbf{W}) = \prod_{i=1}^{N_c} \prod_{t=1}^{T_i} F_t^{(i)} \tag{4.47}$$

The estimated HMM parameters represent the evolution of cell shape over time. Even though the cell shape distribution was estimated as part of the overall HMM, instead of using the results of Section 4.2.2, the distribution parameters do not reveal time evolution properties. The transition matrix  $\mathbf{A}$  is the parameter

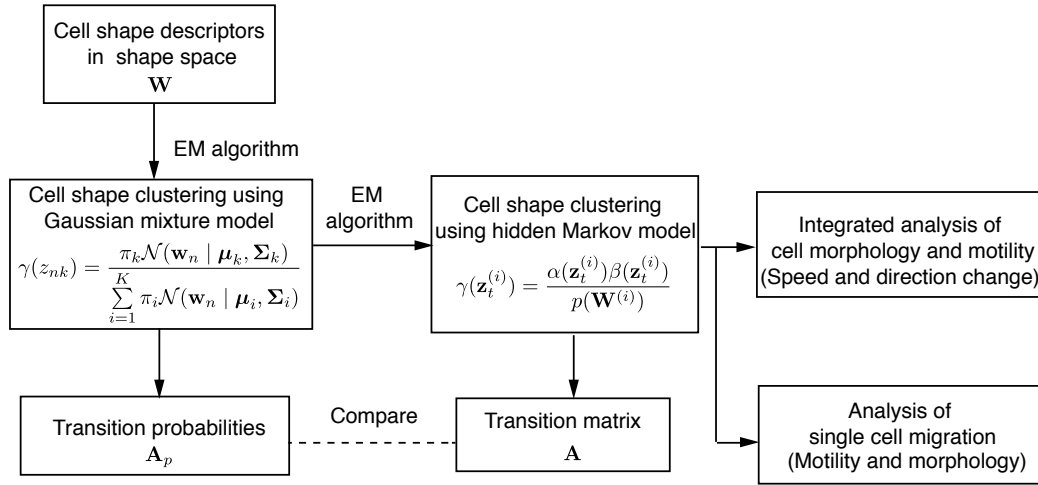


Figure 4.8: Flowchart of cell morphology analysis.

containing this information. However, the diagonal elements of this matrix tend to be large, reflecting the fact that cell shape class transitions are not frequent at the data sampling rates used here. A more appropriate measure under these circumstances is to estimate how long a cell remains in a specific cell shape class before it transitions to another shape class.

This measure is the sojourn time that represents the expected time duration of residence in a particular cell shape class which can be calculated from the transition matrix:

$$\tau_k = \frac{1}{1 - A_{kk}}, \quad k = 1, 2, 3 \quad (4.48)$$

In summary, cell shapes are clustered into round, teardrop and elongated states using HMM, where the transition matrix  $\mathbf{A}$  describes the transition probabilities of states in contiguous moments and sojourn time is used to estimate the expected time of each shape state.

## 4.4 Morphology analysis of breast cancer cells

The morphodynamic modelling algorithm for analysing cell shape evolution developed in the previous sections, summarised in Figure 4.8, is now applied to wild breast cancer cells data described in Table 4.3. This section provides the quantitative analysis of the control cell morphodynamics and links cell morphology to motility characteristics.

**Table 4.3: Experimental information.** The data is used for the demonstration of morphology analysis module.

<b>Experimental data</b>	<ul style="list-style-type: none"> <li>• Cell line: MDA-MB-231 breast cancer cells</li> <li>• Cell collection: control 3-13-2, control 3-13-2, control 3-14-2 <ul style="list-style-type: none"> <li>• Culture condition: tissue culture plastic Dulbecco's Modified Eagle's Medium (DMEM) supplemented 5% heat-inactivated on Fetal Bovine Serum (FBS)</li> </ul> </li> </ul>
<b>Image sequences</b>	<ul style="list-style-type: none"> <li>• Image type: time-lapse grey-scaled image sequence</li> <li>• Resolution: <math>500 \times 664</math></li> <li>• Length: 220 Frames</li> <li>• Time interval: 5 minutes</li> </ul>

#### 4.4.1 Quantitative morphodynamics analysis

The shape descriptors of cells from the data given in Table 4.3 are represented as points in a 2-D shape space (see Figure 4.5a), with the cell shape features distributions modelled as a linear mixture of three Gaussians (see Figure 4.6d). In order to keep a count of the observed cell shape transitions, cell shapes are classified into the shape class with highest membership probability (4.26). The observed number of shape transition between two adjacent frames are counted using the method given in Section 4.3.1, and the results are given in Table 4.4. The counts are normalized by rows to obtain an estimation of the transition matrix  $\mathbf{A}_p$ , given by:

$$\mathbf{A}_p = \begin{bmatrix} 0.8362 & 0.1108 & 0.0530 \\ 0.1089 & 0.8113 & 0.0798 \\ 0.1015 & 0.1909 & 0.7077 \end{bmatrix} \quad (4.49)$$

Note that this transition matrix estimate models the shape sequences as a Markov model but assumes that the shape classes are accurately estimated. In Table 4.4, the counts show that a cell shape state remained in the same state 3 to 10 times over all cell shape transitions. The low cell counts for some shape transitions reveal that potentially the data may be insufficient to accurately model the morphodynamics. With increased amounts of data,  $\mathbf{A}_p$  may become less sensitive to the data variations.

It is clear that a better model representation can be formed by the use of HMM that treats the shape class as a hidden variable and the shape features as observa-

**Table 4.4: Count of shape transitions between adjacent frames**

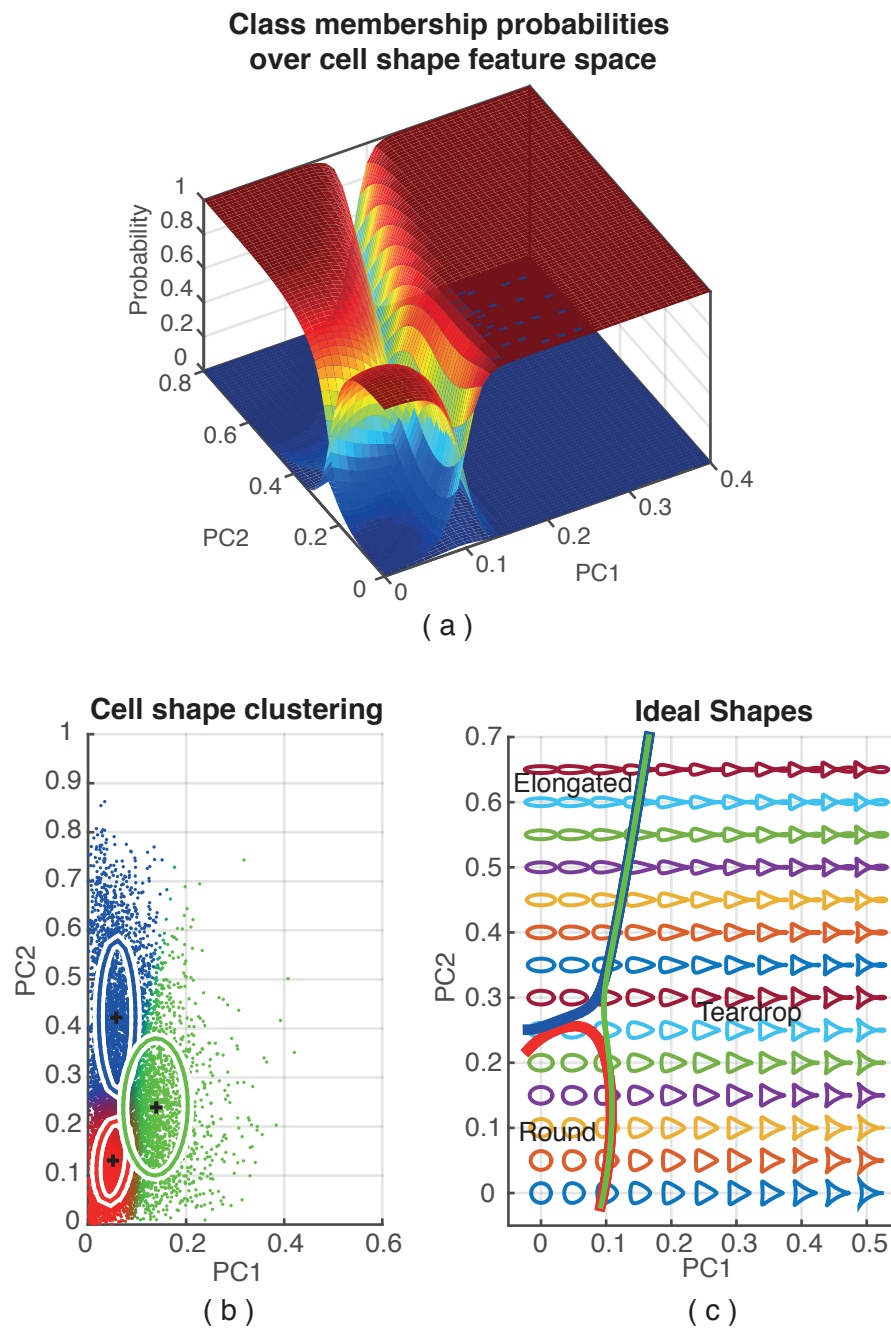
$t-1 \backslash t$	Round (R)	Elongated (E)	Teardrop (T)
R	2920	387	185
E	390	2906	286
T	160	301	1116

tions that are conditioned on the hidden shape class. Hence, the previous analysis was followed with cell morphodynamics being described by a first order HMM, EM based estimation was employed to determine the model parameters. By introducing the first order HMM, cell shape sequences are assumed to be independent of the observations except for those at the most recent observation frame, and the emission probabilities are represented by a mixture Gaussian distribution.

The converged parameters are given in Table 4.5. The cell shape classification is achieved by normalising the mixture distributions, which computes the posterior probability of the features belonging to specific cell shape class,  $\gamma(\mathbf{z}_t)$ . The class membership posterior probabilities over cell shape feature space are shown in Figure 4.9a. The figure shows that the class boundaries are very clearly defined with probabilities dropping and rising sharply. Figure 4.9b presents the data points that are shown by the proportions of the red, blue and green colours associated with the value of membership probabilities. Due to the trends of membership probabilities, prominent colour changes occur at the interaction areas. The clusters intersections in the shape space are shown in Figure 4.9c with the boundary superimposed with the ideal shapes.

**Table 4.5: Parameters of shape feature distribution in the identified HMM.**

Clusters	Means	Covariances	Mixing coefficients
$k = 1$	$\begin{bmatrix} 0.0504 & 0.1298 \end{bmatrix}$	$\begin{bmatrix} 0.0008 & 0.0004 \\ 0.0004 & 0.0048 \end{bmatrix}$	0.3763
$k = 2$	$\begin{bmatrix} 0.0593 & 0.4236 \end{bmatrix}$	$\begin{bmatrix} 0.0010 & -0.0001 \\ -0.0001 & 0.0174 \end{bmatrix}$	0.3274
$k = 3$	$\begin{bmatrix} 0.1375 & 0.2399 \end{bmatrix}$	$\begin{bmatrix} 0.0031 & 0.0001 \\ 0.0001 & 0.0144 \end{bmatrix}$	0.2963



**Figure 4.9: Results of cell clustering based on HMM.** (a) The membership probabilities of different shape class over the shape feature space. (b) The data clustering results. Each point represent a descriptor of cell shape and is coloured by the proportions of red, blue and green colours associated with the associated membership probabilities. (c) The cluster intersections in the shape space and the ideal shapes are located at the coordinates of corresponding descriptors.

The transition matrix is

$$\mathbf{A} = \begin{bmatrix} 0.9766 & 0.0175 & 0.0059 \\ 0.0607 & 0.9094 & 0.0298 \\ 0.3438 & 0.0532 & 0.6029 \end{bmatrix} \quad (4.50)$$

The corresponding Sojourn times are calculated in Table 4.6. Sojourn time is an indicator of the prevalence of cells to maintain a specific shape before any transitions and can characterise morphodynamics differences in different cellular environments. The Sojourn time is also normalised with respect to the round state to see any pattern of relative changes in sojourn time of other cell shapes.

**Table 4.6: The Sojourn time of being each state.**

Cluster (k)	Shape state	Sojourn time (minutes)	Normalised sojourn time
1	Round	214	1
2	Elongated	55	0.2570
3	Teardrop	13	0.0607

The transition matrix in (4.50), indicates that the cells have the highest probability of retaining the same shape state rather than jumping to other states during the 5 minutes interval, whereas the round state has the highest probability (0.9766), followed by the elongated (0.9094) and teardrop shape state (0.6029). When shape state transition occurs, the teardrop shaped cells display a significantly high probability to switch to the round state (0.3438). In respect of sojourn time (Table 4.6), cells are expected to stay in round state for as long as about 4 times of maintaining elongated and about 16 times of maintaining teardrop.

Compared to the approximate transition matrix,  $\mathbf{A}_p$ , in (4.49),  $\mathbf{A}$  displays similar patterns of the probabilities of cell shape transitions. Specifically, the diagonal entries in both matrices are all over 0.5, which indicates the cells have higher probabilities of retaining the shape state instead of transiting to others in the 5 minutes time interval. Furthermore, the diagonal states are identified in the same order of round, elongated and teardrop and cells show the strongest stability of staying in the round state and the least probability of remaining in teardrop shape.

However, there are dissimilarities between  $\mathbf{A}$  and  $\mathbf{A}_p$ . For instance, in HMM, the cells at teardrop state displays a higher probability of switching to round shape than the elongated shape whereas in  $\mathbf{A}_p$ , this characteristic is reversed. This is possibly due to two reasons. Firstly, the shape class boundaries are slightly different



between the two approaches. Secondly, and more importantly, the uncertainties in the estimation is ignored in the approximate transition probability estimation while HMM includes it.

In summary, cells can retain the round shape state for the longest duration while also displaying higher probability to transition to the teardrop shape than the elongated one. Both elongated and teardrop cells have a higher probability to transition to the round shape than transition to the other shapes. The teardrop shaped cells are the most difficult shape state to be retained, indicating that this shape is perhaps a transition state for the cells. These reflect the morphodynamic characteristics during controlled cell migration and this analysis provides a quantitative assessment of morphodynamics.

#### 4.4.2 Integrated motility and morphology analysis

The HMM hidden variable estimates are used in classifying cell shapes according to the shape class with the highest posterior membership probability. However, the highest probability of a cell shape class does not necessarily indicate that this classification has a high degree of confidence or that it is closer to a decision boundary. In order to identify cell motility patterns from those that clearly belong to a specific shape, a threshold on this posterior probability is applied to extract only those cells that can clearly be attributed to a specific cell shape class. These cells will have the posterior probabilities that are greater than the threshold.

The integrated motility and morphology analysis is centred on linking specific cell shapes with features of cell motility. Essentially, it seeks to address whether specific cell shapes are associated with higher migration speed and directional change in motility. Such quantitative integrated analysis can reveal if cellular behaviour can be predicted based on motility and morphology characteristics.

Figure 4.10 gives the cell motile speed distribution in truncated histograms for cells in each of the three shapes. This distribution is repeated with different thresholds to investigate whether there is a strong correlation between motility and morphology. These histograms are normalized by the number of cells in each category so that the histograms represent the class conditional probability.

The elongated cells are characterised by having a distribution with a higher proportion of cells at greater cell speeds of over 8 pixels/5 mins compared to the cells of the other two shapes. The teardrop shaped cells show the highest probability at the speed range [2,4] pixels/5 mins but not at the smallest velocity range, where the round shaped cells have the highest probability. With the increase of the threshold, round cells show a continuous increase of the probability for cell speeds less than 2 pixels/5 mins and the probability is 0.52 when the threshold is

largest at 0.96. Note that the greater the threshold, the more closer the cell shapes chosen in the analysis are to the ideal shapes.

A similar analysis is carried out to analyse cell migration angle change, and its correlation with cell shapes, with the results given in Figure 4.11. As the threshold is increased, the histograms appear to change very little indicating that the distributions are quite stable. The elongated cells exhibit the highest probability when migration angle change is smaller than  $15^\circ$  and bigger than  $110^\circ$ . This indicates that, the elongated cells have the strongest preference to keep moving in the same direction or suddenly orienting in the opposite direction. The teardrop and round cells show similar histogram patterns, but teardrop shaped cells show higher probability of direction change that is smaller than  $15^\circ$ .

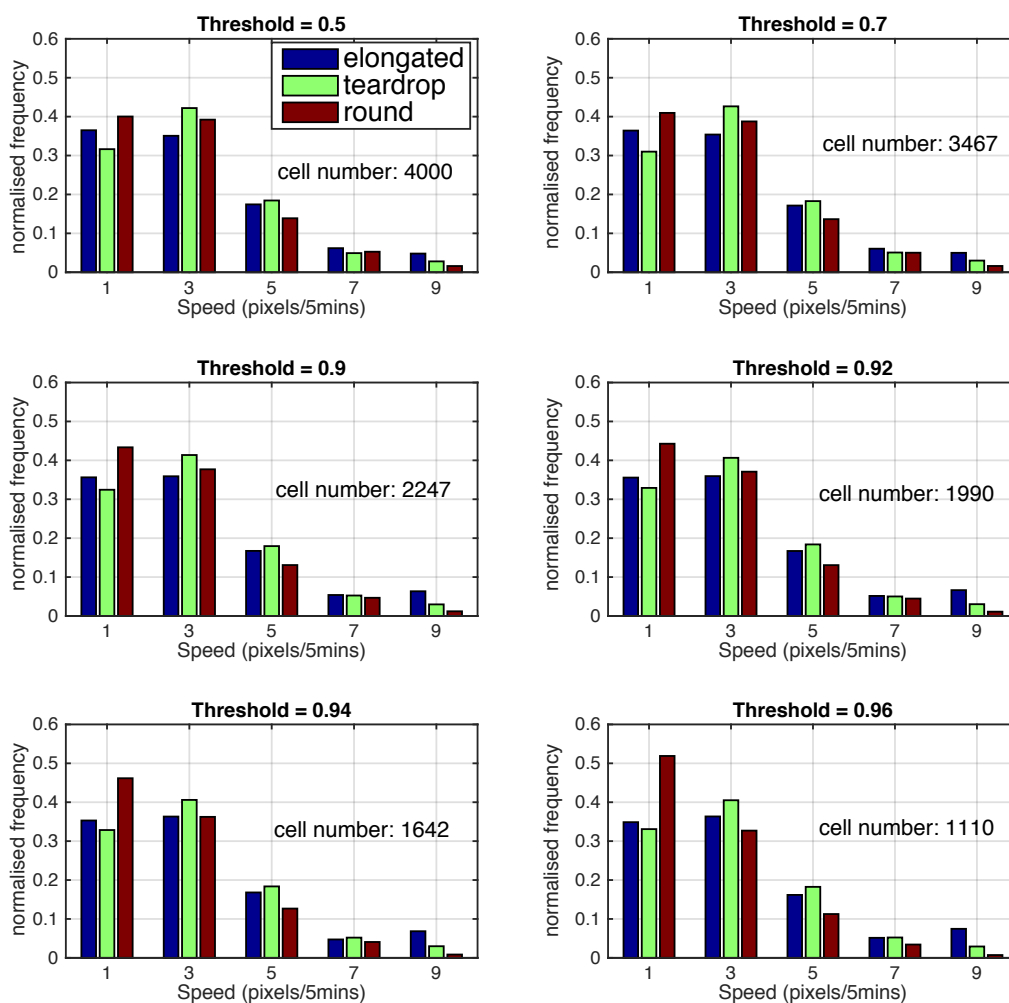
Additionally, cell speed with respect to direction change is analysed in Figure 4.12. The analysis is focussed on the threshold for shape classification being set to 0.96. At this threshold, the cells are closer to their ideal shape and even at this threshold, there are sufficient number of data of cell shape and motility to perform this analysis. In comparison with round cells, teardrop shaped cells move relatively faster. Elongated cells showed a tendency to move forward with a significant speed increase as observed for the range  $[0^\circ, 22.5^\circ]$  (round: 4.28 pixels/5 mins, teardrop: 4.77 pixels/5 mins, elongated: 5.42 pixels/5 mins).

Integrated morphology and motility analysis showed that shape states have distinct motility features. Cells at round shape state move slowly and no preference for direction. In comparison, teardrop shaped cells show a preference for small directional changes,  $[0^\circ, 15^\circ]$ , and move faster along with any orientation. Elongated cells show the highest probability of directional changes smaller than  $15^\circ$ , whereas cells migrate at the highest speed amongst all shapes.

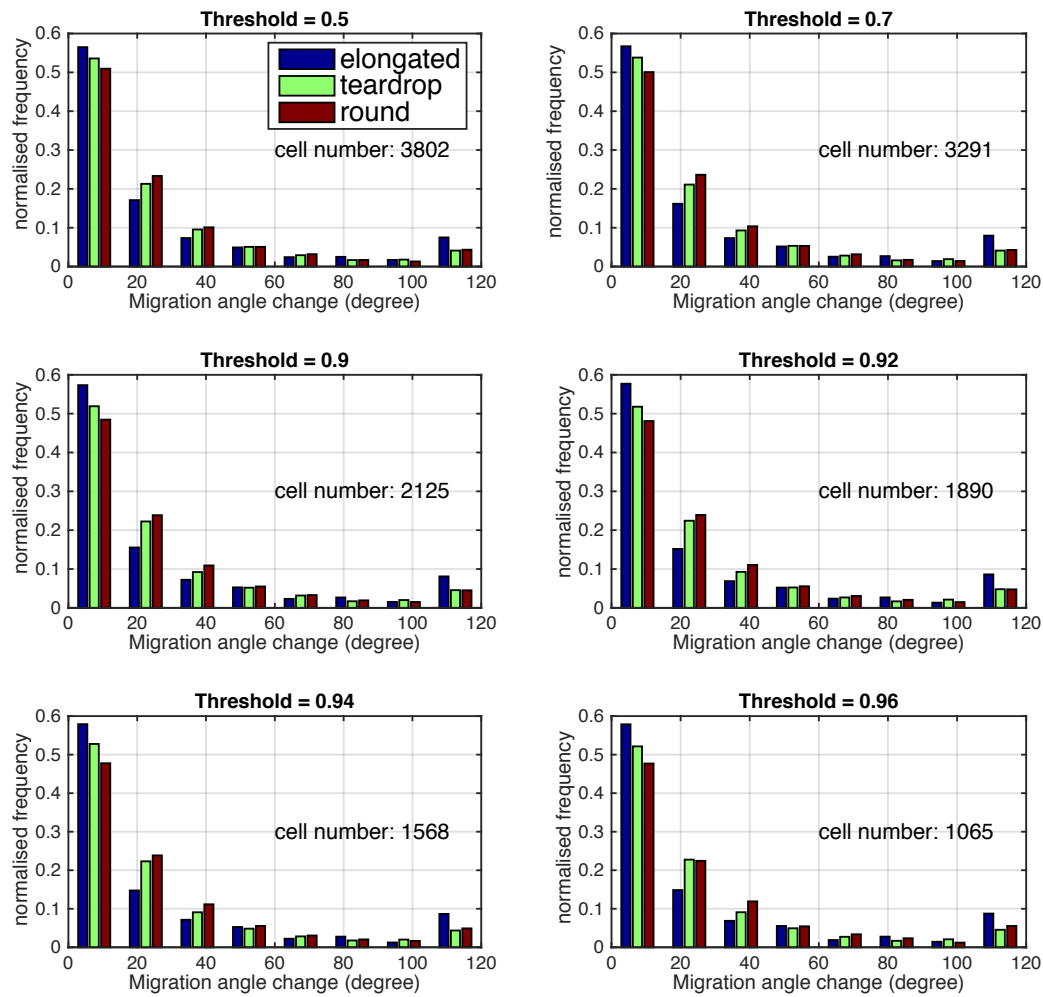
#### 4.4.3 Morphodynamics analysis of single cell migration

The analysis module for cell morphodynamics is based on cell-centric views so that the analysis is amenable for use in the study of single cell migration. A single cell migration with significant directional changes in the observed time-lapse microscopy is extracted. The results of tracking this cell is given in Figure 4.13. Cell boundary tracking showed that cell boundary speed were generally smaller than 12 pixels/frame, except for two sections where cell rear migrated significantly faster than the typical speeds (see Figure 4.13a). The centroid trajectory showed that there were two migration directional changes (see Figure 4.13b).

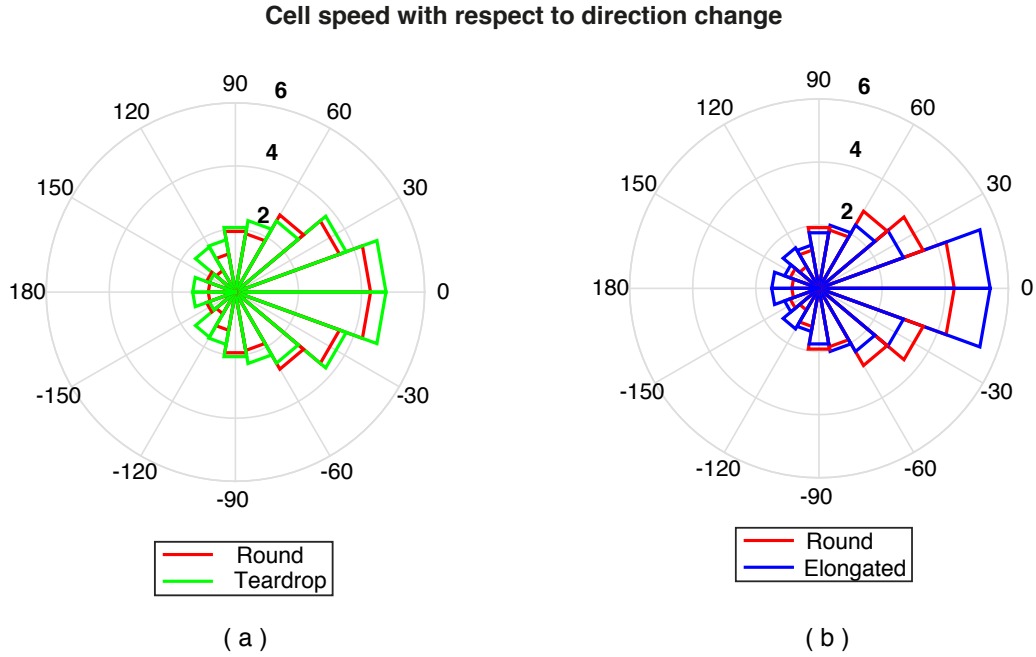
In order to explore how cell shape changes were correlated with motility characteristics of cell speed and cell migration directional change, a quantitative measure of shape change is introduced. This measure is based on the Euclidean dis-



**Figure 4.10: The comparison of cell migration speed of different shapes.** The speed histogram (truncated) of different shape is normalised by the number of cells involved in the shape class. The cell numbers of each analysis are marked in the right of each figure. Cell shape classification is performed with different thresholds: 0.5, 0.7, 0.8, 0.9, 0.94 and 0.96.



**Figure 4.11: The comparison of cell migration angle change for different shapes.** The histogram (truncated) for different shapes are normalised by the number of cells involved in the shape class. The cell numbers of each analysis are marked in the right of each figure.. Cell shape classification is performed with different thresholds: 0.5, 0.6, 0.7, 0.9, 0.92, 0.94 and 0.96.



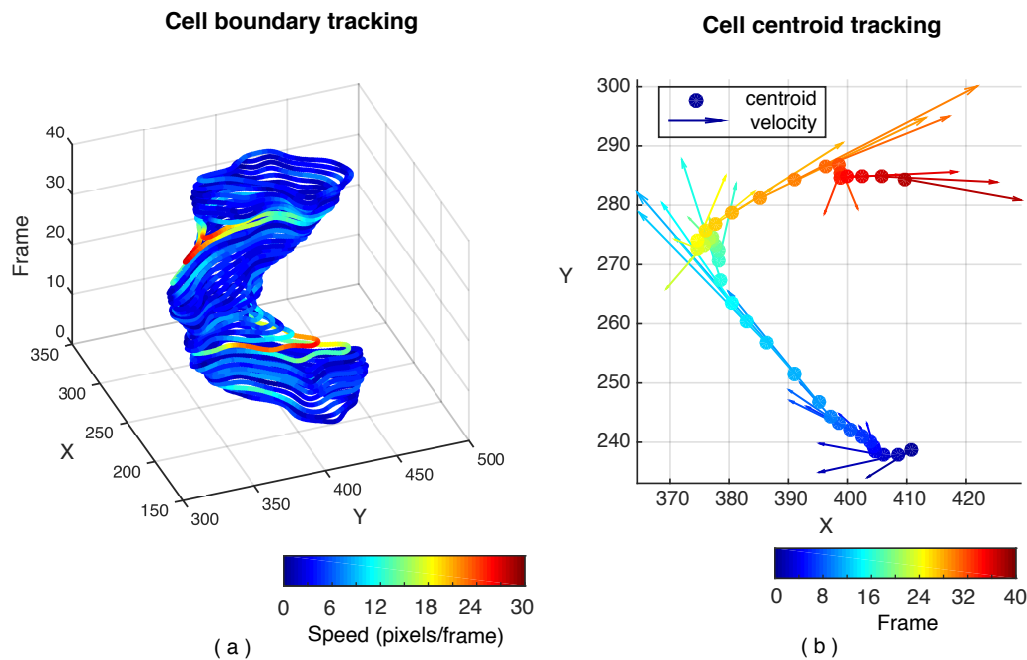
**Figure 4.12: Cell speed with respect to direction change for different shape state.** (a) The comparison of round and teardrop state. (b) The comparison of round and elongated shape state. The threshold, used for shape classification, is set up as 0.96. In the polar charts, angular and radial coordinates represent cell migration direction change ( $^{\circ}$ ) and the corresponding average speed (pixels/5 mins), respectively, additionally, the angle range  $[0^{\circ}, -180^{\circ}]$  are symmetric to the angle range  $[0^{\circ}, 180^{\circ}]$ .

tance in the shape feature space  $PC1 - PC2$ , and given by:

$$\|\Delta PC\| = \sqrt{(w_{t,PC1} - w_{t-1,PC1})^2 + (w_{t,PC2} - w_{t-1,PC2})^2} \quad (4.51)$$

where  $(w_{t,PC1}, w_{t,PC2})$  denotes the 2D shape descriptor of cell in frame  $t$ .

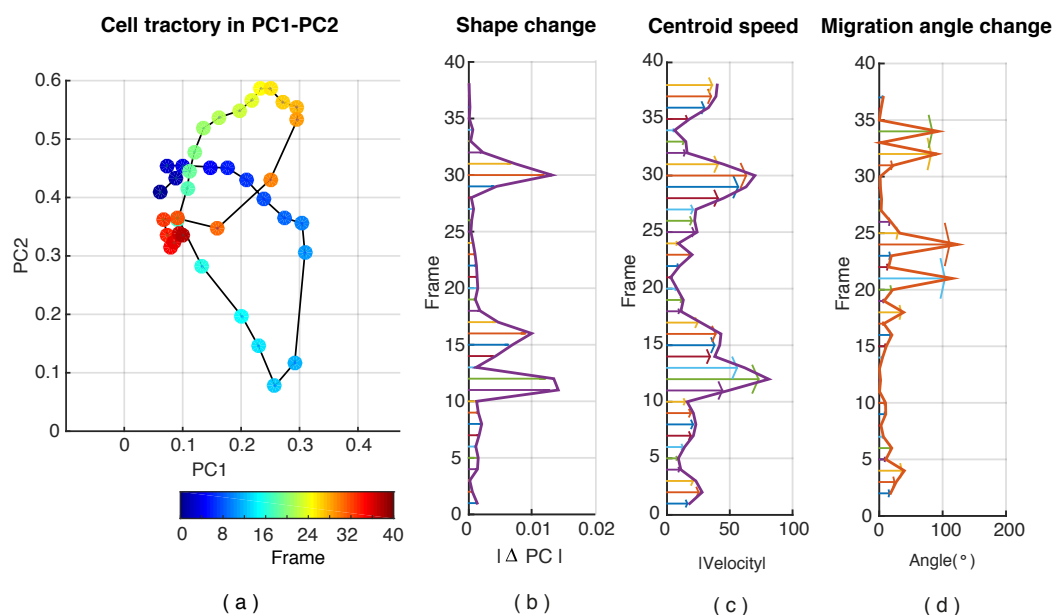
Figure 4.14a shows the cell shape evolution in shape feature space  $PC1 - PC2$  with the trajectory showing that the shape changes are fast at some stages. The cell shape changes, centroid velocity and migration angle changes over time are presented in Figure 4.14b, Figure 4.14c, Figure 4.14d, respectively. The first significant shape change occurred from frame 11 to 17 and cell speed increased simultaneously. The migration angles changed after the cell shape changes and the migration speed became small. Further on, the cell experienced a similar migration process that started with cell shape and speed increasing in frame 29 and followed by migration directional change in frame 32.



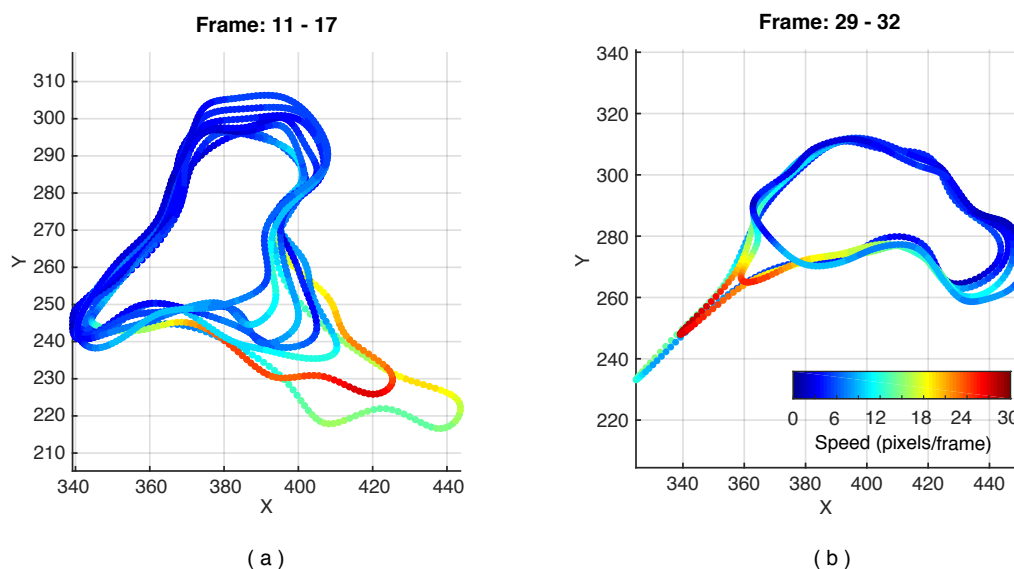
**Figure 4.13: Tracking of single cell migration.** (a) Cell boundary tracking. The speeds of cell boundary correspond to the colour bar. (b) Cell centroid tracking. The centroid velocities are represented as vectors with magnitudes indicated by length and directions indicated by arrows and cell centroids being treated as the origin. The colours indicate the time (frame) index.

Figure 4.15 shows those two periods of change (frames 11 – 17 and 29 – 32) when the cell changed shape significantly and it migrated quickly. In both periods, the cell experienced a tail contraction in which the cell rear was moving with a larger speed than the other parts of the cell boundary. The front of the cell remained mostly unchanged.

In summary, the integrated motility and morphology analysis on this cell reveals that migration directional changes was associated with cell tails shrinking. Cell tails moved quickly, and resulted in fast cell shape change. A slow migration in a new direction generated once the tail contraction was completed. Further single cell studies are required to investigate the cellular behaviours that may be presented in general. More importantly, the quantitative integrated analysis also shows the potential for identifying cellular behavioural characteristics that may have time correlation rather than association of characteristics at a particular time instant.



**Figure 4.14: Motility and morphodynamic characteristics of single cell migration.** (a) Cell shape evolution in shape feature space, where colour is used as time (frame) index. (b) Cell shape changes over time. (c) Cell centroid speed over time. (d) Cell migration angle changes over time.



**Figure 4.15: Shape tracking in single cell migration.** The magnitude of the cell boundary speeds are shown in colours, corresponding to the colour bar, on the cell boundary. (a) For frames 11 – 17 (b) For frames 29 – 32.

## 4.5 Conclusion

In this chapter, a module for cell morphodynamics analysis was proposed. It consisted of cell shape descriptions, shape clustering, shape evolution modelling and integrated analysis of cell morphology with motility at population and single cell migration. The module is designed as an extension to the cell tracking framework and was demonstrated by its application to the image sequences of wild breast cancer cells.

In shape description, the analysis is performed by determine the FDs and their dimension reduced by keeping the principal frequency components. The FDs, as generic spectral shape descriptors, have the physical explanation associated with each of the components as well, and are invariant to shape shifting, rotation and scale change. To only keep the strong frequency components, extracted discriminatory information about the FDs from the many cells used in the analysis. The population of cell shape distribution in this reduced 2-D shape feature space was then modelled by a Gaussian mixture model and assigned to clusters of round, elongated and teardrop shape classes.

For shape evolution over time, a HMM was estimated using the data from the shape feature space. In the process, the Gaussian mixture model was utilised as a prior to generate the emission properties of the HMM. The integrated analysis of cell motility and morphology was then performed by combing shape classification results with cell centroid motility features. The quantitative morphodynamic analysis showed that the round shape is the most stable state of breast cancer cells and that the teardrop shape was the state, which transited more readily. The analysis module demonstrated that it can reveal migration characteristics at a single cell and population levels. Specifically, it is able to capture time correlations associated with specific cellular behavioural patterns.

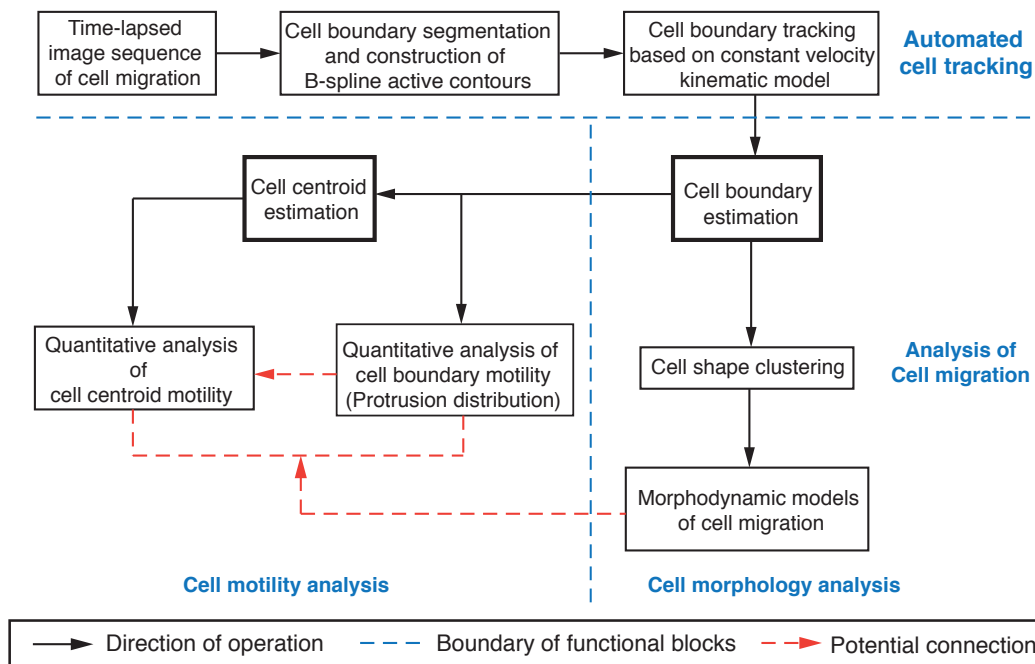


## Chapter 5

# Comparison of MDA-MB-231 Cell Behaviours with Regulating Migration-Associated Proteins

Cell migration is the foundation of tumour invasion, which is the essential reason leading to the death of tumour patients [146]. Vast numbers of drugs have been designed on the basis of either inhibiting or modifying cell migration. Cell migration is an integration of a sequence of locomotion, like morphology polarization, membrane extension and attachments formation and stabilization, and are accomplished by spatial and temporal coordination of distinctive molecular mechanisms [85]. Any change to those molecular structures or assemblies can affect the way that cells migrate. In the cancer research laboratory, different chemical inhibitors that act on migration-associated molecular components have been applied in cell cultivation to identify the underlying molecular functions, at the same time, explore the efficient way to regulate cell migration [171].

Currently, the evaluation of cell behaviour alterations under different cultivation environments are generally performed based on visual inspection or manual counting with applying different chemical indicators [122, 142]. This implies unacceptably high time consuming human effort and can lead to mistakes. With the development of computer vision technology, this requirement for an efficient and systematic cell behaviour analysis system must tackle the complexity of cell migration mechanism and the large amount of types of data that may be derived. By combining the methods discussed in Chapter 3 and Chapter 4, a framework that integrates automatic cell boundary tracking, cell motility and morphology analysis, is constructed and used to assess cell migration (see Figure 5.1). Unlike classical separate cell motility and morphology analysis, the framework is un-



**Figure 5.1: Diagram of the integrated cell migration analysis framework.** The framework consists of automatic cell boundary tracking and subsequent cell motility and morphology analysis. The potential correlation, marked by the red arrows, between cell morphology and motility are also discussed in this chapter.

derpinned by the parametric active contours used for cell boundary description, which intrinsically links cell morphology to motility.

In this chapter, the proposed framework is illustrated to measure the migration of MDA-MB-231 breast cancer cells subjected to experimental regulation of different migration-associated proteins.

ROCK and myosin II are two different functional proteins involved in cell migration due to their contribution to the cell adhesion dynamics, which is an essential process driving cells to migrate, as introduced in Section 2.1.4. Blebbistatin and Y-27632 are chemical inhibitors of ROCK and myosin II, respectively. They have been widely applied to cell cultivations in laboratory to infer the functions of ROCK and myosin II in cell migration [4, 42, 128]. In Section 5.2, the proposed framework is applied to quantitatively compare the motility and morphology of control, blebbistatin and Y-27632 activated cells, in order to explore whether the inhibitions of ROCK and myosin II, which work in the same pathway, regulate cell migration identically.

Apart from the inhibition of functional proteins, cell behaviours are experimen-

tally regulated by control of the migration-associated gene expressions [79, 129]. The expressions of FAK, RhoE and  $\beta$ -PIX, respectively, contribute to different steps of cell focal adhesion continuum, which generates traction stress to drive cell migration [50]. In Section 5.3, MDA-MB-231 cells were cultivated with knocking down genes: FAK, RhoE and  $\beta$ -PIX. The integrated analysis framework is applied to evaluate cell migrations and subsequently to quantitatively compare the regulation functions in the absence of different genes.

## 5.1 Experimental procedures

To validate the assessment framework for cell migration, two groups of experiments were performed by applying chemical inhibitors to influence different protein expressions that are likely to be associated with cell migration.

The cell lines used are commercially available from American Tissue Cell Culture (ATCC, Manassas, VA, USA). The experiments were performed by Dr. Julia E. Sero in the dynamical cell systems team, division of cancer biology, Institute of Cancer Research, London, UK.

### 5.1.1 Introduction of experimental plasmids

In the first experiment, MDA-MB-231 cells were treated by different chemical inhibitors, blebbistatin and Y-27632, both acting on proteins in the same pathway:

- Blebbistatin is a small molecule inhibiting activity of non-muscle myosin II, which plays an essential role in cell adhesion maturation and subsequent cell migration (see Figure 2.4 in Section 2.1.4). In one study, blebbistatin was applied to pancreatic adenocarcinoma to investigate the myosin II functions in cell migration. This study showed that blebbistatin had a negative effect on cell focal adhesion formation, through which the cytoskeleton activity is regulated, inhibiting cell adhesion, migration and spreading [42].
- Y-27632 is a biochemical inhibitor of ROCK, which is similar to myosin II which contributes to cell adhesion maturation. It has been shown that Y-27632 inhibits migration of rat ascites hepatoma cells through the suppression of cell focal adhesion and actin bundle, which are essential for cell movement [63].

Next, MDA-MB-231 were cultivated by knocking down FA associated genes, including FAK, RhoE and  $\beta$ -PIX. These genes have been identified to play a role in cell migration and are briefly described below:

- FAK-encoded protein is mainly expressed during the focal adhesion formation and boundary spreading. Compared to normal cells, the FAK activity is detected to increase in tumour cells, such as prostate, thyroid and breast tumours [150]. Aberrant expression of FAK have been experimented in a variety of cell types and various regulation functions on cells are reported. For example, van der Gaag et al. proves that overexpression of FAK can encourage the cell spreading of melanoma cancer [152]. Ward et al. shows that depletion of FAK inhibits tumour growth of mice ovarian carcinoma cells [159].
- RhoE is a member of Rnd proteins and is involved in many cellular activities, such as cell migration, epithelial polarity and cytoskeletal dynamics [153]. It is proved that the RhoE expression is essential for preventing the transformation of human prostate cancer cells [9]. Beyond prostate cancer cells, the functions and mechanisms of RhoE have not been fully understood since it is found in many cell events [9].
- $\beta$ -PIX is found located in nascent adhesion and is dissociated after adhesion becomes mature. An interaction exists between myosin II, which is essential for FA growth and maturation, and  $\beta$ -PIX. Hence cell migration and morphology are affected by a change to  $\beta$ -PIX. For example, for human corneal epithelial cells, overexpression of  $\beta$ -PIX negatively regulates adhesion maturation, but promotes cell migration [79, 82].

### 5.1.2 Experimental data

MDA-MB-231 breast cancer cells were labelled with CellTracker Orange (excitation wavelength of 561 nm and emission wavelength of 600 nm; Invitrogen) and cultured on tissue culture plastic in DMEM plus 5% heat-inactivated FBS. Time-lapse grey-scale imaging was performed using an Opera Cell::Explorer automated spinning disk confocal microscope (PerkinElmer) with a climate control chamber (37°C, 5% CO<sub>2</sub>). Images of revolution 500 × 664 were acquired at 5 minutes intervals with a 20× water objective lens ( $NA = 0.7$ ) with 2 × 2 binning. Movement correction was applied to time-lapse movies using Velocity (PerkinElmer).

Table 5.1 provides the experimental conditions used and datasets obtained for the comparison of myosin II and ROCK inhibitors. Table 5.2 lists the experimental conditions and datasets obtained for the comparison of different FA associated genes, FAK, RhoE and  $\beta$ -PIX.

Table 5.1: Data used for comparison of blebbistatin and Y-27632.

Chemical inhibitors	Data sets	Interval	Video duration
None (control sets)	live control J21-F1 live control J21-F3 live control J21-F5	5 mins	38 frames
Blebbistatin	live Blebb J18-F5 live Blebb J18-F7 live Blebb J19-F2	5 mins	38 frames
Y-27632	live Y27 J15-F1 live Y27 J15-F7 live Y27 J16-F2	5 mins	38 frames

Table 5.2: Data used for the comparison of focal adhesion associated genes.

Knockdown genes	Data sets	Interval	Video duration
None (Control sets)	control 3-13-2 control 3-14-1 control 3-14-2	5 mins	220 frames
FAK	FAK 5-13-3 FAK 5-13-5 FAK 5-14-5	5 mins	220 frames
RhoE	RhoE 7-13-3 RhoE 7-13-5 RhoE 7-15-4	5 mins	220 frames
$\beta$ -PIX	$\beta$ -PIX 4-13-2 $\beta$ -PIX 4-13-3 $\beta$ -PIX 4-14-1	5 mins	220 frames

## 5.2 Quantitative motility and morphology analysis of MDA-MB-231 cells treated by blebbistatin and Y-27632

Myosin II and ROCK are essential proteins that affect cell migration through participation in cell focal adhesion maturation and growth (see Section 2.1.4). Blebbistatin and Y-27632 are small molecular inhibitors, acting on myosin II and ROCK, respectively, and have been studied mostly for their effects on kinases distributions using immunofluorescence techniques and vision inspections, such as the analysis of inhibition of cell blebbing by Y-27632, and the evaluation of blebbistatin effects on apoptosis [11, 78, 111, 120]. There are currently no systematic computational analysis of cell behaviours, which can be used for the comparison of the effects of blebbistatin and Y-27632 on cell migration in isolation from the complicated molecular process involved in cell signalling in cellular activities.

This section uses the proposed framework in Chapter 3 and Chapter 4 to compare cell migration, regulated by inhibition of myosin II and ROCK functions using blebbistatin and Y-27632, at the population level. The comparison consists of the analysis of cell centroid motility, cell boundary dynamics, cell morphology and morphodynamics.

### 5.2.1 Analysis of cell centroid motility

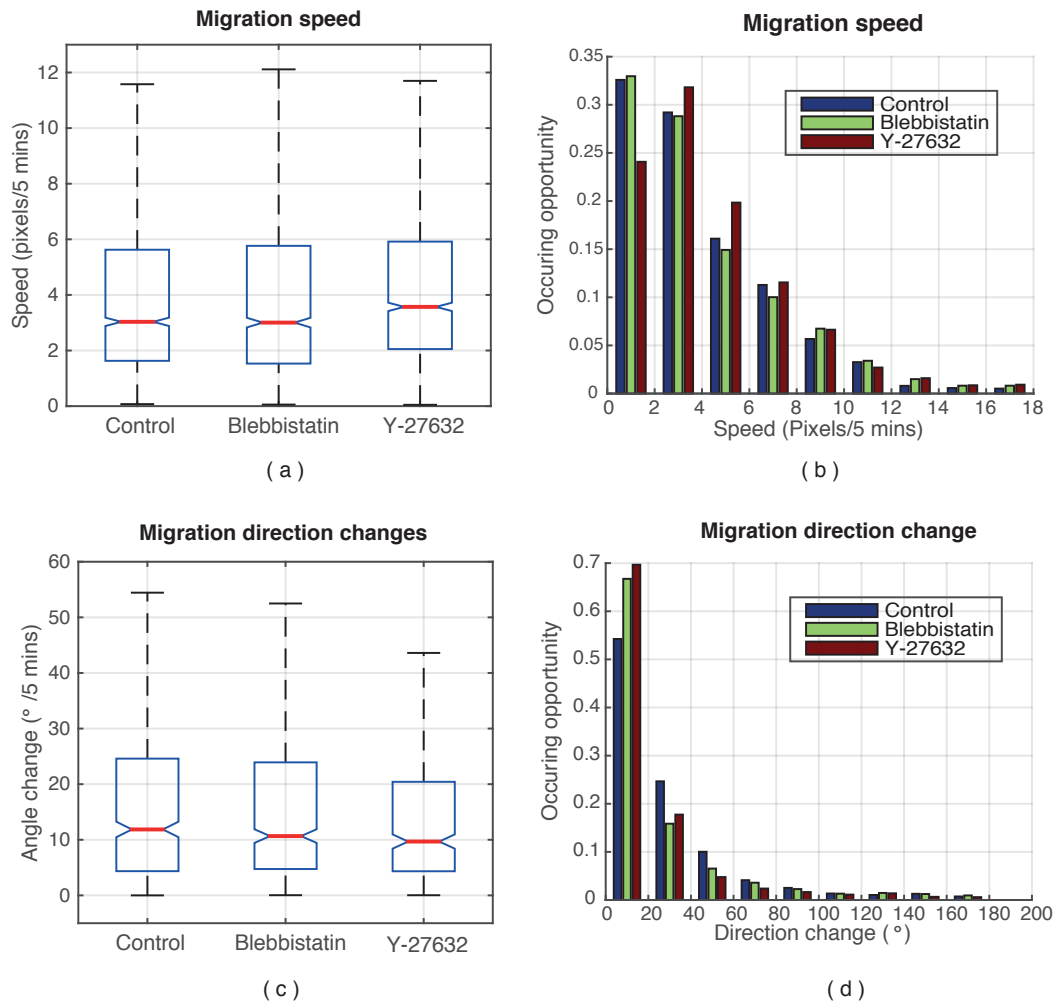
In Section 3.4, cell dynamics are described by cell trajectories and velocities using automated cell boundary tracking. To compare the effects of blebbistatin and Y-27632 on cell motility, the quantitative analysis of cell motility is presented according to movement of cell centroid and boundary dynamics.

Cell centroid motility characteristics, including migration speed, migration direction change, migration length, confinement ratio and average speed in respect of direction change, are calculated separately to characterise cell motility. Both chemical inhibitors suppress cell centroid motility, however they do so in a different way.

Cell migration speeds and direction change are calculated based on cell centroid tracking (Method is given in Section 3.4.1). Blebbistatin and Y-27632 treated cells show a similar average speed to control cells (control: 3.0 pixels/5 min, blebbistatin: 3.0 pixels/ 5 min, Y-27632 3.6 pixels/5 min; see Figure 5.2a). The speed histogram shows that more Y-27632 treated cells migrate faster than 2 pixels/5 min. Blebbistatin-treated cells, however, have similar distributions of migration speeds to the control cells (see Figure 5.2b).

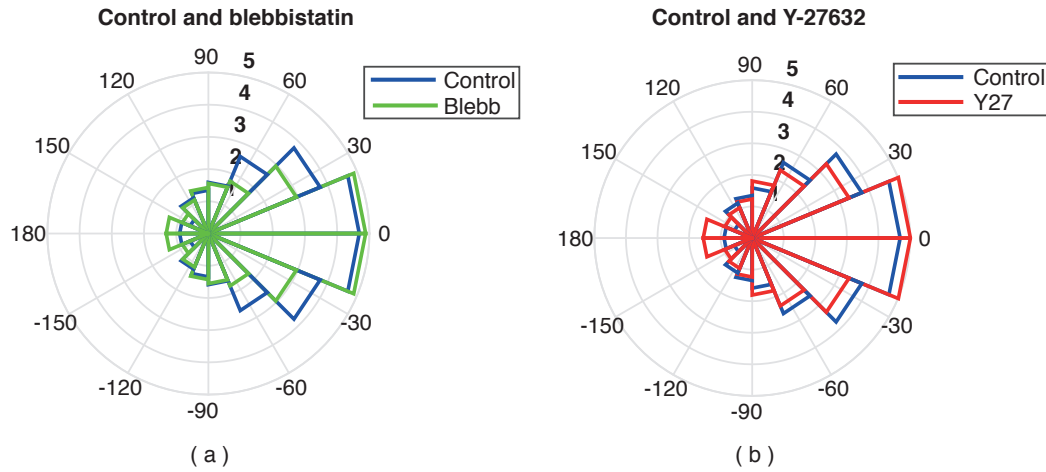
Figure 5.2c shows that Y-27632-treated cells migrate with smaller mean and narrowly distributed direction changes between adjacent frames, compared to con-

control and Blebbistatin-treated cells. In addition, the direction change histogram in Figure 5.2d shows that both inhibitors restrict cell migration direction changes in the 5 min study interval (occurrence of angle change less than  $20^\circ$ : control 54%; blebbistatin 68%; Y-27632 70%).



**Figure 5.2: Quantitative analysis of cell centroid speed and direction change.** (a) Centroid migration speed of cells subjected to blebbistatin and Y-27632 treatments and corresponding speed histogram (truncated) is given in subfigure (b); (c) Centroid migration direction change of cells subjected to treatment of blebbistatin and Y-27632 and the histogram (truncated) is given in subfigure (d). In subfigure (a) and (c), the red lines in the middle of the box correspond to the median values, the lower and upper edges of the box are used to mark the 25% and 75% of the data distributions, respectively. The whiskers outside of the box extend to the extreme values of the data.

Secondly, with the assumption that cell migration speed is associated to the



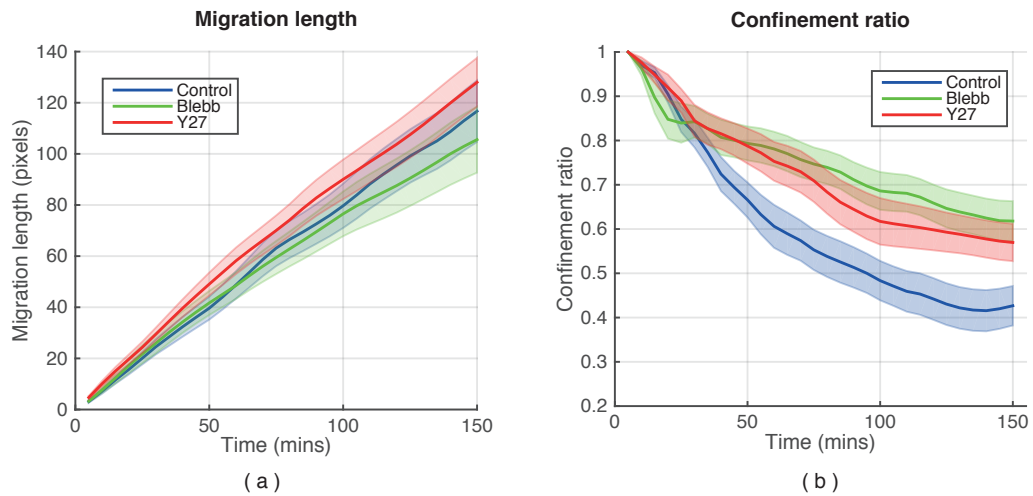
**Figure 5.3: Cell centroid speed in respect to migration direction change.** (a) Comparison between control cells and blebbistatin treated cells. (b) Comparison between control and Y-27632 treated cells. These polar charts are designed up-down symmetrical, angular coordinates represent cell migration direction changes and radial coordinates are the corresponding average centroid speed (pixels/5 mins)

orientation, cell average speeds in respect of migration direction change are presented in a polar chart, where angular and radial coordinates represent migration angle changes and corresponding average speeds, respectively (see Figure 5.3). Compared to the control cells, blebbistatin and Y-27632 mainly maintain cell speed when cells move forward ( $[0^\circ, 20^\circ]$ ). In other directions, blebbistatin strongly suppresses cell speed more than Y-27632. Hence, the control, Y-27632 and blebbistatin treated cells have increasing ability to maintain migration direction while blebbistatin treated cells exhibit a stronger ability to maintain migration direction.

Additionally, cell tracks over 30 frames are extracted to characterise cell ability of directional migration through cell migration length and confinement ratio. Figure 5.4a gives the comparison of cell migration length. Over longer time spans, blebbistatin reduces cell migration distance, Y-27632, on the other hand, improves the migration distance (blebbistatin 105 pixels; control 117 pixels; Y-27632 128 pixels at 150 minutes). Figure 5.4b shows the comparison of cell confinement ratio. Both inhibitors exhibited significantly larger confinement ratio than control cells after the first 50 minutes. This indicates that both inhibitors can promote cell directional migration, with blebbistatin exhibiting stronger effects (150 minutes, blebbistatin 0.62; Y-27632 0.57; control 0.45).

In summary, Y-27632 increased cell translocation whereas blebbistatin inhibited cell translocation, especially when observed over longer time period. How-





**Figure 5.4: Quantitative analysis of cell directional migration.** (a) Comparison of cell migration length over time. (b) Comparison of cell confinement ratio. The solid lines represent the average value and the transparent band is the standard deviation.

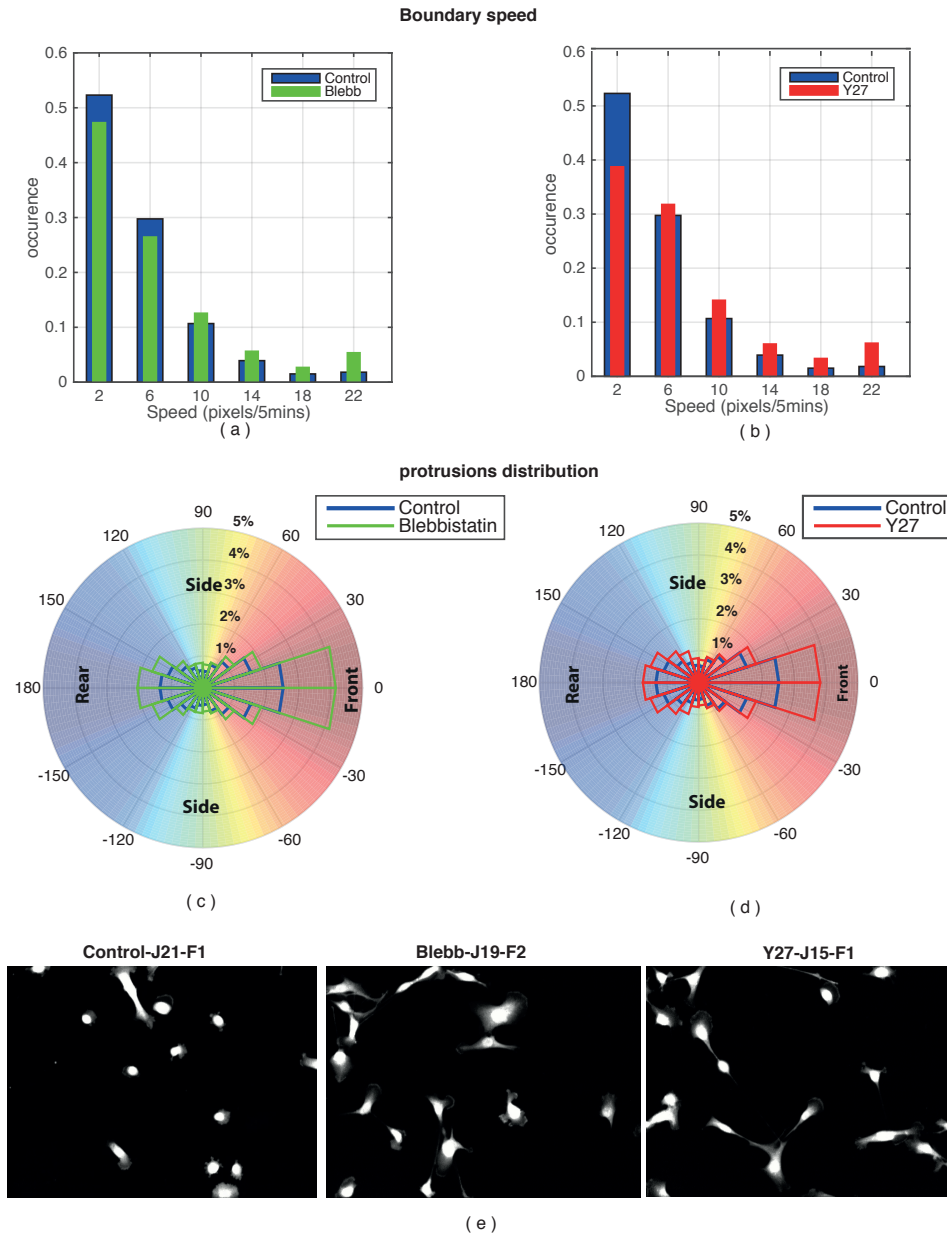
ever, both chemical inhibitors promoted MDA-MB-231 cell directionality in motility. Blebbistatin showed stronger effects on cells so as to maintain migration direction when compared to Y-27632, especially over longer time periods.

### 5.2.2 Analysis of cell boundary motility

Cell boundary dynamics is characterised by boundary velocities and protrusion distributions, which are calculated based on cell boundary tracking (method is given in Section 3.4.2).

Quantitative analysis of cell boundary motility is shown in Figure 5.5a and Figure 5.5b. The speed histograms show that blebbistatin and Y-27632-treated cells have smaller number of cell boundaries moving slower than 4 pixels/5 min (Control: 0.52; blebbistatin: 0.47; Y-27632: 0.39). This indicates that both chemical inhibitors promote cell boundary dynamics, and Y-27632 has a stronger effect than blebbistatin (see Figure 5.5a and Figure 5.5b).

Subsequently, in order to compare the protrusiveness of cells subjected to different treatments, the cell boundary was first identified as the cell boundary adjacent to the protrusions. The length of protrusive cell boundary was normalised against the entire cell boundary length and compartmentalised into the different angles associated with cell orientation (Methods are given in Section 3.4.2 and results are shown in Figure 5.5c and Figure 5.5d). All cells exhibit similar dis-



**Figure 5.5: Quantitative analysis of cell boundary motility.** (a) Speed histogram of cell boundary (control and blebbistatin-treated cells); (b) Speed histogram of cell boundary (control and Y-27632-treated cells). (c) Protrusion distribution of control and blebbistatin treated cells; (d) Protrusion distribution of control and Y-27632 treated cells. The polar histograms are designed up-down symmetric. Angular coordinates represent the angle to cell migration direction. Front, side and rear indicate toward, vertical and opposite direction of cell movement, respectively. Radial coordinates represent the occurrence of boundary in the corresponding angle range adjacent to protrusions and normalised by the length of entire cell boundary. (e) Images extracted from cell migration image sequences. From left to right, they belong to control, blebbistatin and Y-27632 treated cells, respectively.

tributions: Most protrusions were observed at the cell boundaries close to cell front. Fewest protrusions were observed on both sides of the cell boundary. Compared to control cells, blebbistatin and Y-27632 both significantly increased cell protrusions occurring at their front (i.e.,  $0^{\circ}$ – $18^{\circ}$ : control, 2.5%; blebbistatin, 4.2%; Y-27632, 3.8%), as well as a minor increase at their rear. In particular, blebbistatin-treated cells displayed a larger amount of protrusions and greater protrusion size distribution at the front from  $0^{\circ}$  to  $54^{\circ}$ .

In summary, blebbistatin and Y-27632 both increased the activities of the cell boundary, as well as increased the number of protrusions occurring at the cell front. Compared to Y-27632, blebbistatin-treated cells showed an increased size distribution at the front of the cells. This analysis is consistent with the observations in cell migration. Extracted frames from cell migration image sequences are shown in Figure 5.5e.

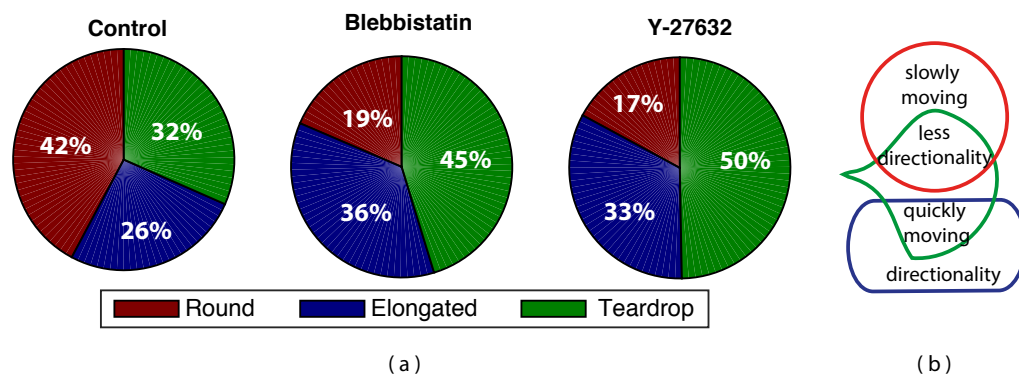
### 5.2.3 Integrated analysis of cell morphology and motility

Cell morphology is analysed based on three different shape states: round, elongated and teardrop, of which the membership probabilities are derived from control datasets given in Figure 4.9 (method is discussed in Chapter 4). Cell morphology is characterised by the population of shape states and is combined with an analysis of centroid motility characteristics.

The shape states population in Figure 5.6a shows that blebbistatin and Y-27632 identically affect cell shape distributions. Both chemical inhibitors decrease the percentage of round shape state (control: 42%, blebbistatin: 19%, Y-27632: 17%), but increase the percentage of the other two states, where the teardrop shape state becomes the most frequent.

In the integrated motility and morphology analysis of control cells (see Section 4.4.1), different shape states had distinct motility features, which are summarised in Figure 5.6b.

A decrease in percentage of cells with round shape, which are characterised by lower amount of movement, indicate more cells are moving faster overall. Elongated cells showed the strongest capability of maintaining direction; thus, an increase in their proportion indicated that directional cell motility had been promoted by a treatment. Blebbistatin treatments showed the strongest effects. The inference of cell motility from cell shape proportions matches the quantitative analysis of cell motility given in Section 5.2.1.



**Figure 5.6: Population analysis of cell shape state.** (a) Proportion of different cell shape states in datasets. From left to right: control, blebbistatin-treated and Y-27632-treated cells. (b) Motility features of different cell shape states.

#### 5.2.4 Analysis of cell morphodynamics

Cell morphodynamics are characterised by the dynamics of shape transitions, which are first modelled as a Markov model using the shape classes identified from Gaussian mixture models. The transition matrix is estimated through the number of shape transition occurring between two adjacent frames. The accuracy of this model is limited by the assumption that shape classes were accurately classified. Therefore, this accuracy will be limited by insufficient data, especially for the shape transitions that occurred less frequently when compared to shapes remaining unchanged. However the counts of shape transitions can reveal if different treatments alter the probabilities of cell shape transitions. In recognition of the strong assumption that limits the accuracy of the results, a HMM that regards shape classes as hidden variables is employed to improve the modelling of shape evolution. The corresponding methods are discussed in Section 4.3.1 and Section 4.3.2, respectively.

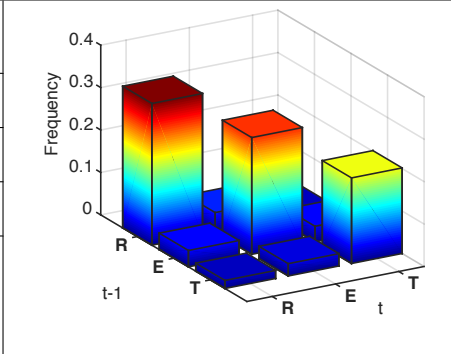
Cell shape evolutions were initially represented as a Markov model, where the shapes were classified according to the Gaussian mixture model identified from control cells (parameters are provided in Table 4.2). Table 5.3 (control), Table 5.4 (blebbistatin) and Table 5.5 (Y-27632) show the counts of cell shape evolutions and show the corresponding 3D bar charts normalised by the sum of all different shape evolutions. These tables show that both chemical inhibitors significantly reduce the proportion of round cells that remain round in the subsequent frame. Maintaining a round shape between two frames was the most frequent shape state evolution for the control cells, but is the least for blebbistatin- and Y-27632-treated cells. In terms of overall transitions, in respect of cells activated by blebbistatin

and Y-27632, the transition states with the highest frequency were those that maintained elongated and teardrop shapes in subsequent frames, respectively.

**Table 5.3: Count of shape transitions between adjacent frames (Control cells).** The corresponding normalised histogram is shown in the right.

$t-1 \backslash t$	R	E	T
R	563	80	31
E	65	464	62
T	32	49	340

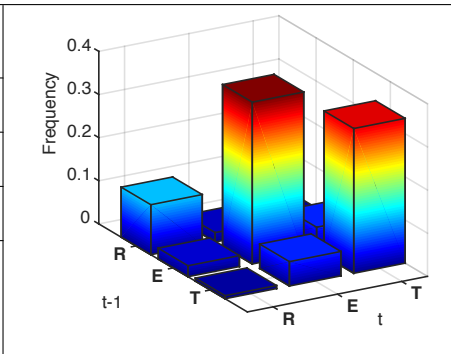
\* R, E, T denote the shape of round, elongated and teardrop, respectively.



**Table 5.4: Count of shape transitions between adjacent frames (Blebbistatin).** The corresponding normalised histogram is shown in the right.

$t-1 \backslash t$	R	E	T
R	164	31	9
E	36	531	79
T	10	80	475

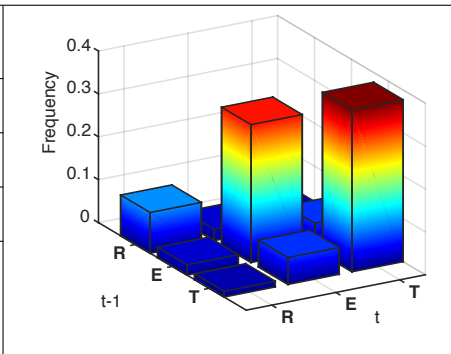
\* R, E, T denote the shape of round, elongated and teardrop, respectively.



**Table 5.5: Count of shape transitions between adjacent frames (Y-27632).** The corresponding normalised histogram is shown in the right.

$t-1 \backslash t$	R	E	T
R	149	42	28
E	38	504	97
T	21	98	592

\* R, E, T denote the shape of round, elongated and teardrop, respectively.



Cell shape evolutions were then represented as a HMM and cell shapes as-

signed as latent variables, with membership probabilities of cell shapes determined from control cells with parameters given in Table 4.5. The identified transition matrices are given in Table 5.6.

In comparison of control cells, blebbistatin and Y-27632 showed similar effects on the cell ability to maintain shapes. Specifically, both inhibitors significantly increased the sojourn time of the teardrop shape state (control: 15 min; blebbistatin: 100 min; Y-27632: 191 min), accompanied with a sojourn time decreasing for the round shape state (control: 753 min; blebbistatin: 34 min; Y-27632: 11 min). This is consistent with the transition results given in Table 5.3, Table 5.4 and Table 5.5. Furthermore, blebbistatin-treated cells had longer sojourn time maintaining the round state compared with Y-27632-treated cells (blebbistatin: 34 min; Y-27632: 11 min). Blebbistatin, at the same time, increased the sojourn time of the elongated state (control: 28 min; blebbistatin: 81 min; Y-27632: 28 min). Y-27632 had a stronger effect in maintaining the teardrop shape state, which had a higher probability and longer corresponding sojourn time.

Any transition with a transition probability greater than 0.1 is deemed significant enough. Control cells most frequently transition to the round state. Blebbistatin treated cells do not show any noticeable shape transition, whereas Y-27632 increases the transition to the teardrop shape state.

In conclusion, Y27 and blebbistatin showed similar effects on the migration of MDA-MB-231 breast cancer cells, although these effects were manifested with distinct features. Both chemical inhibitors promoted cell directional migration, increased cell protrusion distributions, especially at the front of cells, and decreased cell stability of the round shape state. Blebbistatin played a stronger role in the restriction of cell migration direction while Y-27632 treatment resulted in more frequent cell shape transitions.

### **5.3 Quantitative motility and morphology analysis of MDA-MB-231 cells in the absence of focal adhesion associated genes**

FAK, RhoE and  $\beta$ -PIX are three different genes associated with FA and they are expressed at different stages of the adhesion continuum (see Figure 2.4 and Section 2.1.4). The roles that these genes play in cell behaviour such as cell spreading, adhesion and cell invasion have been explored through the enhancement or impairment of gene expressions. The functions are examined by observing the changes in cell phenotypes corresponding to the alteration in gene expressions

Table 5.6: The shape transition relations of breast cancer cells treated with chemical inhibitors blebbistatin and Y-27632.

Treatments	Transition diagram	Sojourn time (minutes)
Control cells	<p>control 0.9934</p> <p>round</p> <p>0.2716 (round to teardrop), 0.0029 (teardrop to round), 0.1696 (round to elongated), 0.0037 (elongated to round)</p> <p>teardrop (0.6665), elongated (0.8197)</p> <p>0.0019 (teardrop to elongated), 0.0107 (elongated to teardrop)</p>	$\tau_R = 753$ (1) $\tau_E = 28$ (0.0372) $\tau_T = 15$ (0.0199)
Blebbistatin	<p>Blebb 0.8548</p> <p>round</p> <p>0.0111 (round to teardrop), 0.0985 (teardrop to round), 0.0057 (round to elongated), 0.0466 (elongated to round)</p> <p>teardrop (0.9500), elongated (0.9382)</p> <p>0.0389 (teardrop to elongated), 0.0571 (elongated to teardrop)</p>	$\tau_R = 34$ (1) $\tau_E = 81$ (2.3824) $\tau_T = 100$ (2.9412)
Y-27632	<p>Y27632 0.5305</p> <p>round</p> <p>0.0074 (round to teardrop), 0.4027 (teardrop to round), 0.0021 (round to elongated), 0.0657 (elongated to round)</p> <p>teardrop (0.9739), elongated (0.8185)</p> <p>0.0187 (teardrop to elongated), 0.1794 (elongated to teardrop)</p>	$\tau_R = 11$ (1) $\tau_E = 28$ (2.5454) $\tau_T = 191$ (17.3636)

\* (·) is the normalised sojourn time with respect to round shape state.

[49]. There have been only a few cases of computational analysis of gene expression associated with cell motility or cell morphology characteristics. A computational analysis framework that can quantify cell motility and morphology characteristics is necessary to compare the influence of expressions of different genes that are associated with cell FA system on cell migration.

In this section, the proposed integrated analysis in Figure 5.1 is used to quantitatively compare cell migration regulated by depletion of genes FAK, RhoE and  $\beta$ -PIX, respectively. The motility is analysed from the view of cell centroid migration as an overview of cell motility and cell boundary movement as a measure of

cell cytoskeleton dynamics. Meanwhile, cell morphology is examined separately in an integrated cell morphology and motility analysis, and the study of cell morphodynamics is examined through the analysis of cell shape evolutions between adjoining frames.

### 5.3.1 Analysis of cell centroid motility

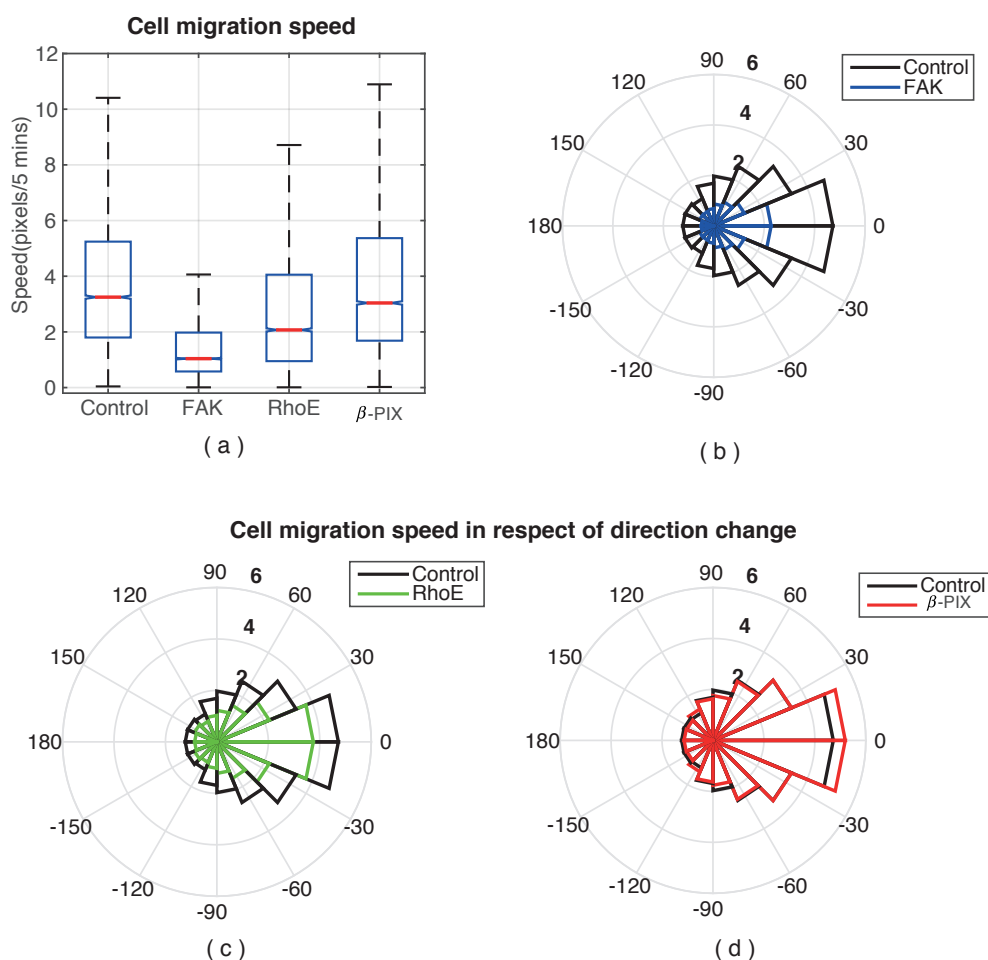
As illustrated in the system diagram in Figure 5.1, automated cell boundary tracking is performed initially to describe the dynamics of multiple cells. Cell centroid motility characteristics are organised based on static and dynamic analysis.

The quantitative analysis of cell centroid speed and migration speed in respect of direction changes are given in Figure 5.7. The distributions of cell migration speeds shows that cell speed is inhibited mostly by the depletion of FAK, as well as a minor inhibition in RhoE knockdown cells but not in the depletion of  $\beta$ -PIX (Median speed in 5 minutes: control 3.25 pixels, FAK 1.04 pixels, RhoE 2.07 pixels,  $\beta$ -PIX 3.04 pixels, see Figure 5.7a). With respect to the direction changes, FAK knockdown cells migrate significantly slower independently of any direction changes, compared to control cells (see Figure 5.7b). RhoE knockdown cells, similarly, migrate slower, but shows less speed inhibition compared to FAK knockdowns (see Figure 5.7c).  $\beta$ -PIX-depleted cells have higher speeds when direction changes are smaller than  $22.5^\circ$  (see Figure 5.7d).

Considering the numbers of cells involved, cell trajectories of 250 minutes are extracted. Cell migration lengths and confinement ratio changes over time are shown in Figure 5.8. Compared to control cells, FAK knockdown cells present a significantly inhibited average migration length. A decrease in rate of change in migration length occurred at 50 minutes, which indicates a stronger suppression in cell migration. Whereas, RhoE and  $\beta$ -PIX knockdowns show a similar migration length to control cells (see Figure 5.8a). The corresponding confinement ratios show that cell directional persistence is significantly reduced by the depletion of FAK and  $\beta$ -PIX, but not RhoE (see Figure 5.8b).

In summary, knockdown of FAK, RhoE and  $\beta$ -PIX affect cell motility differently. The depletion of FAK significantly inhibited both cell migration speed and direction persistence and the inhibition effects were stronger with the increase of cultivated time. The knockdown of RhoE reduced cell migration speed but not direction persistence, whereas the  $\beta$ -PIX knockdowns reduced migration direction persistence but migrated with similar speed to control cells.

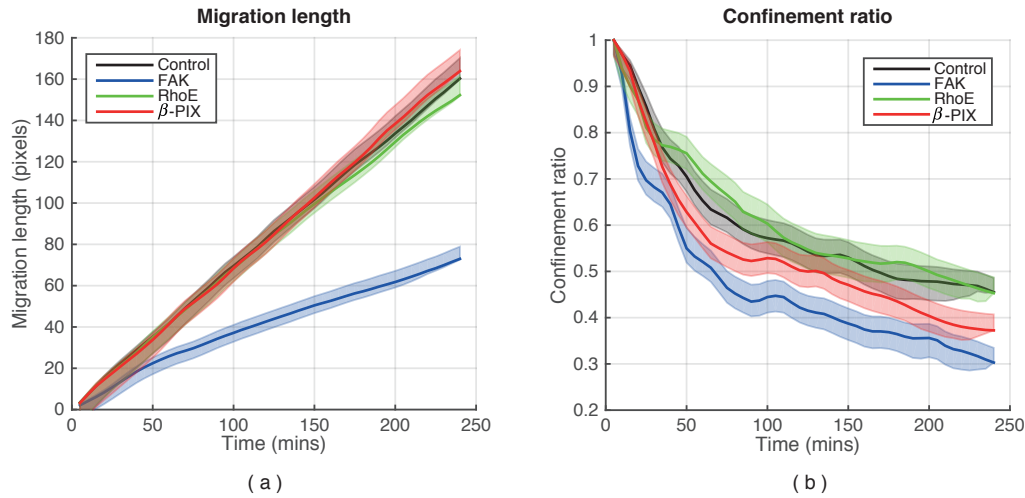




**Figure 5.7: Quantitative centroid motility analysis of breast cancer cells subjected to knockdown of focal adhesion associated genes.** (a) Comparison of cell centroid speeds. Median speeds are marked by the red lines, 25% and 75% of the speed distributions are marked by the lower and upper edges of the boxes and the whiskers extend to the extreme value of cell speed. (b)–(d) Comparison of cell average speed with respect to direction changes of situations: (b) control vs. FAK knockdown cells; (c) control vs. RhoE-depleted cells; (d) control vs.  $\beta$ -PIX. These polar charts are up-down symmetric, angular coordinates are cell migration direction changes ( $^{\circ}$ ) and radial coordinates are the corresponding average cell speeds.

### 5.3.2 Analysis of cell boundary motility

The analysis of cell local dynamics/cell boundary protrusiveness, histogram of cell boundary speeds and normalised protrusion distributions are shown in Figure 5.9 (methods are given in Section 3.4.2).

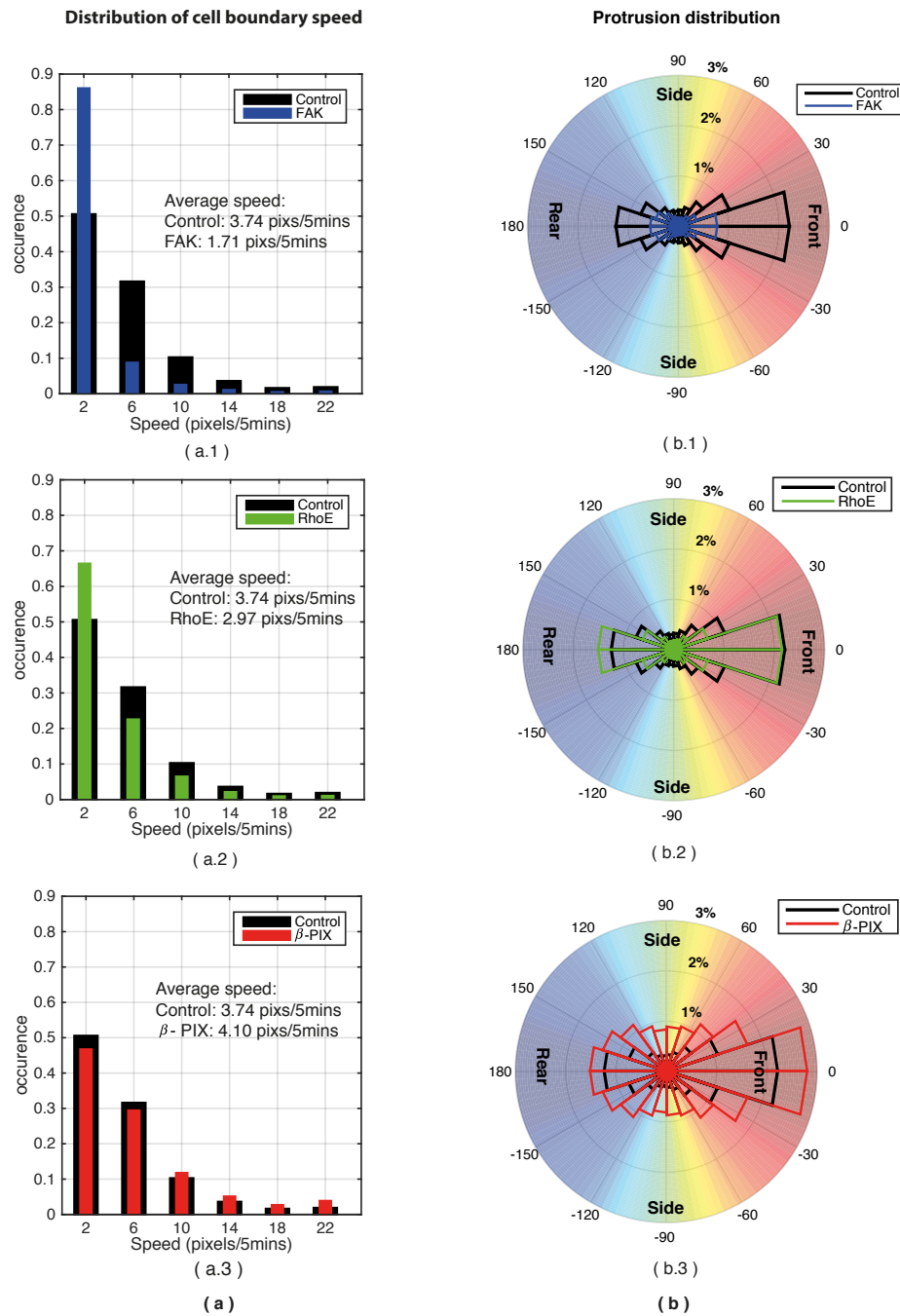


**Figure 5.8: Cell directional migration analysis.** (a) Comparison of cell migration length over time (0–250 mins). (b) Comparison of cell confinement ratio over time (0–250 mins). Solid lines are the mean values and apparent bands are the width of corresponding standard deviation.

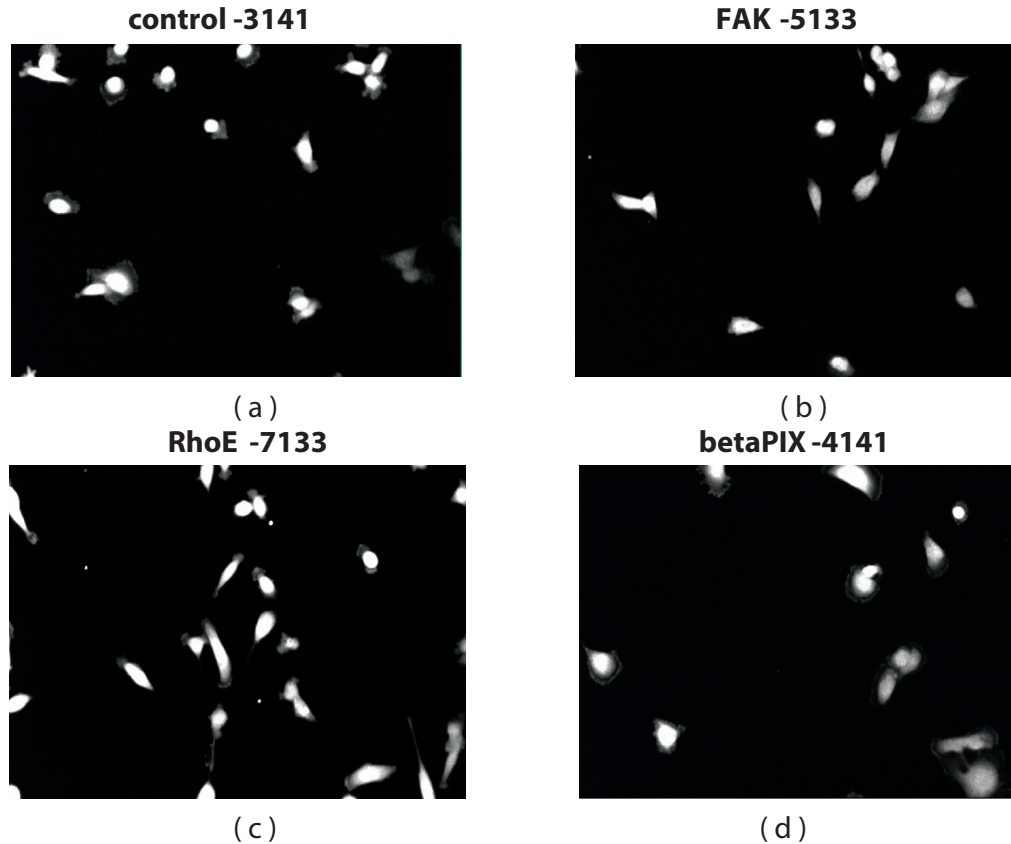
The depletion of FAK significantly restricts cell protrusiveness. In particular, 86% of cell boundaries move with speed slower than 4 pixels per 5 minutes compared to 51% of control cells, hence the average speed is reduced to 1.71 pixels/5 mins from 3.74 pixels/5 mins in control cells (see Figure 5.9a.1). Additionally, distributions of protrusions are reduced independent of any orientation angles (see Figure 5.9b.1), which matches the observations from the migration image sequences in Figure 5.10a and Figure 5.10b.

RhoE knockdown cells exhibit restrictions on cell boundary dynamics only at specific orientations. There are 66% of cell boundaries moving slower than 4 pixels in 5 minutes compared to 51% of control cells (see Figure 5.9a.2). Furthermore, the distribution of protrusions perpendicular to the travel direction are reduced while the protrusions located opposite to the migration direction are increased (see Figure 5.9b.2). This bipolar distributed protrusions is consistent with the observations in Figure 5.10c.

$\beta$ -PIX knockdown cells display a similar boundary speed as control cells (see Figure 5.9a.3). In terms of the distribution of protrusions,  $\beta$ -PIX knockdown cells show larger amount of protrusions independent of orientations (see Figure 5.9b.3). The observations of protrusion consistently display a more uniform distributional pattern (see Figure 5.10c). The protrusion distributions indicate  $\beta$ -PIX knockdown cells are unable to commit to a steady travel direction, which is an explanation of why knockdown of  $\beta$ -PIX persistently reduce cell migration directionality (see



**Figure 5.9: Quantitative analysis of cell boundary motility.** (a) Speed histograms of cell boundaries: (a.1) control vs. FAK knockdown cells; (a.2) control vs. RhoE knockdown cells; (a.3) control vs.  $\beta$ -PIX knockdown cells. (b) Protrusion distribution in respect of cell orientation: (b.1) control vs. FAK knockdown cells; (b.2) control vs. RhoE knockdown cells; (b.3) control cells vs.  $\beta$ -PIX cells. The polar histograms are up-down symmetric, angular coordinates are the angles to cell migration direction and radial coordinates are the percentage of the corresponding area adherent to protrusions over entire cell boundary.



**Figure 5.10: Cell migration images extracted from time-lapse image sequence.** (a) Control cell migration. (b) FAK knockdown cells. (c) RhoE knockdown cells. (d)  $\beta$ -PIX knockdown cells.

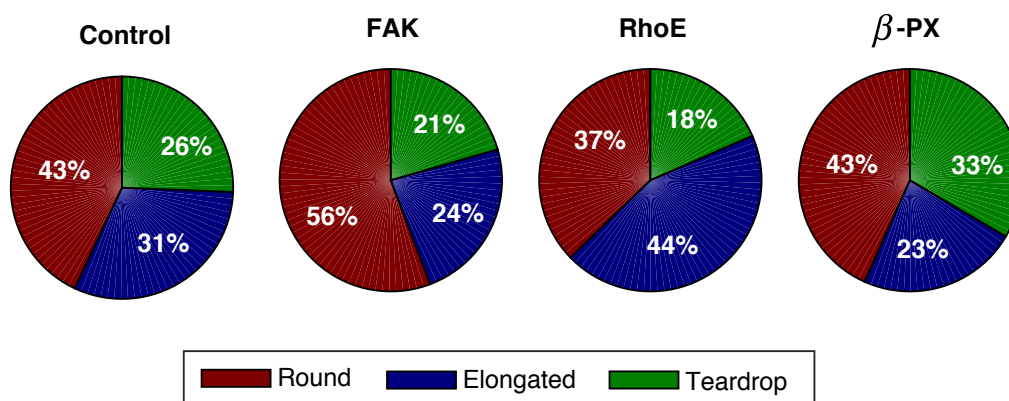
Figure 5.8b).

Depletion of FAK, RhoE and  $\beta$ -PIX, in comparison to control cells, each have distinct boundary motility characteristics, which indicate that those genes affect the cytoskeleton dynamics differently, altering cell morphology characteristics. This is discussed in the next section.

### 5.3.3 Integrated analysis of cell morphology and motility

Cell morphology is characterised by the membership probabilities of different shape states, which are identified from control cells. In this section, ratio of cells being different shape states is computed and is linked to cell motility using the characteristics of different shape states (derived in Section 4.4 and depicted in Figure 5.6b).

The ratios of the cell counts with different shape states for the various treat-



**Figure 5.11: Ratio of cells of different shape states.** From left to right, pie charts represent the ratio of different shape states of control, FAK knockdown, RhoE knockdown and  $\beta$ -PIX knockdown cells.

ments are shown as pie charts in Figure 5.11. Compared to control cells, knocking down FAK significantly increased the percentage of cells in the round shape state (control: 43%; FAK: 56%) with a corresponding drop in the percentages of both elongated and teardrop shape states. The over 50% of round cells indicates that FAK knockdowns should have motility characteristics similar to round shape state, where cells travel slowly and are difficult to maintain a persistent migration direction. This motility inference corresponds to the quantitative analysis of centroid motility shown in Figure 5.7a and Figure 5.8b in Section 5.3.1.

Knockdown of RhoE, distinctively, increased the percentage of cells in the elongated shape state (control: 31%; RhoE: 44%. see Figure 5.11). The elongated cells travelled with strong direction persistence, which can explain why RhoE knockdown cells migrate more slowly than control cells (see Figure 5.7a), but have a similar confinement ratio (see Figure 5.8b).

$\beta$ -PIX knockdown cells had an increased ratio of teardrop shape state, by suppressing the percentage of cells in elongated shape state (see Figure 5.11), which is the shape state with the strongest direction persistence. The decrease in elongated shaped cells indicates generally a lower direction persistence, and a decrease in the cell confinement ratio (see Figure 5.8b).

### 5.3.4 Analysis of cell morphodynamics

The morphodynamic analysis includes the calculation of shape transitions based on a Gaussian mixture model for cell shape classification, as well as the HMM for shape evolution.

**Table 5.7: Count shape transitions between adjacent frames (Control cells for gene tests).** The normalised histogram of counting is shown in the right.

$t-1 \backslash t$	R	E	T
R	2920	387	185
E	390	2906	286
T	160	301	1116

\* R, E, T denote the shape of round, elongated and teardrop, respectively.

The numbers of different shape transitions based on Gaussian mixture model is normalised by their sum, and is represented in a 3D histogram in Table 5.7 for control cells, Table 5.8 for FAK knockdown cells, Table 5.9 for RhoE knockdown cells and Table 5.10 for  $\beta$ -PIX knockdown cells. The depletion of different FA-associated genes reflects the distinctive alteration of cell shape transitions. Control cells, treated as the standard migration benchmark for breast cancer cells, have a high frequency of round shape cells, as well as elongated shape, and not transitioning their shapes at the next frame (Table 5.8). In the absence of FAK, the frequencies of most cell shape evolutions in the 5 minutes intervals were decreased except for round shape state being maintained. Shape transitions, especially, were decreased to a value close to zero. For instance, the numbers of transitions from teardrop to round shape were 46 times and from round to teardrop were 39 times, whereas the number of cells remaining round were 3015 (Table 5.8). The depletion of RhoE increased the maintenance of the elongated cell state rather than the round cells being dominant (Table 5.9). The  $\beta$ -PIX knockdown cells showed a similar pattern to the control cells, but with an increased frequency of elongated cells retaining its shape and decreased frequency of teardrop cells unchanged (Table 5.10).

When the shape transition is modelled through HMM, the transition diagrams and corresponding sojourn times are given in Table 5.11. Compared to control cells, the absence of FAK increased cell stability of all shape states. The probabilities of maintaining the shape states were all over 0.9, hence shape-type transition probabilities are all constrained to be smaller than 0.1. The significantly longer sojourn time of round shape state shows that cells are stuck in the round shape state ( $\tau_R = 676$  minutes;  $\tau_E = 85$  minutes;  $\tau_T = 51$  minutes).

Knockdown of RhoE extended the time cells maintained the elongated shape

**Table 5.8: Count of shape transitions between adjacent frames (FAK knockdown cells).** The normalised histogram of counting is shown in the right.

$t-1 \backslash t$	R	E	T
R	3015	160	39
E	161	1726	98
T	46	92	720

\* R, E, T denote the shape of round, elongated and teardrop, respectively.

**Table 5.9: Count of shape transitions between adjacent frames (RhoE knockdown cells).** The normalised histogram of counting is shown in the right.

$t-1 \backslash t$	R	E	T
R	2396	244	58
E	258	3799	161
T	54	167	727

\* R, E, T denote shape of round, elongated and teardrop

**Table 5.10: Count of shape transitions between adjacent frames ( $\beta$ -PIX knockdown cells).** The normalised histogram of counting is shown in the right.

$t-1 \backslash t$	R	E	T
R	2917	428	201
E	411	2337	405
T	191	398	1756

\* R, E, T denote shape of round, elongated and teardrop.

state and suppressed the time for being in a teardrop state. Elongated shape states had the highest probability of staying unchanged, as well as the longest sojourn time. The teardrop-elongated transition probability, meanwhile, was increased to a notable value 0.1638, compared to 0.0532 for control cells.

The absence of  $\beta$ -PIX, similar to control cells, kept the round shape state as



the state with the highest retention probability, but with a shorter sojourn time (control:  $\tau_R = 214$  min;  $\beta$ -PIX:  $\tau_R = 135$  min). The elongated state, with state self-transition probability 0.7918, was the most unstable state instead of the teardrop shape as it was for control cells. The most frequent shape transitions was from elongated to teardrop shape state ( $p(T_k | E_{k-1}) = 0.1223$ ).

In conclusion, due to the interaction between cell motility and morphology, cell behaviours are affected distinctively by the depletion of FA-associated genes. Knocking down FAK inhibited cell migration generally, so that cells moved slowly with gently spreading boundaries and hardly any shape transitions. RhoE knock-down cells mainly exhibited bidirectional migration with relatively active boundaries at front and rear orientations, and at the same time, these cells had the longest sojourn time in the elongated shape state. Depletion of  $\beta$ -PIX sped up cell movement but inhibited directional travelling by activating cell side-boundary dynamics. Meanwhile, cells were retained for a longer time in the teardrop than the elongated shape state.

## 5.4 Conclusion

An integrated framework that combines automated cell boundary tracking, quantitative analysis of cell motility and morphology can give insights into cell behaviour under different conditions. In this chapter, this integrated framework was used to analyse the regulated migration of MDA-MB-231 breast cancer cells subjected to treatments with different chemical inhibitors and with aberrant expressed genes.

The population level analysis showed that the inhibition of ROCK (blebbistatin treatment) and myosin II (Y-27632 treatments) induced some alterations on cell behaviours that were similar. Compared to wild MDA-MB-231 cells, Y-27632 and blebbistatin both promoted cell directional persistence during migration and increased the cytoskeletal dynamics, evidenced by the accelerated cell boundary velocity and increased protrusion distributions, especially at the cell leading edge. Moreover, blebbistatin and Y-27632 significantly reduced cell stability of round states in terms of less amount of round-shaped cells being observed and the reduced time in which the cells maintained their round shape state.

Cells in round shape state were characterised as being slow and migrating without directional preference. The morphological alterations that resulted in different distribution of cell shapes still showed evidence that for individual cell shapes, the motility characteristics were consistent with the findings in Chapter 4. In addition, the regulation of cell migration suggested that the chemotaxis



Table 5.11: The shape transition relations of breast cancer cells with different gene knocked-down.

Drugs	Transition diagram	Sojourn time (minutes)
Control cells	<p>control 0.9766</p> <p>round</p> <p>0.3438 (round to teardrop), 0.0059 (teardrop to round), 0.0607 (round to elongated), 0.0175 (elongated to round)</p> <p>teardrop (0.6029)</p> <p>0.0532 (teardrop to elongated), 0.0298 (elongated to teardrop)</p> <p>elongated (0.9094)</p>	$\tau_R = 214 (1)$ $\tau_E = 55 (0.2570T)$ $\tau_T = 13 (0.0607T)$
Cells with FAK knockdown	<p>FAK 0.9926</p> <p>round</p> <p>0.0699 (round to teardrop), 0.0030 (teardrop to round), 0.0426 (round to elongated), 0.0044 (elongated to round)</p> <p>teardrop (0.9010)</p> <p>0.0291 (teardrop to elongated), 0.0160 (elongated to teardrop)</p> <p>elongated (0.9413)</p>	$\tau_R = 676 (1)$ $\tau_E = 85 (0.1257)$ $\tau_T = 51 (0.0754)$
Cells with RhoE knockdown	<p>RhoE 0.9567</p> <p>round</p> <p>0.0766 (round to teardrop), 0.0052 (teardrop to round), 0.0144 (round to elongated), 0.0381 (elongated to round)</p> <p>teardrop (0.7596)</p> <p>0.1638 (teardrop to elongated), 0.0072 (elongated to teardrop)</p> <p>elongated (0.9784)</p>	$\tau_R = 115 (1)$ $\tau_E = 231 (2.001)$ $\tau_T = 21 (0.1826)$
Cells with $\beta$ -PIX knock-down	<p>BetaPIX 0.9629</p> <p>round</p> <p>0.0901 (round to teardrop), 0.0273 (teardrop to round), 0.0359 (round to elongated), 0.0098 (elongated to round)</p> <p>teardrop (0.8923)</p> <p>0.0175 (teardrop to elongated), 0.1223 (elongated to teardrop)</p> <p>elongated (0.7918)</p>	$\tau_R = 135 (1)$ $\tau_E = 24 (0.1778)$ $\tau_T = 46 (0.3407)$

\*  $(\cdot)$  is the normalised sojourn time with respect to round shape state.

of MDA-MB-231 breast cancer cells, which has been proposed to be essential for cancer metastasis [158], may be controlled by regulating the ROCK and myosin II in the cell.

However, distinct degrees of the effects on cell behaviours induced by blebbistatin and Y-27632 were noticed. Specifically blebbistatin showed stronger promotion of cell directional persistence while Y-27632 showed a larger effect on the cell being able to maintain itself in the round shape state. This may be caused by the underlying mechanisms of cell migration involving ROCK and myosin II or the indirect influence of a difference in chemical inhibitor concentration in cell culture. More laboratory experience and integrated analysis are therefore suggested to identify how MDA-MB-231 breast cancer cells respond to the different chemical inhibitor concentrations.

Next, the computational analysis of FA-associated gene expression showed that FAK, RhoE and  $\beta$ -PIX induced different effects on MDA-MB-231 breast cancer cell behaviours. Compared to the wild cells, knockdown of FAK inhibited cell motility and morphology simultaneously, evidenced by the decreased cell centroid migration speed and directional persistence, reduced cytoskeleton speed, reduced protrusion distributions, significantly increased presence of round shape and the time duration for cells to maintain their round shape state, and the reduced shape transition probabilities. Depletion of RhoE slowed down cell migration and was observed to have distributed protrusions at cell leading and contracting edges that were similar. Meanwhile, cell elongated shape state is dominant, including longer sojourn time and increased occurrence. Knockdown of  $\beta$ -PIX inhibited cell directional migration by equally increasing cell protrusion distributions at cell periphery. With regard to cell morphology, reduced  $\beta$ -PIX expression increased the cell being in the teardrop shape state, while the elongated shape state decreased.

The comparison of MDA-MB-231 breast cancer cell migration with depletion of FAK, RhoE and  $\beta$ -PIX showed that the knockdown of RhoE and  $\beta$ -PIX inhibited different aspects of cell motility while the knockdown of FAK generally reduced cell motility. Cells with reduced RhoE or  $\beta$ -PIX expression had higher morphodynamic variability compared to the reduction of FAK expression. This indicates that FAK expression is essential for MDA-MB-231 breast cancer cell migration. Meanwhile some studies showed that FAK is detected more in breast tumour cells than healthy cells [150]. The FAK is suggested as being a potential therapeutic target for breast cancer treatment.

## Chapter 6

# Conclusions and Future Work

Cell motility and morphology properties are a combined reflection of cellular and extracellular environments, and they mutually affect each other during cell migration. Compared to healthy breast cells, the cancer cells show abnormal motility and morphology properties, which are generally associated with cancer invasion and metastasis [48]. This addresses the importance of cell motility and morphology analysis for a better understanding of cancer cell behaviour. The traditional separate cell motility and morphology analysis is restrictive in the insight that can be obtained if cell behaviours influence cell motility and morphology simultaneously. In this thesis, an integrated computational framework is proposed to describe cell motility and morphology simultaneously as well as the analysis carried out in parallel. The framework is built up on a cell-centric view, which is able to characterise cell dynamics in respect of single cell tracks and can be aggregated to the population-level analysis. The main conclusions of the research investigations are provided below along with some directions for future work.

### 6.1 Summary and conclusions

The computational framework proposed in the thesis consists of modules for an automated multiple cell boundaries tracking, cell motility and morphology characterisation at the same time, integrated study of cell motility and morphology and cell morphodynamics analysis. Its utility was demonstrated by applying it to high throughput screening images of time lapse microscopy data of breast cancer cells in different drug environments.

The problem of cell motility and morphology analysis was viewed as a signal processing problem involving signal tracking, state estimation and classification. Hence, the thesis first reviewed the discrete-time kinematic model structures, static

and state estimation methods that are relevant to automated cell tracking. The tracking process, aimed at reproducing cell spatial and temporal dynamics, either addressed the problem of high-throughput tracking regardless of cell morphology features or focused on cell morphology modelling. The separate cell motility and morphology processing restricts the study of the cell behaviours relevant to cell motility and morphology, the obstacle which this thesis aimed to overcome.

In Chapter 3, an automated spatio-temporal tracking of cell contours, which inherently link cell morphology and motility, was presented. Firstly, the marker-controlled watershed algorithm, which is qualified to separate closely located objects, was applied to cell boundary segmentation of the time-lapse grey-scale image sequences. Secondly, the B-spline active contours with cubic basis functions were constructed for the cell boundary representation. The linear-in-parameter formulation and the off-line calculated basis functions of B-splines translated the problem of cell boundary tracking to a reduced dimensional linear state estimation problem. The tracking was then performed by estimating and smoothing the control points of cell contours through a DWNA kinematic model. The KF and fixed interval RTS smoother were respectively applied to the time forward estimation and backward smoothing. The cell correspondence problem was, meanwhile, solved by the NN algorithm.

The above tracking system was demonstrated by applying it to the time-lapse grey-scale image sequences of wild MDA-MB-231 breast cancer cells. This module was connected to the quantitative cell motility analysis module, which produced quantitative analysis outputs such as velocity, migration direction change, confinement ratio, orientation and distribution of cell protrusion in respect to orientation. The quantitative results showed that cell migration directions, generally, changed gradually over 5 minutes without any sudden changes. The distribution of cell protrusion showed that more protrusions were located at the migration front. While these observations are generally known, these analyses provide the first quantified evidence of such behaviour.

In Chapter 4, a module of morphology analysis was built as an extension of the tracking system proposed in Chapter 3. Initially, the FD combined with PCA were applied to characterise the cell contour features in a 2-D shape space. According to the feature distributions and corresponding physical explanations of FD, cell shapes were assigned to teardrop, elongated and round shape classes by a model of Gaussian mixture distributions. These cell clusters were analysed, and integrated with cell motility features to explore cell motility patterns for the different morphological classes. The cell morphodynamics were then characterised by an HMM initialised with the Gaussian mixture distribution. The estimated transition

matrix gives the transition probabilities between different shape states at adjacent moments and additionally, the state sojourn time, which is the expected time of a cell retaining the same shape class.

The above techniques were applied to both single cell and population-level analysis of cell migration. The simultaneous analysis of cell motility and morphology of single cell migration indicated that the cell migration direction change was lagging behind the shape change, as well as velocity changes. The quantitative integrated analysis showed that cells of different shapes presented different motility features. Round cells were characterised by the slowest migration speed but with a wide range of directional changes. The teardrop cells, on the other hand, migrated quickly and with direction changes that had a relatively narrow distribution. Elongated cells showed the strongest ability to either keep its original orientation or migrate backwards. In terms of the analysis of morphodynamics, the sojourn time decrease was found to be ordered as round, elongated and teardrop states. The teardrop, as the least stable state, showed a significant probability of transition to the round shape state which was the most stable state.

In Chapter 5, the framework integrated the cell tracking and analysis, which was proposed in Chapters 3 and 4, was applied to compare the effects of chemical inhibitors, blebbistatin and Y-27632, on breast cancer cell migration. The quantitative analysis showed that blebbistatin and Y-27632 had similar but distinct effects on breast cancer cell migration. Both chemical inhibitor treatments increased a cell's ability to forward migrate and the distribution of its protrusion at the migration front. Blebbistatin treated cells showed faster migration speeds while Y-27632 treated cells showed more restrictions on migration directional changes. In terms of morphodynamics, both treatments increased the probability of being in the teardrop shape state, but decreased the stability of being in the round shape state. Y-27632 showed a stronger effect on the sojourn time of being in the round shape state which is about 2 times when treated by blebbistatin. These results are based on a small population of cells and so would require analysis over a larger number of cells before a firm conclusion can be drawn.

Subsequently, the integrated framework was applied to analyse the effects of knocking down the FA associated genes. The results showed that the depletion of FAK, RhoE and  $\beta$ -PIX affect cell behaviours differently. FAK knockdown significantly restricted cell motility of both the centroid and the cytoskeleton. RhoE knockdown cells showed slower migration, but stronger capacity for directional persistence. The  $\beta$ -PIX knockdown cells showed similar motility pattern as control cells except that there was an increase of protrusion distribution across the whole cytoskeleton. The cell morphodynamics of FA associated gene knockdown cells,

when compared to control cells, showed that, (i) the depletion of FAK especially influences the cells to predominantly be in the round shape state; (ii) the depletion of RhoE increased the stability of cells to be in the elongated shape state; and (iii) the  $\beta$ -PIX knockdown increased the probability of a cell being in the teardrop shape state. These findings support the assertion that FAK is a potential therapeutic target for breast cancer treatment.

In conclusion, the proposed computational framework permits the integrated analysis of cell motility and morphology through an automated cell boundary tracking system. Its application to high throughput time lapse microscopic image sequences showed its ability to quantify and compare cell behaviours through the automatic extraction of motility, morphology and morphodynamic features. This indicates its capacity for giving an insight into cell-associated hypotheses associated with unknown proteins, chemical treatments and molecular mechanisms.

## 6.2 Future work

The migration of breast cancer cells is accompanied with simultaneous changes of motility and morphology. The proposed novel framework performs the analysis of cell motility and morphology in parallel, combined with quantitative analysis of cell morphodynamics. The thesis has laid a platform for such integrated analysis but further extensions and new directions of research can take this further, suggestions of which are outlined below.

- The tracking results, given in Chapter 3, have shown that the tracking system is adapted for the scenario where cells are moving slowly, without cross migration and cell divisions. To handle this extended scenario, alternative statistical data association methods can be applied in place of the NN algorithm. For example, the Joint Probability Data Association Filter (JPDAF) [135] and Multiple Hypothesis Tracking (MHT) [18] are two commonly applied and robust data association methods to handle the multiple target tracking problem with missing data, object occlusion and so on. This upgrade for the tracking system will provide a better analysis of cell lineages, as well as increase the analysis accuracy of single cell behaviours.
- The cell shapes classification, presented in Chapter 4, is performed in the 2-D shape space extracted from cell boundary FD that are insensitive to shape shift, rotation and scale. A more precise classification for the continuously changed cell shape [174] may be performed by involving more morphology

coefficients that are sensitive to the cellular environment changes, such as cell areas and cell image intensity.

- In addition, analysing cell motility in respect of the shape states obtained from HMM can be used as prior knowledge to construct hybrid model structures (IMM) so that cell behaviours are estimated simultaneously with cell shape state estimation.
- Currently, the applications, presented in Chapter 5, focus on the analysis of breast cancer cell migration. Other cell types, such as leukocytes in the immune system, are also analysed for their motility and morphology. Similar time lapse image sequences of cells are generated under different drug treatments and so their analysis can also be carried out using the developed integrated analysis framework.
- When integrated with different level cell-associated data, this framework can be further applied to study the multi-scale aspects of living systems characteristics, such as the molecule-cell (e.g., protein-cell) and the cell-tissue systems [147].

In summary, these future work are built on the integrated cell motility and morphology analysis framework that is proposed in this thesis and shows its potential for adoption beyond the analysis of breast cancer cells.

# Bibliography

- [1] J.J. Agresti, E. Antipov, A.R. Abate, K. Ahn, A.C. Rowat, J.C. Baret, M. Marquez, A.M. Klibanov, A.D. Griffiths, and D.A. Weitz. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences*, 107(9):4004–4009, 2010.
- [2] M. Al-Bayati and A. El-Zaart. Mammogram Images Thresholding for Breast Cancer Detection Using Different Thresholding Methods. *Advances in Breast Cancer Research*, 02(03):72–77, 2013.
- [3] J. Aldrich. Doing Least Squares: Perspectives from Gauss and Yule. *International Statistical Review*, 66(1):61–81, April 1998.
- [4] J.S. Allingham, R. Smith, and I. Rayment. The structural basis of blebbistatin inhibition and specificity for myosin II. *Nature Structural & Molecular Biology*, 12(4):378–379, 2005.
- [5] K.W. Anna Haeger. Collective cell migration: Guidance principles and hierarchies. *Trends in Cell Biology*, 25(9), 2015.
- [6] C. Bakal, J. Aach, G. Church, and N. Perrimon. Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology. *Science*, 316(5832):1753–1756, June 2007.
- [7] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. Wiley: New York, April 2004.
- [8] N. Barker and C. Iacobuzio-Donahue. The Hidden Beauty in Biomedical Imaging. *Journal of Visual Communication in Medicine*, 38(3-4):220–227, October 2015.
- [9] J. Bektic, K. Pfeil, A.P. Berger, R. Ramoner, A. Pelzer, G. Schäfer, K. Kofler, G. Bartsch, and H. Klocker. Small G-protein RhoE is underexpressed in



- prostate cancer and induces cell cycle arrest and apoptosis. *The Prostate*, 64 (4):332–340, September 2005.
- [10] J.B. Beltman, A.F. Marée, and R.J. de Boer. Analysing immune cell migration. *Nature Reviews Immunology*, 9(11):789–798, 2009.
- [11] K.A. Beningo, K. Hamao, M. Dembo, Y.I. Wang, and H. Hosoya. Traction Forces of Fibroblasts are Regulated by the Rho-Dependent Kinase but not by the Myosin Light Chain Kinase. *Archives of Biochemistry and Biophysics*, 456 (2):224–231, December 2006.
- [12] M. Bertalmio, G. Sapiro, and G. Randall. Morphing active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):733–737, July 2000.
- [13] O. Bertrand, R. Queval, and H. Maître. Shape interpolation using fourier descriptors with application to animation graphics. *Signal Processing*, 4(1): 53–58, 1982.
- [14] S. Beucher and C. Lantuejoul. Use of Watersheds in Contour Detection. In *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation, Rennes, France.*, September 1979.
- [15] S. Beucher and C. Lantuejoul. Use of watersheds in contour detection. In *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation, Rennes, France.*, September 1979.
- [16] J. Bilmes. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical report, International Computer Science Institute, 1998.
- [17] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, August 2006.
- [18] S.S. Blackman. Multiple hypothesis tracking for multiple target tracking. *Aerospace and Electronic Systems Magazine, IEEE*, 19(1):5–18, 2004.
- [19] A. Blake and M. Isard. *Active Contours*. Springer London, London, 1998.
- [20] H. Blom. An efficient filter for abruptly changing systems. In *The 23rd IEEE Conference on Decision and Control, 1984*, pages 656–658, December 1984.
- [21] M.V. Boland and R.F. Murphy. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics*, 17(12):1213–1223, 2001.

- [22] J.E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965.
- [23] R.G. Brown and P.Y.C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering with MATLAB Exercises, 4th Edition*. Wiley Global Education, January 2012.
- [24] C. Solomon and T. Breckon. Wiley: Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab - Chris Solomon, Toby Breckon, December 2010. URL <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470844728.html>. Accessed: 2-16-05-11.
- [25] B. Cabieses, H. Tunstall, and K. Pickett. Understanding the Socioeconomic Status of International Immigrants in Chile Through Hierarchical Cluster Analysis: a Population-Based Study. *International Migration*, 53(2):303–320, 2015.
- [26] C. Cadart, E. Zlotek-Zlotkiewicz, M. Le Berre, M. Piel, and H.K. Matthews. Exploring the Function of Cell Shape and Size during Mitosis. *Developmental Cell*, 29(2):159–169, April 2014.
- [27] Cancer Research UK. Breast cancer statistics | Cancer Research UK, 2015. URL <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer>. Accessed: 2016-04-12.
- [28] J. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, November 1986.
- [29] S.P. Carey, A. Rahman, C.M. Kraning-Rush, B. Romero, S. Somasegar, O.M. Torre, R.M. Williams, and C.A. Reinhart-King. Comparative mechanisms of cancer cell migration through 3d matrix and physiological microtracks. *American Journal of Physiology - Cell Physiology*, 308(6):C436–C447, March 2015.
- [30] A.F. Chambers, A.C. Groom, and I.C. MacDonald. Metastasis: dissemination and growth of cancer cells in metastatic sites. *Nature Reviews Cancer*, 2(8):563–572, 2002.
- [31] K. Chang. Nobel Laureates Pushed Limits of Microscopes. *The New York Times*, October 2014.

- [32] V. Cherkassky and F.M. Mulier. *Learning from data: concepts, theory, and methods*. John Wiley & Sons, 2007.
- [33] C.K. Choi, M. Vicente-Manzanares, J. Zareno, L.A. Whitmore, A. Mogilner, and A.R. Horwitz. Actin and  $\alpha$ -actinin orchestrate the assembly and maturation of nascent adhesions in a myosin II motor-independent manner. *Nature Cell Biology*, 10(9):1039–1050, 2008.
- [34] J.W. Choi, T.K. Whangbo, and C.G. Kim. A contour tracking method of large motion object using optical flow and active contour model. *Multimedia Tools and Applications*, 74(1):199–210, November 2013.
- [35] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5):564–577, 2003.
- [36] J.S. Condeelis, J.B. Wyckoff, M. Bailly, R. Pestell, D. Lawrence, J. Backer, and J.E. Segall. Lamellipodia in invasion. In *Seminars in Cancer Biology*, volume 11, pages 119–128, 2001.
- [37] D.L. Coutu and T. Schroeder. Probing cellular processes by long-term live imaging – historic problems and current solutions. *Journal of Cell Science*, 126(17):3805–3815, September 2013.
- [38] I.J. Cox. A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision*, 10(1):53–66, 1993.
- [39] A.P. Dempster. Upper and Lower Probabilities Induced by a Multivalued Mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, 1967.
- [40] F.J. Doyle and J. Stelling. Systems interface biology. *Journal of The Royal Society Interface*, 3(10):603–616, October 2006.
- [41] M.K. Driscoll, J.T. Fourkas, and W. Losert. Local and global measures of shape dynamics. *Physical Biology*, 8(5):055001, 2011.
- [42] M.S. Duxbury, S.W. Ashley, and E.E. Whang. Inhibition of pancreatic adenocarcinoma cellular invasiveness by blebbistatin: a novel myosin II inhibitor. *Biochemical and Biophysical Research Communications*, 313(4):992–997, 2004.
- [43] G.A. Einicke, editor. *Smoothing, Filtering and Prediction - Estimating The Past, Present and Future*. InTech, February 2012.
- [44] I. Färber, S. Günemann, H.P. Kriegel, P. Kröger, E. Müller, E. Schubert, T. Seidl, and A. Zimek. On using class-labels in evaluation of clusterings.

- In *MultiClust: 1st international workshop on discovering, summarizing and using multiple clusterings held in conjunction with KDD*, page 1, 2010.
- [45] J. Fisher and T.A. Henzinger. Executable cell biology. *Nature Biotechnology*, 25(11):1239–1249, November 2007.
- [46] C. Frantz, K.M. Stewart, and V.M. Weaver. The extracellular matrix at a glance. *Journal of Cell Science*, 123(24):4195–4200, 2010.
- [47] P. Friedl and D. Gilmour. Collective cell migration in morphogenesis, regeneration and cancer. *Nature Reviews Molecular Cell Biology*, 10(7):445–457, 2009.
- [48] P. Friedl and K. Wolf. Tumour-cell invasion and migration: diversity and escape mechanisms. *Nature Reviews Cancer*, 3(5):362–374, May 2003.
- [49] V. Gabarra-Niecko, M.D. Schaller, and J.M. Dunty. FAK regulates biological processes important for the pathogenesis of cancer. *Cancer and Metastasis Reviews*, 22(4):359–374, 2003.
- [50] M.L. Gardel, B. Sabass, L. Ji, G. Danuser, U.S. Schwarz, and C.M. Waterman. Traction stress in focal adhesions correlates biphasically with actin retrograde flow speed. *The Journal of Cell Biology*, 183(6):999–1005, December 2008.
- [51] U. Genschel and W.Q. Meeker. A Comparison of Maximum Likelihood and Median-Rank Regression for Weibull Estimation. *Quality Engineering*, 22(4): 236–255, October 2010.
- [52] Z. Ghahramani and M.I. Jordan. Factorial Hidden Markov Models. *Machine Learning*, 29(2-3):245–273, November 1997.
- [53] J.M. Grange, J.L. Stanford, and C.A. Stanford. Campbell De Morgan’s ‘Observations on cancer’, and their relevance today. *Journal of the Royal Society of Medicine*, 95(6):296–299, January 2002.
- [54] R.A. Guler, S. Tari, and G. Unal. Landmarks inside the shape: Shape matching using image descriptors. *Pattern Recognition*, 49:79–88, January 2016.
- [55] S.I. Hajdu. A note from history: Landmarks in history of cancer, part 1. *Cancer*, 117(5):1097–1102, March 2011.
- [56] P. Hansen and B. Jaumard. Cluster analysis and mathematical programming. *Mathematical Programming*, 79(1-3):191–215, 1997.

- [57] R.P. Hertzberg and A.J. Pope. High-throughput screening: new technology for the 21st century. *Current Opinion in Chemical Biology*, 4(4):445–451, 2000.
- [58] R.M. Hoffman. Green fluorescent protein imaging of tumour growth, metastasis, and angiogenesis in mouse models. *The Lancet Oncology*, 3(9):546–556, 2002.
- [59] P. Hogeweg. The Roots of Bioinformatics in Theoretical Biology. *PLoS Computational Biology*, 7(3), March 2011.
- [60] Y. Hu. Iterative and recursive least squares estimation algorithms for moving average systems. *Simulation Modelling Practice and Theory*, 34:12–19, May 2013.
- [61] S. Hugues, L. Fetler, L. Bonifaz, J. Helft, F. Amblard, and S. Amigorena. Distinct T cell dynamics in lymph nodes during the induction of tolerance and immunity. *Nature Immunology*, 5(12):1235–1242, December 2004.
- [62] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the Hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850–863, 1993.
- [63] F. Imamura, M. Mukai, M. Ayaki, and others. Y-27632, an Inhibitor of Rho-associated Protein Kinase, Suppresses Tumor Cell Invasion via Regulation of Focal Adhesion and Focal Adhesion Kinase. *Japanese Journal of Cancer Research*, 91(8):811–816, 2000.
- [64] A.K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.
- [65] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S.L. Schmid, and G. Danuser. Robust single-particle tracking in live-cell time-lapse sequences. *Nature Methods*, 5(8):695–702, August 2008.
- [66] K.K. Jena. Application of a Nobel Approach for Edge Detection of Images. *Journal of Image Processing and Artificial Intelligence*, 1(1), 2015.
- [67] Z. Ji, Y. Xia, Q. Sun, G. Cao, and Q. Chen. Active contours driven by local likelihood image fitting energy for image segmentation. *Information Sciences*, 301:285–304, April 2015.
- [68] K. Jo, K. Chu, and M. Sunwoo. Interacting Multiple Model Filter-Based Sensor Fusion of GPS With In-Vehicle Sensors for Real-Time Vehicle Positioning.

- IEEE Transactions on Intelligent Transportation Systems*, 13(1):329–343, March 2012.
- [69] K.A. Joshi and D.G. Thakore. A survey on moving object detection and tracking in video surveillance system. *International Journal on Information Theory*, pages 2231–2307, 2012.
- [70] S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477–482, March 2000.
- [71] S.J. Julier and J.K. Uhlmann. New extension of the Kalman filter to nonlinear systems. In *AeroSense'97*, pages 182–193. 1997.
- [72] S. Julier and J. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, March 2004.
- [73] V. Kadirkamanathan, S.R. Anderson, S.A. Billings, X. Zhang, G.R. Holmes, C.C. Reyes-Aldasoro, P.M. Elks, and S.A. Renshaw. The Neutrophil's Eye-View: Inference and Visualisation of the Chemoattractant Field Driving Cell Chemotaxis In Vivo. *PloS one*, 7(4):e35182, 2012.
- [74] K. Kameshwaran and K. Malarvizhi. Survey on Clustering Techniques in Data Mining. *International Journal of Computer Science and Information Technologies*, 5(2):2272–2276, 2014.
- [75] N. Kantas, A. Doucet, S.S. Singh, J. Maciejowski, and N. Chopin. On Particle Methods for Parameter Estimation in State-Space Models. *Statistical Science*, 30(3):328–351, August 2015.
- [76] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, January 1988.
- [77] K. Keren, Z. Pincus, G.M. Allen, E.L. Barnhart, G. Marriott, A. Mogilner, and J.A. Theriot. Mechanism of shape determination in motile cells. *Nature*, 453(7194):475–480, May 2008.
- [78] K. Kim, O. Ossipova, and S.Y. Sokol. Neural Crest Specification by Inhibition of the ROCK/Myosin II Pathway. *STEM CELLS*, 33(3):674–685, March 2015.
- [79] K. Kimura, S. Teranishi, T. Orita, H. Zhou, and T. Nishida. Role of  $\beta$ -Pix in Corneal Epithelial Cell Migration on Fibronectin. *Investigative Ophthalmology & Visual Science*, 52(6):3181–3186, 2011.

- [80] G. Kitagawa. A Self-Organizing State-Space Model. *Journal of the American Statistical Association*, 93(443):1203–1215, 1998.
- [81] T. Kucukyilmaz. Parallel K-Means Algorithm for Shared Memory Multiprocessors. *Journal of Computer and Communications*, 02(11):15–23, 2014.
- [82] J.C. Kuo, X. Han, C.T. Hsiao, J.R. Yates III, and C.M. Waterman. Analysis of the myosin-II-responsive focal adhesion proteome reveals a role for  $\beta$ -Pix in negative regulation of focal adhesion maturation. *Nature Cell Biology*, 13(4):383–393, 2011.
- [83] S. Lamouille, J. Xu, and R. Derynck. Molecular mechanisms of epithelial–mesenchymal transition. *Nature Reviews. Molecular Cell Biology*, 15(3):178–196, March 2014.
- [84] E.Y. Lasfargues and L. Ozzello. Cultivation of human breast carcinomas. *Journal of the National Cancer Institute*, 21(6):1131–1147, December 1958.
- [85] D.A. Lauffenburger and A.F. Horwitz. Cell Migration: A Physically Integrated Molecular Process. *Cell*, 84(3):359–369, February 1996.
- [86] M. Le Borgne, E. Ladi, I. Dzhagalov, P. Herzmark, Y.F. Liao, A.K. Chakraborty, and E.A. Robey. The impact of negative selection on thymocyte migration in the medulla. *Nature Immunology*, 10(8):823–830, 2009.
- [87] S.J. Lee, Y. Motai, and H. Choi. Tracking human motion with multichannel interacting multiple model. *Industrial Informatics, IEEE Transactions on*, 9(3):1751–1763, 2013.
- [88] O.V. Leontieva, V. Natarajan, Z.N. Demidenko, L.G. Burdelya, A.V. Gudkov, and M.V. Blagosklonny. Hypoxia suppresses conversion from proliferative arrest to cellular senescence. *Proceedings of the National Academy of Sciences*, 109(33):13314–13318, August 2012.
- [89] C. Li, C. Xu, C. Gui, and M.D. Fox. Level set evolution without re-initialization: a new variational formulation. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 430–436. 2005.
- [90] C. Li, C. Xu, C. Gui, and M.D. Fox. Distance regularized level set evolution and its application to image segmentation. *Image Processing, IEEE Transactions on*, 19(12):3243–3254, 2010.

- [91] H. Li, B. Manjunath, and S. Mitra. A contour-based approach to multisensor image registration. *IEEE Transactions on Image Processing*, 4(3):320–334, March 1995.
- [92] K. Li, E.D. Miller, M. Chen, T. Kanade, L.E. Weiss, and P.G. Campbell. Cell population tracking and lineage construction with spatiotemporal context. *Medical Image Analysis*, 12(5):546–566, 2008.
- [93] W. Li, L. Fu, B. Niu, S. Wu, and J. Wooley. Ultrafast clustering algorithms for metagenomic sequence analysis. *Briefings in Bioinformatics*, 13(6):656–668, January 2012.
- [94] G. Liasis and S. Stavrou. Building extraction in satellite images using active contours and colour features. *International Journal of Remote Sensing*, 37(5):1127–1153, March 2016.
- [95] C. Liu, K. Zhang, H. Xiong, G. Jiang, and Q. Yang. Temporal skeletonization on sequential data: patterns, categorization, and visualization. *Knowledge and Data Engineering, IEEE Transactions on*, 28(1):211–223, 2016.
- [96] Y.J. Liu, M. Le Berre, F. Lautenschlaeger, P. Maiuri, A. Callan-Jones, M. Heuzé, T. Takaki, R. Voituriez, and M. Piel. Confinement and Low Adhesion Induce Fast Amoeboid Migration of Slow Mesenchymal Cells. *Cell*, 160(4):659–672, February 2015.
- [97] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *“Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics”*, pages 281–297. Oakland, CA, USA., 1967.
- [98] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan. Interacting multiple model methods in target tracking: a survey. *IEEE Transactions on Aerospace and Electronic Systems*, 34(1):103–123, January 1998.
- [99] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.
- [100] E. Meijering, O. Dzyubachyk, and I. Smal. Methods for cell and particle tracking. *Methods Enzymol*, 504(9):183–200, 2012.
- [101] T.P. Minka. From Hidden Markov Models to Linear Dynamical Systems. Technical report, Tech. Rep. 531, Vision and Modeling Group of Media Lab, MIT, 1999.



- [102] P. Moallem, H. Tahvilian, and S.A. Monadjemi. Parametric active contour model using Gabor balloon energy for texture segmentation. *Signal, Image and Video Processing*, 10(2):351–358, January 2015.
- [103] T. Moeslund. Canny Edge Detection. *Laboratory of Computer Vision and Media Technology*, 2009.
- [104] M.R. Mofrad and R.D. Kamm. *Cytoskeletal mechanics: models and measurements in cell mechanics*. Cambridge University Press, 2006.
- [105] T.K. Moon. The expectation-maximization algorithm. *Signal Processing Magazine, IEEE*, 13(6):47–60, 1996.
- [106] B.S. Morse. Lecture 18: Segmentation (Region Based). *Brigham Young University*, 2000, 1998.
- [107] B.S. Morse. Lecture 4: Thresholding. *Brigham Young University*, 2000.
- [108] T. Morzy, M. Wojciechowski, and M. Zakrzewicz. Pattern-oriented hierarchical clustering. In *Advances in Databases and Information Systems*, pages 179–190. 1999.
- [109] National Cancer Institute. What Is Cancer | National Cancer Institute, 2016. URL <http://www.cancer.gov/about-cancer/what-is-cancer>. Access: 2016-04-11.
- [110] N. Ofer, A. Mogilner, and K. Keren. Actin disassembly clock determines shape and speed of lamellipodial fragments. *Proceedings of the National Academy of Sciences*, 108(51):20394–20399, 2011.
- [111] M. Ohgushi, M. Matsumura, M. Eiraku, K. Murakami, T. Aramaki, A. Nishiyama, K. Muguruma, T. Nakano, H. Suga, M. Ueno, T. Ishizaki, H. Suemori, S. Narumiya, H. Niwa, and Y. Sasai. Molecular Pathway and Cell State Responsible for Dissociation-Induced Apoptosis in Human Pluripotent Stem Cells. *Cell Stem Cell*, 7(2):225–239, August 2010.
- [112] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.
- [113] I. Pal, A.W. Robertson, U. Lall, and M.A. Cane. Modeling winter rainfall in Northwest India using a hidden Markov model: understanding occurrence of different states and their dynamical connections. *Climate Dynamics*, 44(3-4):1003–1015, May 2014.

- [114] F. Papenmeier, H.S. Meyerhoff, G. Jahn, and M. Huff. Tracking by location and features: Object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1):159–171, 2014.
- [115] K.K. Parker, A.L. Brock, C. Brangwynne, R.J. Mannix, N. Wang, E. Ostuni, N.A. Geisse, J.C. Adams, G.M. Whitesides, and D.E. Ingber. Directional control of lamellipodia extension by constraining cell shape and orienting cell tractional forces. *The FASEB Journal*, 16(10):1195–1204, 2002.
- [116] J.T. Parsons, A.R. Horwitz, and M.A. Schwartz. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nature Reviews Molecular Cell Biology*, 11(9):633–643, 2010.
- [117] E. Persoon and K.S. Fu. Shape discrimination using fourier descriptors. *IEEE transactions on pattern analysis and machine intelligence*, (3):388–397, 1986.
- [118] J. Pfanzagl. *Parametric Statistical Theory*. Walter de Gruyter, January 1994.
- [119] D.L. Pham, C. Xu, and J.L. Prince. Current methods in medical image segmentation 1. *Annual Review of Biomedical Engineering*, 2(1):315–337, 2000.
- [120] J. Piltti, M. Varjosalo, C. Qu, J. Häyrynen, and M.J. Lammi. Rho-kinase inhibitor Y-27632 increases cellular proliferation and migration in human foreskin fibroblast cells. *PROTEOMICS*, 15(17):2953–2965, September 2015.
- [121] Z. Pincus and J.A. Theriot. Comparison of quantitative methods for cell-shape analysis. *Journal of Microscopy*, 227(2):140–156, August 2007.
- [122] S. Pinner and E. Sahai. PDK1 regulates cancer cell motility by antagonising inhibition of ROCK1 by RhoE. *Nature Cell Biology*, 10(2):127–137, 2008.
- [123] Rahman Farnoosh and Arezoo Hajrajabi. Estimation of parameters in the state space model of stochastic RL electrical circuit. *COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 32(3):1082–1097, May 2013.
- [124] H.E. Rauch, C.T. Striebel, and F. Tung. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450, 1965.
- [125] S.I. Rennard, N. Locantore, B. Delafont, R. Tal-Singer, E.K. Silverman, J. Vestbo, B.E. Miller, P. Bakke, B. Celli, P.M.A. Calverley, H. Coxson, C. Crim, L.D. Edwards, D.A. Lomas, W. MacNee, E.F.M. Wouters, J.C. Yates, I. Coca,

- and A. Agustí. Identification of Five Chronic Obstructive Pulmonary Disease Subgroups with Different Prognoses in the ECLIPSE Cohort Using Cluster Analysis. *Annals of the American Thoracic Society*, 12(3):303–312, February 2015.
- [126] C.C. Reyes-Aldasoro, S. Akerman, and G.M. Tozer. Measuring the velocity of fluorescently labelled red blood cells with a keyhole tracking algorithm. *Journal of Microscopy*, 229(1):162–173, 2008.
- [127] A.J. Ridley, M.A. Schwartz, K. Burridge, R.A. Firtel, M.H. Ginsberg, G. Borisy, J.T. Parsons, and A.R. Horwitz. Cell migration: integrating signals from front to back. *Science*, 302(5651):1704–1709, 2003.
- [128] K. Riento and A.J. Ridley. Rocks: multifunctional kinases in cell behaviour. *Nature Reviews Molecular Cell Biology*, 4(6):446–456, 2003.
- [129] K. Riento, R.M. Guasch, R. Garg, B. Jin, and A.J. Ridley. RhoE binds to ROCK I and inhibits downstream signaling. *Molecular and Cellular Biology*, 23(12):4219–4229, 2003.
- [130] X. Rong Li and Y. Bar-Shalom. Tracking in clutter with nearest neighbor filters: analysis and performance. *Aerospace and Electronic Systems, IEEE Transactions on*, 32(3):995–1010, 1996.
- [131] Saed Sayad. An introduction to data mining, 2011. URL [http://www.saedsayad.com/clustering\\_hierarchical.htm](http://www.saedsayad.com/clustering_hierarchical.htm). Access: 2016-03-31.
- [132] H. Sailem, V. Bousgouni, S. Cooper, and C. Bakal. Cross-talk between Rho and Rac GTPases drives deterministic exploration of cellular shape space and morphological heterogeneity. *Open Biology*, 4(1):130132, January 2014.
- [133] S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, September 2013.
- [134] N. Schneider and D.M. Gavrila. Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study. In J. Weickert, M. Hein, and B. Schiele, editors, *Pattern Recognition*, number 8142 in Lecture Notes in Computer Science, pages 174–183. Springer Berlin Heidelberg, September 2013.
- [135] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In *IEEE International Conference on Robotics and Automation, 2001. Proceedings 2001 ICRA*, volume 2, pages 1665–1670 vol.2, 2001.

- [136] J.L. Semmlow. *Biosignal and biomedical image processing: MATLAB-based applications*, volume 22. Marcel Dekker Inc, Madison Avenue, New York, 2004.
- [137] J.E. Sero, C.K. Thodeti, A. Mammoto, C. Bakal, S. Thomas, and D.E. Ingber. Paxillin mediates sensing of physical cues and regulates directional cell motility by controlling lamellipodia positioning. *PloS ONE*, 6(12):e28303, 2011.
- [138] J.E. Sero, H.Z. Sailem, R.C. Ardy, H. Almuttaqi, T. Zhang, and C. Bakal. Cell shape and the microenvironment regulate nuclear translocation of NF- $\kappa$ B in breast epithelial and tumor cells. *Molecular Systems Biology*, 11(3):790, 2015.
- [139] J.V. Small, T. Stradal, E. Vignal, and K. Rottner. The lamellipodium: where motility begins. *Trends in Cell Biology*, 12(3):112–120, 2002.
- [140] B. Snijder, R. Sacher, P. Rämö, E.M. Damm, P. Liberali, and L. Pelkmans. Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*, 461(7263):520–523, September 2009.
- [141] C. Solomon and T. Breckon. *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons, 2011.
- [142] A.V. Somlyo, D. Bradshaw, S. Ramos, C. Murphy, C.E. Myers, and A.P. Somlyo. Rho-Kinase Inhibitor Retards Migration and in Vivo Dissemination of Human Prostate Cancer Cells. *Biochemical and Biophysical Research Communications*, 269(3):652–659, 2000.
- [143] L. Song, E.J. Hennink, I.T. Young, and H.J. Tanke. Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. *Biophysical Journal*, 68(6):2588, 1995.
- [144] D. Spasova, D. Spasov, A. Stoilova, B. Atanasova, L. Mihajlov, and N. Valkova. Application of cluster analysis for evaluation of new Bulgarian and Macedonian cotton varieties and lines. *Bulgarian Journal of Agricultural Science*, 22(1):125–130, 2016.
- [145] J. Steinier, Y. Termonia, and J. Deltour. Smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry*, 44(11):1906–1909, September 1972.
- [146] B. Stewart, C.P. Wild, and others. World cancer report 2014. *World*, 2015.
- [147] M.A. Stolarska, Y. Kim, and H.G. Othmer. Multi-scale models of cell and tissue dynamics. *Philosophical Transactions of the Royal Society of London A*:

- Mathematical, Physical and Engineering Sciences*, 367(1902):3525–3553, September 2009.
- [148] T. Stradal, K.D. Courtney, K. Rottner, P. Hahne, J.V. Small, and A.M. Pendergast. The Abl interactor proteins localize to sites of actin polymerization at the tips of lamellipodia and filopodia. *Current Biology*, 11(11):891–895, 2001.
- [149] Z. Szallasi, J. Stelling, and V. Periwal. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. MIT Press, 2010.
- [150] L. Taliaferro-Smith, E. Oberlick, T. Liu, T. McGlothen, T. Alcaide, R. Tobin, S. Donnelly, R. Commander, E. Kline, G.P. Nagaraju, L. Havel, A. Marcus, R. Nahta, and R. O'Regan. FAK activation is required for IGF1r-mediated regulation of EMT, migration, and invasion in mesenchymal triple negative breast cancer cells. *Oncotarget*, 6(7):4757–4772, February 2015.
- [151] J.P. Thiery. Epithelial–mesenchymal transitions in tumour progression. *Nature Reviews Cancer*, 2(6):442–454, 2002.
- [152] E.J. van der Gaag, M.T. Leccia, S.K. Dekker, N.L. Jalbert, D.M. Amodeo, and H.R. Byers. Role of zyxin in differential cell spreading and proliferation of melanoma cells and melanocytes. *Journal of Investigative Dermatology*, 118(2):246–254, 2002.
- [153] F.M. Vega and A.J. Ridley. Rho GTPases in cancer cell biology. *FEBS Letters*, 582(14):2093–2101, June 2008.
- [154] O. Virtajoki, P. Fra, and others. Fast pairwise nearest neighbor based algorithm for multilevel thresholding. *Journal of Electronic Imaging*, 12(4):648–659, 2003.
- [155] Y. Vlasov. Silicon photonics for next generation computing systems. In *European Conference on Optical Communications*, pages 1–2, Brussels, Sept. 2008.
- [156] C. Wählby, J. Lindblad, M. Vondrus, E. Bengtsson, and L. Björkesten. Algorithms for cytoplasm segmentation of fluorescence labelled cells. *Analytical Cellular Pathology*, 24(2):101–111, 2002.
- [157] B. Wang and S. Fan. An improved CANNY edge detection algorithm. In *2009 second international workshop on computer science and engineering*, pages 497–500. 2009.

- [158] S.J. Wang, W. Saadi, F. Lin, C.M.C. Nguyen, and N.L. Jeon. Differential effects of EGF gradient profiles on MDA-MB-231 breast cancer cell chemotaxis. *Experimental Cell Research*, 300(1):180–189, 2004.
- [159] K.K. Ward, I. Tancioni, C. Lawson, N.L.G. Miller, C. Jean, X.L. Chen, S. Uryu, J. Kim, D. Tarin, D.G. Stupack, S.C. Plaxe, and D.D. Schlaepfer. Inhibition of focal adhesion kinase (FAK) activity prevents anchorage-independent ovarian carcinoma cell growth and tumor progression. *Clinical & Experimental Metastasis*, 30(5):579–594, December 2012.
- [160] G. Welch and G. Bishop. An Introduction to the Kalman Filter. Technical report, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1995.
- [161] WHO. WHO | The top 10 causes of death, 2015. URL <http://www.who.int/mediacentre/factsheets/fs310/en/>.
- [162] WHO. WHO | Fact sheet on the World Malaria Report 2014, 2016. URL [http://www.who.int/malaria/media/world\\_malaria\\_report\\_2014/en/](http://www.who.int/malaria/media/world_malaria_report_2014/en/). Access: 2016-04-11.
- [163] K. Wolf, I. Mazo, H. Leung, K. Engelke, U.H. von Andrian, E.I. Deryugina, A.Y. Strongin, E.B. Bröcker, and P. Friedl. Compensation mechanism in tumor cell migration mesenchymal–amoeboid transition after blocking of pericellular proteolysis. *The Journal of Cell Biology*, 160(2):267–277, 2003.
- [164] E.C. Woodhouse, R.F. Chuaqui, and L.A. Liotta. General mechanisms of metastasis. *Cancer*, 80(S8):1529–1537, 1997.
- [165] P. Wu, X. Li, J. Kong, and J. Liu. Heterogeneous Multiple Sensors Joint Tracking of Maneuvering Target in Clutter. *Sensors (Basel, Switzerland)*, 15(7):17350–17365, 2015.
- [166] Y. Xiao, Y. Zhang, J. Ding, and J. Dai. The residual based interactive least squares algorithms and simulation studies. *Computers & Mathematics with Applications*, 58(6):1190–1197, September 2009.
- [167] Z. Xin, S. Jiang, P. Jiang, X. Yan, C. Fan, S. Di, G. Wu, Y. Yang, R.J. Reiter, and G. Ji. Melatonin as a treatment for gastrointestinal cancer: a review. *Journal of Pineal Research*, 58(4):375–387, May 2015.
- [168] Y. Xiong and D.Y. Yeung. Mixtures of ARMA models for model-based time series clustering. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 717–720. 2002.

- [169] C. Xu, A. Yezzi Jr, and J.L. Prince. On the relationship between parametric and geometric active contours. In *Signals, Systems and Computers, 2000. Conference Record of the Thirty-Fourth Asilomar Conference on*, volume 1, pages 483–489, 2000.
- [170] R. Xu, D. Wunsch, and others. Survey of clustering algorithms. *Neural Networks, IEEE Transactions on*, 16(3):645–678, 2005.
- [171] D. Yamazaki, S. Kurisu, and T. Takenawa. Regulation of cancer cell motility through actin reorganization. *Cancer Science*, 96(7):379–386, 2005.
- [172] F. Yang, C. Venkataraman, V. Styles, V. Kuttenger, E. Horn, Z.v. Guttenberg, and A. Madzvamuse. A Computational Framework for Particle and Whole Cell Tracking Applied to a Real Biological Dataset. *Journal of Biomechanics*, February 2016.
- [173] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys (CSUR)*, 38(4):13, 2006.
- [174] Z. Yin, A. Sadok, H. Sailem, A. McCarthy, X. Xia, F. Li, M.A. Garcia, L. Evans, A.R. Barr, N. Perrimon, C.J. Marshall, S.T.C. Wong, and C. Bakal. A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes. *Nature Cell Biology*, 15(7):860–871, July 2013.
- [175] B.J. Yoon. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics*, 10(6):402–415, September 2009.
- [176] W. Yu, X. Tian, Z. Hou, Y. Zha, and Y. Yang. Multi-scale mean shift tracking. *IET Computer Vision*, 9(1):110–123, 2015.
- [177] X. Yun and Z.L. Jing. Kernel joint visual tracking and recognition based on structured sparse representation. *Neurocomputing*, 2016.
- [178] R. Zaidel-Bar and B. Geiger. The switchable integrin adhesome. *Journal of Cell Science*, 123(9):1385–1388, 2010.
- [179] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, January 2004.
- [180] M. Zhao, B. Song, J. Pu, T. Wada, B. Reid, G. Tai, F. Wang, A. Guo, P. Walczysko, Y. Gu, and others. Electrical signals control wound healing through phosphatidylinositol-3-OH kinase- $\gamma$  and PTEN. *Nature*, 442(7101):457–460, 2006.

- [181] Q. Zhong, A.G. Busetto, J.P. Fededa, J.M. Buhmann, and D.W. Gerlich. Unsupervised modeling of cell morphology dynamics for time-lapse microscopy. *Nature Methods*, 9(7):711–713, July 2012.
- [182] X. Zhou and S.T. Wong. Informatics challenges of high-throughput microscopy. *Signal Processing Magazine, IEEE*, 23(3):63–72, 2006.
- [183] X. Zhou, K.Y. Liu, P. Bradley, N. Perrimon, and S.T. Wong. Towards Automated Cellular Image Segmentation for RNAi Genome-Wide Screening. In J.S. Duncan and G. Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005: 8th International Conference, Palm Springs, CA, USA, October 26-29, 2005, Proceedings, Part I*, pages 885–892. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [184] C. Zimmer, B. Zhang, A. Dufour, A. Thebaud, S. Berlemont, V. Meas-Yedid, and J.C.O. Marin. On the digital trail of mobile cells. *IEEE Signal Processing Magazine*, 23(3):54–62, May 2006.
- [185] C. Zimmer and J.C. Olivo-Marin. Coupled parametric active contours. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(11):1838–1842, 2005.
- [186] C. Zimmer, E. Labruyere, V. Meas-Yedid, N. Guillen, and J.C. Olivo-Marin. Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: A tool for cell-based drug testing. *Medical Imaging, IEEE Transactions on*, 21(10):1212–1221, 2002.
- [187] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.