Modelling and Identification of Neutrophil Cell Dynamic Behaviour

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

Xiliang Zhang

Department of Automatic Control and Systems Engineering

April 2016

To my family and all my friends

Acknowledgements

I'd like to thank the following people:

- Professor Visakan Kadirkamanathan, Professor Stephen A. Billings, Doctor Sean Anderson, and Professor Stephen A. Renshaw.
- My father, my mother, and other family members.
- My girlfriend, all my classmates and friends.

I appreciate the help of the research staff, especially Professor Visakan Kadirkamanathan. I cannot image what would have happened without the help and encouragement of Professor Visakan Kadirkamanathan. Professor Visakan Kadirkamanathan taught me a lot, including how to cogitate the problem, guided me to the right direction in research, and gave me many suggestions especially when I modified the thesis. I also appreciate the help from Professor Stephen A. Billings. Professor Stephen A. Billings gave me many ideas about my research target and modified my upgrade report and other reports carefully. I am thankful to Doctor Sean Anderson, too. Doctor Sean Anderson helped me a lot in programming and modifying the program. In addition, Doctor Sean Anderson also gave me a large number of ideas about the research target. Finally, I owe my thanks to Professor Stephen A. Renshaw. Professor Stephen A. Renshaw provided a large number of data and explained the medical process and mechanism to me explicitly.

My gratitude goes to my father and my mother. My parents taught me, helped me, and encouraged me. I was so lucky to have such a kind family. Without my parents, I cannot finish my PhD. I also appreciate all the other family members, including my uncles, aunts, brothers, and sisters.

I am also thankful to my friends and classmates. Firstly, I want to appreciate my girlfriend's patience and kindness. She encouraged me to finish the thesis when I was in despair. I also need to appreciate the help from my classmates working together and friends in Sheffield. They are Professor Ziqiang Lang, Dr. Ling Zhao, Dr. Dazhi Jiang, Dr. Yuzhu Guo, Dr. Liang Zhao, Dr. Zuoyi Song, Dr. Yu Zhou, Dr. Pengfei Guo, Dr. Xueyan Zhao, Dr. Fei He, Dr. Qian Lu, Dr. Shu Wang, Dr. Yang Li, Dr. Rui Wang, Ms Yu Liu, Mr. Jia Zhao, Dr. Xiaokai Nie, Dr. Wei Li, Mr. Youchen Wang, Ms Yang Zhang, Ms Zhenni Wang, Ms Shu Jiang, Mr. Minghe Ye, Mr. XinXin Yao, Dr. Andrew Hills, Dr. Parham Aram, Dr. Tara Baldacchino, Dr. Yashar Baradaranshokouhi, Dr. Geoff Holmes, Dr. Kirubhakaran Krishnanathan, Dr. Lingli Shen, and other members of Visakan's group. They helped me a lot in my life in UK.

Finally, I will appreciate all the members mentioned above again for their help in my academic and general life in UK.

Abstract

This thesis is focussed on the shape analysis and tracking of neutrophil cells to facilitate the understanding of their behaviour. Neutrophil, one of the important type of white blood cell, protects humans and animals from infections and inflammations. The underlying mechanism is believed to be that when inflammation happens, a changing chemotaxis field causes neutrophils to move to inflammatory sites. During this migration process, neutrophils also change shapes. For example, pseudopod protrusions are formed on the boundary in response to the local gradient of the chemotactic field. After inflammation resolution, some neutrophils go to "sleep" and some move back but what drives these mechanisms are still not fully understood. If the mechanisms were known, it would be helpful to accelerate or slow down the process of treating some diseases.

The thesis attempts to provide a quantitative analysis based on time lapse microscopic images of *in vivo* zebrafish animal models. The underlying premise governing the analysis is to identify if cell states from their motion and shape.

The thesis begins with cell centroid tracking and uses three common kinematic models commonly used in the target tracking literature. The interacting multiple model framework, which is underpinned by multiple Kalman filters, is used to determine probabilities of most likely model to explain the cell motility pattern. These different models are then compared to identify if the motion pattern (motility) can be attributed to different cell behaviours. This is then followed by cell shape tracking to characterise not only the cell shape but also to identify regions of protrusions. It addresses the problem of estimating the chemotactic field that acts to recruit neutrophil cells to the inflammation sites based only on observed cell tracks, without any direct measurement associated with the external cell environment. By assuming that the cell velocity is proportional to the local chemotactic gradient, a least squares method in combination with the Kalman filter, was used to estimate this field. Results on a set of real data show the estimated field.

Cell shapes were modelled as B-spline parametric active contours. By casting the parametric active contour model in state space form, Kalman filter was employed to track the shapes. Shape tracking required solving the problem of cell boundary association between two time frames which required identification of correspondence points at cell boundaries. This required improvement to a nearest neighbour filter method to give continuity of the cell shapes of the same neutrophils across the different frames. Characterisation of cell shape was carried out by employing Fourier descriptors from which two features, magnitude of the highest descriptor and the magnitude of the lowest frequency, were chosen as proxy for

associating cell shapes with cell states. These features for instance are associated with cell roundedness. Both tracking methods, cell centroid and shape tracking, are employed on the real data and results shown to demonstrate the effectiveness of the methods.

This thesis makes the following novel contributions. Firstly, a new framework was established to solve the cell centroid and shape tracking towards characterising cell shape behaviour. Secondly, the nearest neighbourhood filter employed to solve the association problem was improved to solve the problem of neutrophil cells disappearing and reappearing in image frames. Thirdly, the low frequency Fourier descriptor, combined with other methodologies, was successfully implemented in detecting the modes of neutrophil behaviour. In addition, the chemotactic field was estimated by using the centroid velocities. Furthermore, the multiple model filter was used for behavioural mode detection of cells.

Key Words

Neutrophils, the Kalman Filter, Nearest Neighbourhood Filter, B-Spline, Fourier Descriptor, Multiple Model Filter, Velocity Estimation, Low Frequency Descriptor

Papers Arising from This Thesis

V. Kadirkamanathan, S. Anderson, S. Billings, X. Zhang, G. Holmes, C. Reyes-Aldasoro, P. Elks, and S. Renshaw. The Neutrophil's Eye-View: Inference and Visualisation of the Chemoattractant Field Driving Cell Chemotaxis *In Vivo*. *Plos One, 7, 2012*.

Abbreviations

Abbreviation	Description
KF	Kalman filter
MM	multiple model
UKF	unscented Kalman filter
HIV	human immunodeficiency virus
NETs	neutrophil extracellular traps
G-CSF	granulocyte colony stimulating factor
GFP	green fluorescent protein
RFP	red fluorescent protein
NMR	nuclear magnetic resonance
FFT	fast Fourier transform
NNF	nearest neighbourhood filter
INNF	improved NNF
LFD	low frequency descriptor
DFT	discrete Fourier transform
EKF	extended Kalman filter
FDM	finite difference method
FEM	finite element method
PF	particle filters
SMC	sequential Monte Carlo
CT	computerized tomography
AR	autoregressive
LS	least square
CSS	curvature scale space
FD	Fourier descriptor
WD	wavelet descriptor
SFD	short-time Fourier descriptor
CCH	chain code histogram
NCCH	normalized CCH
JPDA	joint probabilistic data association
IMM	interacting multiple model
IMBM	interacting multiple bias model
IMAM	interacting multiple acceleration model
UAV	unmanned aerial vehicles

Contents

List of Figures

Chapter 1

Introduction

This thesis is mainly focused on applying tracking methodologies to neutrophils' problems, including centroid tracking, shape tracking, dynamic behaviour analysis, together with uncovering the process and mechanism of neutrophils when they respond to inflammation.

1.1 Background

Before presenting the research on neutrophils, relevant medical background detail will be considered, starting from the immune system.

Immune systems are highly significant for biological systems since they can protect health by ingesting invaders. They can identify the non-self (foreign bodies that are not from the host), find them, destroy or remove them. However, immune systems are not effective against all foreign bodies such as the human immunodeficiency virus (HIV).

Usually, there are three steps in respect of the working process of an immune system [14]. Firstly, the immune system distinguishes the non-self. Then it destroys all invaders, most of which are dangerous, such as bacteria. Finally, the immune system eradicates diseased cells or tissues, such as cancer.

The immune response of higher vertebrates, such as human beings, can be grouped into two types, that is, innate immunity and acquired immunity [14]. The former, also called natural immunity, has an effective defence before an invader appears with all animals and plants having innate immunity [98]. There are more than 30 proteins that can usually recognise and eliminate invaders in higher vertebrates [14]. If the innate immunity cannot destroy invaders, the acquired immunity comes into effect and relies on lymphocytes, which can be divided into B and T classes. The lymphocytes remain inactive until they meet antigens. The ac-

quired immunity response is very slow since its processes are complex. However, a trait called immunologic memory can reduce the response time, by recognising the same invader and triggering the response [14]. The immune system, with other systems such as the vascular, propagates and matures the inflammatory response.

Inflammation is a complex biological process in which harmful stimuli are eliminated to heal the host. The mechanism of inflammation has been described by Henson [66]. Usually, inflammation is beneficial to the host. For example, Haslett [64] reported the mechanism of the process in which the apoptosis neutrophils in inflamed tissues are cleaned. However, sometimes, it will cause an allergic reaction and attack the tissues or cells of the host. Haslett [63] put forward the idea that if neutrophilic inflammation were not resolved, it would cause irreparable tissue damage.

Inflammation can be divided into two classes: acute and chronic. The former is the process during which plasma and leukocytes (especially granulocytes) move to the stimuli site in increasing numbers. Usually, acute inflammation lasts a short time. However, chronic inflammation, where destruction and healing happen at the same time, lasts longer.

According to Lieschke [96], *in vivo* models are used in the only effective studies on inflammation. *In vivo* models are different from *in vitro* ones. The former uses a living organism while the latter uses a dead organism. Understanding human inflammation mechanisms requires the identification of a substitute animal for *in vivo* models. The neutrophil analysis in humans is the ultimate goal to finding treatments for humans; however, it is not easy to carry out experiments on human beings. As such, having an animal for an *in vivo* model which is wild, visualizable and contains many individual immune cells such as neutrophils is necessary, since *in vivo* models are significant to research into inflammation.

The traditional animal models used in research are Drosophila and mice [95]. Recently, powerful substitutes have emerged, such as zebrafish. Lieschke [95] reported that the traditional animal models, such as Drosophila and mouse, are not as good as the zebrafish in respect of complexity and suitability. Given the transparency of zebrafish larvae, they can easily be stained with fluorescent proteins to observe the immune system cells *in vivo* [128]. There are additional advantages of utilizing zebrafish as a model, according to Thisse and Zon [153], such as reducing the space required for feeding. Applications of zebrafish to *in vivo* investigation include Jong and Zon [78] who showed the process of hematopoiesis and Crowhurst and Lieschke [34] who undertook research on leukocytes. Neutrophils, a type of leukocytes, can be stained by fluorescent protein and have the fastest response of all leukocytes. Using a zebrafish, an *in vivo* model can, to some degree, show in-

flammation. As such, neutrophils in zebrafish *in vivo* are an excellent testing bed for the study of inflammation.

Neutrophils, also called neutrophil granulocytes, are a type of white blood cells. White blood cells, also referred to as Leukocytes (also spelled "Leucocytes"), are cells that protect human's body from a large number of diseases. Neutrophils account for approximately 50% to 70% of human white blood cells, while only 10% to 25% exist in mice [106]. According to [89], neutrophils move to the inflammatory site and clean the infection from both the experimental data and the clinical data. Furthermore, research [162] and [38] showed that some human immunodeficiency is caused by a decreasing number of neutrophils. Therefore neutrophils play an important role in protecting the body by preventing the invading pathogens, making them highly significant for humans and other animals. However, excessive neutrophils will also lead to some tissue damage [139].

Initially, neutrophils were only thought to be useful to pro-inflammatory effects. However, increasingly, experimental data has shown that there are other actions, such as anti-inflammatory [89]. Another effect shown by Fournier and Parkos [48] was that neutrophils are significant for wound debridement. Additionally, according to Jablonska et al. [73], neutrophils have a positive effect on the growth and spread of tumour.

Since too many neutrophils can damage tissues, numbers should be controlled. Often, bio-chemical mediators can increase or decrease the number of neutrophils by controlling the response of phagocytes [139].

Neutrophils are generated in the bone marrow. The lifecycle of neutrophils is very short, no greater than 12 hours for mice and 5.4 days for human beings [122]. Some mature neutrophils can live outside the bone marrow in tissues in certain circumstances and stay in organs, such as the liver or lungs [146] [120]. They are helpful to some degree, but they can also damage other cells [89]. Usually, neutrophils die in two ways: a normal death and the neutrophil extracellular traps (NETs). The former means that neutrophils are cleared in organs, such as the liver, bone marrow [68] [141], and vasculature. NETs will be introduced later in this chapter. Once neutrophils died, signals were given out which attracted monocytes and these then "ate" the dead neutrophils and other dead cells in tissue [143].

Research [112] [44] [67] has explained of the process in which the mechanism kills microbes. Usually, there are three kinds of protection mechanisms (both intracellular and extracellular) when neutrophils work, namely:

- Phagocytosis (the process of swallowing virus)
- Degranulation (the process of delivering soluble anti-microbial)

- NETs

The phagocytosis process is when neutrophils "eat" the virus, after which reactive oxygen species or antibacterial proteins will kill the viruses, as per the literature [27]. This process belongs to the intracellular mechanism.

In the degranulation process, antibacterial proteins are released into the surrounding environment, killing the virus. This process belongs to the extracellular mechanism.

NETs are an extension of antimicrobials [27]. First reported by Brinkmann et al. [28], it was believed that when inflammation occurs, neutrophils can release intracellular chemical substances, into extracellular circumstances and this is called NETs. This process can either kill pathogens [58], called apoptosis, or not kill [105], called necrosis. The reason has not been clear until recently, and the literature [89] offers some hypotheses. However, according to Saitoh et al. [135], NETs is ineffective in respect of HIV. The mechanism of NETs formation is still not clear. This process belongs to the extracellular mechanism.

Neutrophils can generate lethal hits against microbes, but at the same time, they may kill other tissues. Neutrophils are mostly dormant cells. When infection happens, neutrophils are activated and increased by the tuning act of granulocyte colony stimulating factor (G-CSF). Usually, the number of neutrophils increases, while the production of G-CSF decreases.

The zebrafish is transparent and therefore can be observed in brightfield intravital microscopy. Historically, intravital microscopy has been used to observe some thin and transparent tissues [103]. The advantage of this method is that it can observe mobile cells, such as neutrophils. Atherton and Born [8] pioneered this work. About 25 years ago, it became popular and widely used [155]. More recently, thicker tissue and higher resolution have been possible via powerful microscopies. In this current study, both high resolution neutrophils and low resolution were used for different goals. Fluorescence microscopy was also involved in the pre-processing work. Neutrophils and macrophage were labelled by green fluorescent protein (GFP) and red fluorescent protein (RFP), respectively.

It should be noted, particularly in respect of this research, that neutrophils are special cells. Firstly, they always change their shapes to respond to the chemotaxis change, while most of the other cells cannot change shapes. For this reason, online shape tracking was required as part of this current study. Furthermore, the motion and morphology of neutrophils are important to understand the process of neutrophils responding to chemotaxis. Secondly, the pseudopod always emerges during movement to the inflammatory site, while most other cells do not have pseudopods. Therefore, shape tracking, especially on pseudopods, is signif-

icant to understand the mechanism of neutrophil movement. Finally, neutrophils in either tail transection experiments or the normal situation, have several modes which are active, "dead", and the mode changing between the other two. The "dead" mode means the neutrophils move randomly around some small areas. The numbers of neutrophils in the three modes differ between cases. For example, in the tail transection experiment, the number of active neutrophils was obviously larger than that of the other two modes. It is believed that because of the three modes, neutrophils must maintain an optimum number since too many or too few are likely to cause damage. Therefore, the mode detection of neutrophils does help to understand their complex mechanism.

During the whole programme, many methodologies were applied. Most are well known and frequently used in other areas. For example, the Kalman filter (KF) was used in the Apollo program. As such, the related works and applications will be introduced and reviewed.

Computer vision, as an important area, was firstly introduced. According to [140], computer vision mainly deals with images or videos, including processing and analysis. It usually acquires real, high-dimensional data. After being processed and analysed, the data is transformed into special forms which can be understood by computers. The process can be generalised as "looking", "thinking", and "deciding". Its origin can be traced back to the 1960s. However, it underwent rapid development in the 1980s [140]. Computer vision includes video tracking, image restoration, and its applications encompass navigation, and , most significantly, medical image processing, which is highly significant for patients' diagnoses [74]. For instance, Nuclear Magnetic Resonance (NMR) is always used to check tumours inside the body. Medical image processing has been used in this research, for data pre-processing. This application area mainly relies on the information extracted from images, which are usually X-ray or tomography. Machine vision is perhaps another prominent area, different from medical image processing. Usually, programs are preinstalled to finish particular tasks. For example, a robot arm can be programmed to catch a specific tool. In addition and as a new subject, autonomous vehicles have become increasingly popular with the application in unmanned aerial vehicles (UAV) being one of the most famous. More related works, including edge detection and threshold, are referred to in the next chapter. Applications of B-Spline and Fourier descriptor (FD) will also be addressed.

Another important area is aerospace engineering namely aircrafts and spacecrafts [20]. Usually it is structured in aeronautical engineering and astronautical engineering. Aerospace engineering dates back one hundred years ago with Sir

George Cayley regarded as the pioneer in this area, with the Wright brothers having flown successfully in 1903. Aerospace engineering developed quickly, due to military reasons and its wide applications include control engineering. Aircraft and spacecraft are the most famous applications. For example, the moon have not seem mysterious since the spacecraft landed on it in Apollo program. Radar and target tracking are also important applications, such as bearings-only tracking [130]. Those algorithms in radar and target tracking area has been used in this research, such as the Kalman filter in both centroid tracking and shape tracking. More related works, including the nearest neighbourhood filter (NNF), are referred to in the following chapters.

Recently, cell tracking has become more popular. It was first applied more than 30 years ago and has developed rapidly in the last 10 to 15 years. More advanced tracking methodologies have been applied, including particle filters (PF) used in mitotic cell tracking [124]. E. Meijering and Smal plotted a figure to show the increment of research on cells and tracking in recent years [42]. Cell tracking also covers a large number of areas, such as image processing, in addition to having a wide range of applications. The NNF is used in a large number of cell tracking applications, because it can effectively fix the index and keep the continuity. It is also suitable for online tracking, another reason for its wide use. Therefore, it has been also used in this research. Velocity estimation is another application, with instantaneous velocity having been measured by J.B. Beltman and de Boer [76]. However, the velocities on the shape boundary contain more useful information. Therefore, they have been estimated in this research in order to further uncover the mechanism.

1.2 Motivation

The project merits investigation because the morphology and motion research on neutrophils could help to further understand the inflammation mechanism and processes. A more thorough understanding of the processes and mechanism could lead to new therapeutic functions for both animals and humans.

The morphology and motion of neutrophils were obtained from processing the initial data by mathematical and engineering techniques. The initial data was received from pre-processing data, which was translated from raw time lapse images *in vivo* and the time lapse images *in vivo* were obtained from observing zebrafish. The whole process has developed a new methodology for researching and analysing neutrophils and other cells.

The interest can be generalised as the following aspects. Firstly, it is believed

that a field, similar to a magnetic field, makes neutrophils move to the inflammatory site. This was invisible but could be expressed, just as the magnetic field was expressed by the distribution of iron sand. Therefore, the problem could be simplified to another mathematical expression. This mathematical expression could be considered as a mathematical model, which was established by using engineering techniques. It is interesting not only because it is a new application, but also because the medical phenomenon of the neutrophil moving mechanism is solved. Secondly, during the moving process of neutrophils to the inflammatory site, the neutrophils' morphology changed. For instance, pseudopods emerged in some frames and the shapes of most of the neutrophils changed in each frame. It is believed that the morphology changes were driven by the mechanism. Therefore, further understanding of the morphology changing will help to better understand the mechanism. The problem was consequently transformed into the tracking of the morphology changing. The problem of tracking the morphology changing can be effectively solved by the new framework of shape tracking in this thesis. Finally, an effective tool to analyse the changes of the morphology is needed with the low frequency descriptor (LFD) being considered.

The challenges were listed as follows. Firstly, the pre-processing of the raw data was the challenging. The raw data was figures from the Volocity 5, shown in Figure 1.1. Since the data was three-dimensional images, they had to be transformed to two-dimensional binary matrices, which were imported in Matlab. Secondly, traditional mass point tracking methodologies were applied to the new application in neutrophils and other cells. However, the mathematical model of neutrophil centroid trajectory tracking was difficult to establish. The motion of neutrophils was complex and, therefore, it cannot be expressed as any commonly used model. As such, establishing the mathematical model became the challenge. Thirdly, the mathematical model for the shape tracking presented additional difficulty. As the morphology of neutrophils changed in each frame, the boundary points were difficult to fix in the same position. In addition, indices of shape boundary points rotated in different frames. The above reasons led to the challenge of establishing the mathematical model. Fourthly, maintaining the continuity of the same neutrophil in different frames was another issue. The traditional multi targets tracking had the ability but whether the neutrophils were the same, or not, in the two consecutive frames could not be guaranteed. Therefore, it had to be combined with other techniques to solve the problem. Finally, determining an effective tool to analyse the characteristics was difficult. The FD was a good choice, however, it is usually used in a one dimensional area. Therefore, the final challenge was to analyse the neutrophils' characteristics using an effective tool.

Figure 1.1: **The initial neutrophils T00001C02Z004**

This is the raw data from the software called Volocity 5 (Improvision, Perkin Elmer Inc.). It is from the tail experiment. It records the first zebrafish one hour after the experiment. The name T00001C02Z004 is the combination of different important information, where T00001 means the first frame; C02 means channel two; and Z0004 means the fourth height. It is the low resolution data. In the figure, the white areas are neutrophils and all the other black areas are background. It should be noted that neutrophils are minimal. However, the figure is the best one in the raw data. This figure will be preprocessed and the details will be introduced in the next chapter.

1.3 Rationale for Research Study

The initial step in solving the problem was to prove the existence of the field. As with iron sand in the magnetic field, neutrophils should be considered as mass points. The neutrophils' dynamic distribution had to prove the existence of the chemotaxis field. Therefore, it was thought traditional centroid tracking methodologies would help. However, all the traditional tracking methodologies require an accurate mathematical model and having considered the motion of neutrophils, none of the regularly used mathematical models was deemed suitable. Therefore, a combination of the commonly used models was required. Generally, the multiple model (MM) filter is used to establish the mathematical model of neutrophils. Then the KF is used to estimate both the position and the velocity of neutrophils. Information about the positions and velocities, as per the iron sand, were thought to be relevant in respect of proving the existence of the chemotaxis field and estimating the chemotaxis field.

After determining the centroid information, ascertaining the shape information is key. Therefore, the second step is to track and estimate the neutrophils' shape. Given the knowledge of traditional tracking, the aim was to consider the shape boundary as the combination of a large number of mass points. Nevertheless, guaranteeing that a point on the boundary is the same point in the next frame generates an additional challenge, yet a methodology from computer vision could offer a solution. The parameter active contour can transfer the position to control points and vice versa. Therefore, the shape information is transferred to the information of control points. The combination of tracking and parameter active contour could estimate the positions of the shape and other information of the feature. The only problem would be the continuity of the neutrophils. Based on continuity, the NNF was requested and it was hoped the combination of the above methodologies would solve the shape tracking problem.

An effective analysis tool was needed for both the centroid and the shape information. Therefore, the final step was to analyse those data. In addition, the analysis tool was required to be combined the tracking methodologies above. FD was a good choice, however, LFD was more suitable, since it could detect the mode of neutrophils in addition.

After undertaking the above three steps, the new framework had been established to solve the problem and allowed for a better understanding of the neutrophils' mechanism.

1.4 Organization

Having considered the rationale underpinning the research, this dissertation has been divided into three parts: the centroid tracking, the shape tracking and the shape analysing. It comprises seven chapters: the introduction, literature review, the centroid tracking, the shape tracking, the shape descriptor, the combining methodology and the conclusion. A synopsis of each chapter follows.

In this introduction chapter, the medical background especially of neutrophils, has been introduced in detail. The motivation for the research has been expounded and the thesis' structure presented. Innovation and contributions to research are given and finally the data is introduced including the raw data.

The literature review focuses on the core academic knowledge of signal and image processing. With relevant background details having been discussed in the introduction, several specific methods are presented, including filter based signal processing, model based signal process, and image processing. The descriptors are also introduced in detail. Finally, the pre-processing is presented.

The centroid tracking chapter is considered the basic research target, completed in the first year of this PhD. The key assumption is to consider a neutrophil as a mass point. Based on this assumption, a huge number of methods and engineering techniques could be used to track and analyse the trajectory, such as the KF [130]. However, neutrophils always change their modules caused by the changing chemotaxis. Therefore, multiple models were established to adapt to the changing of modules and probability was calculated to determine which model was to be selected in each frame. By mathematical derivation, the velocity square (v^2) has a relationship with the chemotaxis field. Therefore, it is also introduced in this chapter, as an analysis tool for the chemotaxis field estimation. The chapter has been organised as follows. Background of tracking algorithms is introduced in the introduction section. In the next section, the MM algorithm and interacting multiple model (IMM) algorithm are introduced. The derivation of both algorithms is complex with a large amount of probability calculation, so this is not covered. The derivation details can be found in relevant papers and books, such as [10]. In the results section, this thesis' the high resolution data is presented. The KF and the MM filter were implemented in this section and the parameters, such as $\boldsymbol{Q}$, tested to discover the "best". Finally, the analysis undertaken to estimate the chemotaxis field with the conclusion presented at the end of the chapter.

The shape tracking chapter is considered to be an upgraded form of centroid tracking and the second year of this PhD was devoted to it. Based on the state space model, the neutrophil shape boundary can be viewed as a combination of

all the points on the shape boundary with position (and velocity) as the measurement output. Therefore, the problem changed to a familiar case, as per the centroid tracking. However, problems still exist, such as fixing the index of boundary points and determining the state is. B-Spline, as an effective tool, is introduced to consider the "control points" as the state, and its potential to retrieve the shape. However, the index must be known, otherwise it can only work in individual frames and cannot work in continuous ones. The NNF solved this problem by determining the index in the different frames. Finally, the KF was used to estimate the state, which replaced the control points and calculated the velocity of the boundary. The above methods were combined as the methodology of shape tracking. This not only reduced the calculation complexity but also "read" the velocity (which had not previously been considered) as a tool to analyse neutrophils, part of the work in the subsequent chapter. The shape tracking chapter is organised as follows. In the introduction, the background of the shape tracking algorithm is introduced. In the second section, a high order state space model (also called the augmented state space model), B-Spline, the NNF and the KF for the velocity estimation are proposed with formulae. The improved NNF (INNF) was also introduced to solve the problem of neutrophils disappearing in some frame k and reappearing in the subsequent frame $k+1$. In the results section, B-Spline, NNF, INNF, and the KF are implemented with low resolution data. Estimated velocity on the boundary is also retrieved. A conclusion to this section is provided.

The shape descriptor chapter is considered as a tool to analyse the results from shape tracking chapter. It comprised the third year work of this PhD. In the previous chapter, the information of shape boundaries was obtained. The focus of this chapter is the analysis of the shapes. The FD as an effective tool was selected for analytical work. In addition to the FD, other descriptors were implemented, such as the LFD. The pseudopod was also analysed. The organization of the chapter is as follows. In the introduction, the FD is detailed. The height descriptor is introduced as well and the radius descriptor is mentioned in respect of future work. In the results' section, The FD is implemented. As the FD is the frequency domain of a two-dimensional shape without physical meaning, tests are undertaken to verify hypotheses and basic properties. After verification, the shape descriptor is combined with shape tracking to obtain the results. The height descriptor was implemented to determine the relationship between the height of the FD and the roundness. LFD is also implemented with a very good result being achieved, since it, to some degree, can classify the neutrophil module. In the pseudopod analysis, an active positive moving neutrophil was found. The FD and shapes are compared with the subsequent two frames. Finally, a conclusion section is presented at the

end of this chapter.

The combining methodology chapter has its focus on assimilating and synthesising the previous chapters to present a new methodology, which can deal with similar problems or other cells. In this chapter, only results are given. The chapter is divided into the following sections. The first offers a comparison between the mode from centroid tracking and LFD which was undertaken to ascertain if a relationship exists. It combined mode detection obtained from the MM filter in chapter three and the shape descriptor in chapter five. In the second section, centroid velocity was combined with LFD to determine the connections. It combines the velocity estimation from chapter three with the shape descriptor from chapter five. Then shape velocity distribution was compared with LFD. It links the shape velocity estimation from shape tracking in chapter four to the shape descriptor in chapter five. However, from the above comparisons, it has been extremely difficult to illustrate any relationship separately, due to the complexity of neutrophils. Therefore, the combination of all the methods would appear to be the best solution to this complex issue. In the neutrophil tracking and analysis section, as mentioned above, the combination result was presented as a new methodology. The concluding section of this chapter is given as a summary.

The conclusions' chapter is the final one of this thesis. It provides a summary section and suggestions for future work. In the summary, all the results, innovation, and contributions are noted. In the future work section, methods which can be improved and new ideas are proposed. These include improving the INNF to determine the index of neutrophils even when the neutrophil disappears in frame k and reappears in frame $k+i$, where i is a positive integer greater than 1.

1.5 Innovation and Contributions

This section considers the innovation and contributions offered during this PhD research. The chemotaxis field was estimated using centroid velocity and it formed a part of the paper [82]. The MM filter was implemented and used for cell behavioural mode detection based on neutrophil centroid information. The traditional NNF was improved to solve the problem of neutrophils disappearing and reappearing. Shape tracking techniques were implemented and the information of shape boundary velocity and the behaviour of the characteristics were extracted, based on shape information. New descriptors were applied, such as LFD. Neutrophil features were identified by using a shape descriptor. The information on centroid and shape was integrated for neutrophil behavioural characteristics.

This work's most important contribution is the establishment of a framework

to track both the centroid and shape of neutrophils systematically. It also tracked the velocities on the shape boundary of neutrophils. Furthermore, the framework can automatically analyse the features of neutrophils. In addition, it is also suitable for other cells and it, therefore, formed a methodology to deal with tracking and analysis of cells.

1.6 Data Introduction

1.6.1 Ethics Statement

All the experiments on zebrafish were performed according to guidelines and legislation set out in UK law in the Animals (Scientific Procedures) Act 1986. Ethical approval was given by the University of Sheffield Local Ethical Review Panel.

1.6.2 Data Acquisition

The data used in the thesis was from the zebrafish larva transection experiment, shown in Figure 1.2. The experiment is the same as one described in [128] and [82]. Figure 1.2 illustrated that the tail fin of a zebrafish larva was transected and therefore inflammation occurred. Due to the inflammation, all the neutrophils moved to the inflammatory site. In the experiment, the neutrophils were stained with GFP, as shown in Figure 1.2. The video microscopy data was recorded to describe the experimental process. Finally, the process was performed in Volocity 5 (Improvision, Perkin Elmer Inc.) by time lapse images. The raw data of the first zebrafish in the tail transection experiment from Volocity 5 is shown in Figure 1.3 and the neutrophils are shown in Figure 1.1.

The video frames were first transformed to a sequence of three-dimensional binary images automatically in Volocity 5. Then those images (raw data) were compressed into two-dimensional binary images. After a double threshold method, the two-dimensional binary images were transformed to the spare matrices. The above process was pre-processing which will be explained in the next chapter and the data obtained is referred to as the initial data in this thesis.

Figure 1.2: The experiment on zebrafish.

The tail fin of zebrafish larva was transected and all the neutrophils stained with GFP moved to the inflammatory site in the experiment. This experiment mainly focused on the neutrophil response to the inflammation and was the same as described in [128] and [82].

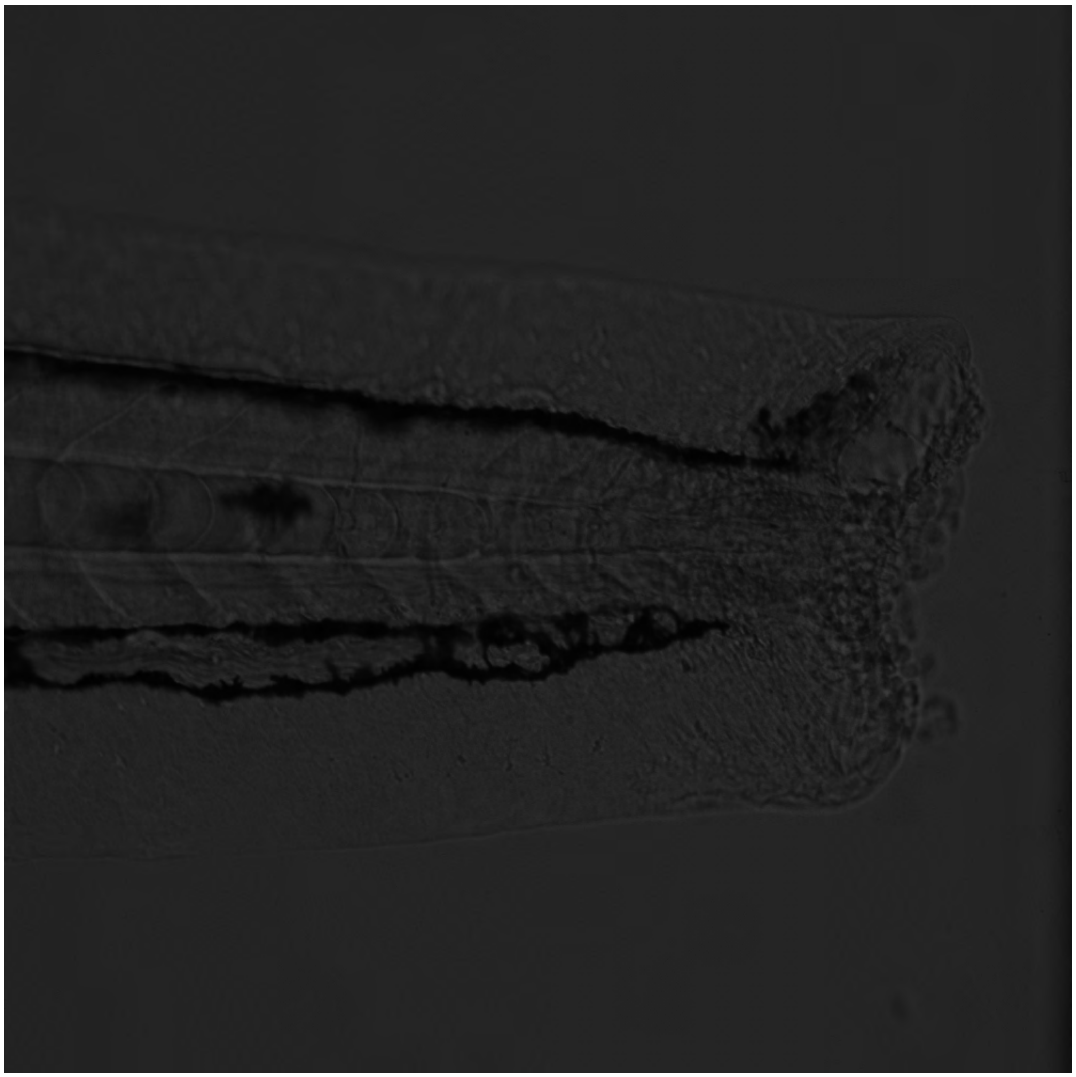

Figure 1.3: **The initial zebrafish T00001C01Z004**

This is the raw data from the software called Volocity 5 (Improvision, Perkin Elmer Inc.). It is from the tail experiment. It records the first zebrafish one hour after the experiment. The name T00001C01Z004 is the combination of different important information, where T00001 means the first frame; C01 means channel one; and Z0004 means the fourth height. In the figure, the outline, the skeletons, and the vessels of the zebrafish are clear. Even the transection is clear on the right.

Chapter 2

Signal and Image Processing

The previous chapter mainly focused on the introduction of the medical background. The research motivation and the dissertation's structure were also explained, together with its innovation and contributions. Finally, the data acquisition was illustrated with raw data images and the image of the transection experiment. This chapter covers the literature review and the data pre-processing. The literature review consists of three parts: signal processing, image processing and descriptors. The data pre-processing has been used to study the data presented in the previous chapter.

2.1 Introduction

According to Moura [111], signal is defined as physical information utilized to present the change in time domain and/or space domain at the beginning. Signal was also used to show abstract information or a sequence, such as the combination of genetic codes. Processing is a type of operation, such as filtering. Therefore, signal processing is an operation utilized to deal with signals.

The history of signal processing is very long with the earliest signal processing dating back to classical numerical analysis, 400 years ago according to Oppenheim and Schafer [113]. The range of signal processing applications is considerable, including engineering and medicine. It is applied not only in military and industrial areas but also in consumer electronics with home entertainments, such as TVs using signal processing [114]. As technology advances, wireless communication equipment, such as smart phones, increasingly rely on signal processing. Oppenheim et al. believed in the future, signal processing would play an increasingly important role in consumer, industrial and governmental areas [114]. As such, the importance of signal processing is assumed.

According to Oppenheim et al., signal processing is a sophisticated process including representation and transformation [114]. Usually, digital signal processing and analog signal processing exist. The latter applies when both the input and output are continuous time values. Digital signal processing is a discrete time signal, whereby either the input or the output is a sequence instead of a continuous waveform. Signal processing was analogue because of the continuous time system technique at the beginning; however, with the development of computers and techniques, such as fast Fourier transform (FFT), digital signal processing developed rapidly. Currently, signal processing always means the digital type. The foundation of digital signal processing is sampling. Based on the sampling technique, a significant amount of continuous time input can be sampled into a sequence. This is the input of digital signal processing. Then after processing, the discrete time output can be retrieved to a continuous time signal, as the system's output. In most engineering cases, the system is a real time one. This leads to the requirement that the sampling rate must be kept the same between both input and output. Spectral analysis is another area of signal processing, with its core idea being to estimate parameters and evaluate the frequency response. Traditional signal processing deals with one dimensional signals. However, with techniques advancing, multi-dimensional signals are required in some areas, for example, medical imaging. Recently, not only traditional engineering but also biomedical engineering have increasingly focused on signal processing and signal processing has gained in popularity.

Image processing is the upgraded form of signal processing. Its input is image and its output is either image or feature. Image processing always considers the input as a two-dimensional signal while traditional signal processing is single input. As such, most signal processing techniques can be used in image processing. Other techniques, for image processing only, such as edge detection, are presented later in this chapter. Image processing, when mentioned, always means digital image processing, but there is also analog image processing. Since its cost is low and the algorithm is simple, image processing is increasingly popular in many areas.

2.2 Filter based Signal Processing

Filter based signal processing refers to signal processing using filter techniques. Usually, it is related to frequency domain, however, in some special cases, filter based signal processing works in a spatial domain, such as smoothing a two-dimensional vertical signal (in special cases, it is referred to as an image). The frequency filter can be implemented either in the transfer function or in frequency

domain methods, such as the Fourier transform.

2.2.1 Transfer Function

The transfer function, also called a system function, shows the relationship between the input and the output in the s-plane. It is based on the Laplace transform. It is usually used in a single input single output linear time invariant system and expressed as $G(s) = \frac{Y(s)}{X(s)}$, where $Y(s)$ and $X(s)$ are the Laplace transform of output and input; $s = \sigma + j * \omega$ is the complex Laplace operator and usually, $\sigma = 0$, that is, $s = j * \omega$, which changes the Laplace transform to the Fourier transform with the frequency ω. It is commonly used as a filter in signal processing and famous filters include the Butterworth filter, the Chebyshev filter (Type I), the Chebyshev filter (Type II) and the Gaussian filter with overviews of these presented below.

The Butterworth filter has a property of maximally flat frequency response in the passband. This filter is widely used in frequency analysis especially when plotting Bode plots. Butterworth's pioneering work included showing the low-pass filter and providing a general definition of filter, which deleted noise frequencies and had uniform sensitive wanted frequency.

The Chebyshev filter was based on the Butterworth filter but with modifications in sharp cutoff and flat in stopband (Type I) or flat in passband (Type II). The Chebyshev filter was named after Pafnuty Chebyshev since the mathematical characteristics were derived from Chebyshev polynomials.

The Gaussian filter refers to one where the impulse response is or approximates to a Gaussian function. The Gaussian filter has minimum group delay and no overshoot to a step input. Based on the properties above and the properties of Gaussian function, the Gaussian filter is an ideal time domain filter and commonly used in many areas, such as telecommunication systems.

2.2.2 Frequency Domain

The frequency domain, different from time domain, always refers to the mathematical analysis based on frequency. For example, in time domain, a signal is analysed based on a time variable, while in the frequency domain, a signal is analysed by different frequencies. In the frequency domain, the spectrum or spectral density is always used as a presentation of a signal. Another commonly used tool is the power spectral density. There are two parameters, which are magnitude and phase, in the frequency domain. Magnitude is the amplitude of the complex signal and phase is the angle of the signal. Fourier transform is widely used in the frequency domain, because it has the ability to change from time to frequency and

vice versa. Z transform and wavelet transform are widely used in signal processing and image processing. Laplace transform, as the more general form of Fourier transform, is used in control systems and electronic circuits.

2.2.3 DFT and FFT

Discrete Fourier transform (DFT) is one of the most effective and commonly used algorithms in digital signal processing, according to Lyons [100]. DFT is a straightforward method to discretize the continuous time Fourier transform into the discrete frequency sequence. The mathematical expression of DFT is as follows:

$$T_k = \sum_{j=0}^{N-1} t_j e^{-i2\pi jk/N} \tag{2.1}$$

where T_k is the Fourier sequence with frequency k; t_j is the j^{th} sampling value from the continuous time variable; N is the length of the sampling sequence; i is the complex symbol with $i^2 = -1$.

It should be noted that the input and output of DFT are finite. DFT is linear, invertible and periodical Fourier transform, which means the inverse of DFT exists. The output of DFT represents different frequencies and both are orthogonal.

The disadvantage of DFT is its expensive cost. However, in 1965, Cooley and Tukey [33] first proposed an algorithm, FFT, which reduces the calculation time and increases the efficiency even for thousands of points DFT. FFT is equal to DFT when the length N of DFT is an integral power of two. However, the calculation cost of FFT is obviously cheaper than that of DFT, especially when N is considerable. An example was given in [100] to illustrate the effect of FFT in reducing calculation time and increasing efficiency. The core idea of FFT is to divide DFT into odd sequences and even sequences, then to transform the sequences with mathematics and Euler's formula. There are many FFT algorithms, such as the Cooley-Tukey FFT algorithm and the Prime-factor FFT algorithm.

2.3 Model based Signal Processing

Model, in this current work, means a mathematical model and is a description of a system using mathematics. A simple mathematical model always helps to reduce the complexity of the target. Therefore, a simple and effective model plays a highly significant role in researching the behaviour and making a prediction.

The commonly used model in classical control theorem is the transfer function mentioned above. Until the 1960s, with the development of mathematics and

modern control theorem, the state space model became increasingly important and popular. The state space model represents the input output relationship with first order differential equations (continuous time system) or difference equations (discrete time system) for a linear system. The advantage of state space is that it can represent a multiple input multiple output system and it can deal with a time domain system directly instead of using Laplace transform and the inverse of Laplace transform. It also reduces the computation cost and time. The commonly used formula is as follows:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{u} + \boldsymbol{w} \\ \boldsymbol{y} &= \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u} + \boldsymbol{v} \end{aligned} \tag{2.2}$$

for continuous time system and

$$\begin{aligned} \boldsymbol{x_{k+1}} &= \boldsymbol{F}\boldsymbol{x_k} + \boldsymbol{u_k} + \boldsymbol{w_k} \\ \boldsymbol{y_k} &= \boldsymbol{H}\boldsymbol{x_k} + \boldsymbol{D}\boldsymbol{u_k} + \boldsymbol{v_k} \end{aligned} \tag{2.3}$$

for discrete time system.

Where $\dot{\boldsymbol{x}}$ is the first order derivative of $\boldsymbol{x}$; $\boldsymbol{x}$ and $\boldsymbol{x_k}$ are the states which have different physical meanings based on different systems; $\boldsymbol{A}$ and $\boldsymbol{F}$ are system matrices which decide the system characteristic; $\boldsymbol{u}$ and $\boldsymbol{u_k}$ are the control variables; $\boldsymbol{w}$, $\boldsymbol{v}$, $\boldsymbol{w_k}$, and $\boldsymbol{v_k}$ are white noises; $\boldsymbol{y}$ and $\boldsymbol{y_k}$ are the outputs of the system; $\boldsymbol{C}$ and $\boldsymbol{H}$ are the observation matrices; $\boldsymbol{D}$ is the feedforward matrix and always equal to zero or zero matrix; $\boldsymbol{x_{k+1}}$ is the state at time $k+1$.

In addition to the state space model representation, commonly used physical models are needed. Generally, there are three basic and commonly used physical models, namely: the constant acceleration model, the constant velocity model and the random walk model.

The constant acceleration model is one of the most commonly used physical models. It is called the constant acceleration model because the expected acceleration remains unchanged, although, the acceleration is randomly perturbed from this constant value by white (Gaussian) noise. As such, there are three dimensions (3D or 3 states, that is, position, velocity, and acceleration) in each coordinate axis. As only a discrete time system has been used in this thesis, the continuous time system is not discussed in this current work. The discrete time model with state space representation is as follows:

$$\boldsymbol{x_{k+1}} = \boldsymbol{F}\boldsymbol{x_k} + \boldsymbol{w_k} \tag{2.4}$$

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F_{11}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F_{22}} \end{pmatrix} \tag{2.5}$$

$$\boldsymbol{F_{11}} = \boldsymbol{F_{22}} = \begin{pmatrix} 1 & h & \frac{h^2}{2} \\ 0 & 1 & h \\ 0 & 0 & 1 \end{pmatrix} \tag{2.6}$$

$$\boldsymbol{cov}(\boldsymbol{w}) = \begin{pmatrix} \boldsymbol{Q_{11}} & \boldsymbol{Q_{12}} \\ \boldsymbol{Q_{21}} & \boldsymbol{Q_{22}} \end{pmatrix} \tag{2.7}$$

$$\boldsymbol{Q_{11}} = \boldsymbol{Q_{12}} = \boldsymbol{Q_{21}} = \boldsymbol{Q_{22}} = \begin{pmatrix} \frac{h^5}{20} & \frac{h^4}{8} & \frac{h^3}{6} \\ \frac{h^4}{8} & \frac{h^3}{3} & \frac{h^2}{2} \\ \frac{h^3}{6} & \frac{h^2}{2} & h \end{pmatrix} \tag{2.8}$$

where h is sampling time, $\boldsymbol{cov}(\boldsymbol{w})$ is the discrete time covariance of the white noise input.

The output function is

$$\boldsymbol{y_k} = \boldsymbol{H}\boldsymbol{x_k} + \boldsymbol{v_k} \tag{2.9}$$

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{2.10}$$

The constant velocity model is also a commonly used physical model and the velocity is expected to maintain a fixed constant value. White (Gaussian) noise which disturbs the constant value exists. There are two dimensions (2D or 2 states, that is, position and velocity) in each coordinate axis. The discrete time model with state space representation is the same as the constant acceleration, however, the parameters have different sizes and expressions as follows:

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F_{11}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F_{22}} \end{pmatrix} \tag{2.11}$$

$$\boldsymbol{F_{11}} = \boldsymbol{F_{22}} = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \tag{2.12}$$

$$\boldsymbol{cov}(w) = \begin{pmatrix} \boldsymbol{Q_{11}} & \boldsymbol{Q_{12}} \\ \boldsymbol{Q_{21}} & \boldsymbol{Q_{22}} \end{pmatrix} \tag{2.13}$$

$$\boldsymbol{Q_{11}} = \boldsymbol{Q_{12}} = \boldsymbol{Q_{21}} = \boldsymbol{Q_{22}} = \begin{pmatrix} \frac{h^3}{3} & \frac{h^2}{2} \\ \frac{h^2}{2} & h \end{pmatrix} \tag{2.14}$$

The output parameter is

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.15}$$

The random walk is a mathematical concept, first introduced by Pearson in 1905 [118]. There are many kinds of random walk models [32]. The model used in the program is nearly the same as the already discussed models. There is only one dimension (1D or 1 state, that is, position) in each coordinate axis. The change of the position is caused by the white (Gaussian) noise. The parameters are as follows:

$$\boldsymbol{x_k} = \left(s_{x_k}, s_{y_k}\right)^T \tag{2.16}$$

$$\boldsymbol{F} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.17}$$

$$\boldsymbol{cov}(w) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.18}$$

The output parameter is

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.19}$$

Model based signal processing is a type of tracking algorithm based on the three mathematical models above, with state space representation applying to signal processing, as per the application of the KF in signal processing.

2.3.1 Kalman Filter

The KF is named after the Hungarian-American scientist Rudolf Emil Kalman. The KF's main contribution is that it solves the problem of filtering in the prescience of stochastic disturbance and noise, by using the state space model representation. Therefore, it is widely used in aerospace and tracking, such as the Apollo program. The phase-locked loop (which appears in most electronic communication equipment, such as radio) is the most commonly used continuous time KF in the applications. However, in most applications, the discrete time KF is widely used based on the calculations by computers. The history of tracking and the KF is summarised below.

The pioneering work of statistical signed process first appeared in astronomy work [52]. Fisher [47] proposed the maximum likelihood estimate, which can solve problems effectively. A famous Russian mathematician, Kolmogorov [90], published a book about probability and random process theory in 1933. Wiener contributed (the Wiener-Hopf integral equation and its solution, and the Wiener filter), to the linear filters and prediction for random processes in [156]. The Wiener filter was the first statistical filter and the predecessor of the KF. However, the Wiener filter was based on the transfer function, which led to complex computing. From the basic work, some extensions were done. The finite-memory of the Wiener-Hopf integral equation was solved by Zadeh and Ragazzini [161]. Another simple solution of the Wiener-Hopf integral equation was given by Bode and Shannon [24]. The non-stationary case was studied by Booton [26]. Wiener also solved the nonlinear filter problem in [157]. Fourier analysis and the transfer function (frequency domain) were widely used to obtain useful information. However, this is a extremely difficult method because of the calculation complexity and difficulty of understanding physical meanings. As mathematics, especially probability and random process knowledge, advanced, tracking algorithms based on the pioneering work for time domain were determined. Tracking means filters. The earliest and best filter for the linear Gaussian model case is the KF, first developed by Kalman in 1960 [83]. The KF uses several first order differential equations (state space representation) instead of high order differential equations (the transfer function representation, also known as the Wiener filter) and solves those first order differential equations in parallel. However, it was not accepted in1960. Until the Apollo program, scientists had focused on the KF. The algorithm is very simple, therefore it has been widely used. It is an optimal recursive Bayesian estimator. For linear and Gaussian cases, no nonlinear filter can improve on the KF, however, for nonlinear or non-Gaussian cases, the KF loses efficacy.

The expression $\boldsymbol{x} \sim N(0, \boldsymbol{Q})$ means the random variable $\boldsymbol{x}$ satisfy the normal

distribution with mean 0 and variable $\boldsymbol{Q}$.

Consider a discrete time system [5] given by

$$\begin{aligned} \boldsymbol{x_{k+1}} &= \boldsymbol{F x_k} + \boldsymbol{w_k} \\ \boldsymbol{y_k} &= \boldsymbol{H x_k} + \boldsymbol{v_k} \end{aligned} \tag{2.20}$$

where $\boldsymbol{w_k} \sim N(0, \boldsymbol{Q})$ is the signal noise; $\boldsymbol{v_k} \sim N(0, \boldsymbol{R})$ is the observer noise; $\boldsymbol{F}$ is the discrete time system matrix; $\boldsymbol{H}$ is the discrete time observation matrix; $\boldsymbol{x_k}$ is the state vector at time k; $\boldsymbol{y_k}$ is the output vector at time k.

The main iterative formulae of the KF are as follows:

$$\boldsymbol{P_{k|k-1}} = \boldsymbol{F P_{k-1} F^T} + \boldsymbol{Q} \tag{2.21}$$

$$\boldsymbol{K_k} = \boldsymbol{P_{k|k-1} H^T} (\boldsymbol{H P_{k|k-1} H^T} + \boldsymbol{R})^{-1} \tag{2.22}$$

$$\boldsymbol{P_k} = \boldsymbol{P_{k|k-1}} - \boldsymbol{K_k H P_{k|k-1}} \tag{2.23}$$

$$\hat{\boldsymbol{x}}_k = \boldsymbol{F} \hat{\boldsymbol{x}}_{k-1} + \boldsymbol{K_k} (\boldsymbol{y_k} - \boldsymbol{H F} \hat{\boldsymbol{x}}_{k-1}) \tag{2.24}$$

where $\hat{\boldsymbol{x}}_k$ is the estimated state vector at time k; $\boldsymbol{K_k}$ is called the Kalman gain; $\boldsymbol{P_{k|k-1}}$ is the covariance of the state vector at time k given the output from $\boldsymbol{y_1}$ to $\boldsymbol{y_{k-1}}$; $\boldsymbol{P_k}$ is the covariance of the state vector at time k given the output from $\boldsymbol{y_1}$ to $\boldsymbol{y_k}$.

In the KF formulae, the parameters $\boldsymbol{Q}$ and $\boldsymbol{R}$ represent the signal noise covariance and the observer noise covariance respectively, which are not known but the design value of $\boldsymbol{Q}$ and $\boldsymbol{R}$ need only to be in the same scale as the true values. Theoretically, these two parameters need to be semi-positive definite or positive definite. However, in practice, sometimes even when the noise covariance satisfies the condition of non-negativity, the KF performance can suffer. The reason for this is that considerable or minimal eigenvalues of the covariance cause the matrix inversion to be singular and the iteration cannot proceed. Therefore, these two parameters are highly significant when the KF is implemented.

The state estimation problem can be divided into three parts based on the data information:

Filter: The filter uses the measurement data information up to time t. The filter recovers information from the noise corrupted measurement. This filter is always used in discrete systems. However, there is also a continuous time KF and

Algorithm 2.1 Kalman Filter Algorithm

Assume the initial condition $\hat{\boldsymbol{x}}_{\mathbf{0|0}}$ and covariance matrix $\boldsymbol{P}_{\mathbf{0|0}}$ are normal distribution
for $i = 1$ to N **do**
  Calculate the estimated state prediction $\hat{\boldsymbol{x}}_{\boldsymbol{i|i-1}}$ and the prediction of covariance matrix $\boldsymbol{P}_{\boldsymbol{i|i-1}}$
  Calculate Kalman gain $\boldsymbol{K_i}$, the upgrade covariance matrix $\boldsymbol{P_i}$ and the state estimator $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$
end for

the solution to the continuous time KF is related to the continuous time Riccati equation.

Smoother: The smoother uses the measurement data information up to time t_m, which is greater than time t. Therefore, the immeasurable information can be smoothed and the result should be more accurate.

Predictor: The predictor uses the measurement data information up to time t_m to forecast future information at time t.

These can be summarised as

- filter, when $t = t_m$
- smoother, when $t < t_m$
- predictor, when $t > t_m$

The KF algorithm was shown in Algorithm 2.1.

Based on the simple KF formula, other filters were developed to give an equivalent implementation, such as the information filter, and the square root filter. The former, utilizing the Matrix Inversion Lemma, uses an information matrix instead of the inverse covariance matrix and makes it computationally simpler and solvable for specific classes of problems. The latter, first proposed by Potter [13], decomposes the error covariance matrix into the square root form. Anderson [4] made some extensions on decomposing the covariance update formula. Morf and Kailath [110] also undertook some extensions on calculating the filter gain with different formulations.

2.3.2 EKF, UKF and Other Filters

Since many real systems are nonlinear, a transform of the KF was developed, called the extended Kalman filter (EKF). The main idea of the EKF is that it linearizes and approximates the nonlinear term or function by using Taylor expansion. The

EKF usually means the first order case. There are other orders of EKF in the literature [75] [134]. Nevertheless, an obvious disadvantage is the expensive computation cost, especially when the order of Taylor expansion is high. It requires a huge number of calculations to compute the covariance and gain of the EKF. Furthermore, the EKF will diverge when the initial guess is wrong, while KF will rectify the initial guess. Therefore, high order EKF is rarely applied. In addition to EKF, there are many nonlinear filters including optimal nonlinear filters and suboptimal nonlinear filters. Grid-based methods were one of the optimal nonlinear filters. The finite difference method (FDM), the finite element method (FEM), and so forth, belong to grid-based methods. However, the disadvantages are that these methods must guarantee the number of finite states, and the computation cost increases with the number of states. Beneš and Daum filters are also optimal nonlinear filters. The Daum filter was the upgraded version of the Beneš filter. Both are used for finite dimensions. Schmidt [137] implemented the Daum filter as an application.

The unscented Kalman filter (UKF) was a representative of suboptimal nonlinear filters. The difference between EKF and UKF is that EKF approximates the nonlinear function, while UKF approximates posterior probability density. UKF uses statistical linearization and the formulae and the derivation were illustrated in [130]. UKF seemed to be more robust than EKF [79], however, Gustafsson and Hendeby showed that the performance of second order EKF was better than that of UKF [60]. Several variants of UKF exist. For instance, Julier and Uhlmann [80] modified the traditional unscented transform (UT) into the scaled UT.

The MM filter, as a type of Gaussian sum filter, belonged to suboptimal nonlinear filter and will be introduced in the next chapter. PF [125] have been popular recently. PF, as a kind of suboptimal nonlinear filter, was the extension of sequential Monte Carlo (SMC). They calculated the probability densities of particles. Hammersley and Morton [61] first proposed the idea of sequential importance sampling, which is the core idea of SMC. The ideas were developed in the literature [62] and [1]. However, because of computation, they were not acceptable until Gordon et al. [54] proposed the resampling step. Since then, PF has developed rapidly. However, the disadvantage is that PF needs a sufficiently large number of particles and a very good estimating state of the past. Other commonly used filters include the fixed coefficient filter (also called the α - β filter), which was the first target tracking algorithm. It is simple, but not accurate.

2.4 Image Processing

Image processing usually means digital image processing which is a digital image processed by computer. According to Gonzalez and Woods [53], there are three kinds of image processing, low level, mid level, and high level. Low level image processing is usually called pre-processing. It mainly focuses on noise reducing and contrast increasing. The characteristic of low level image processing is that both the input and output are images. Mid level image processing focuses on special tasks, such as segmentation. The characteristic is that though the input is image, the output is special features from the initial image input. High level image processing tries to make the image or the features understood by humans after processing.

Image processing stemmed from newspaper industry in the 1920s. Pictures were encoded and transmitted by submarine cable then decoded at the receiving end, which reduced transmission time considerably. Printing methods, such as selecting the distribution of gray level, appeared to improve the quality of digital images. By late 1921, a new technique, perforated tapes, had replaced traditional recovering technique. Later and to current time, with the development of computers and data technology, such as storage and display, image processing has developed rapidly. The application of image processing is extensive and in many fields. It has been used in astronavigation area since the 1960s. One famous example was the image of the moon taken from *Ranger* 7. Image processing has also been used in medicine from 45 years ago. An example is computerized tomography (CT), which was invented during that time.

2.4.1 Segmentation

Segmentation is a process which partitions an image into several regions or objects to be analysed easily. It is usually divided into several parts according to the results. Some of the results are contours from using edge detection methods. Some are intensity from using threshold. Others are colours. There are a huge number of methods, which have not been introduced in this current work, with the exception of edge detection and threshold.

Edge Detection

Usually, there are three types of edges: step edge, ramp edge and roof edge. The first is the edge where intensity changes from 1 to 0 (or from 0 to 1) ideally and suddenly at the distance of one pixel. A ramp edge is one where intensity changes from 1 to 0 (or from 0 to 1) slowly and where distance is a set more than one pixel.

It appears in the real situation because of the noise in the image. Therefore, the edge could be any point in the set. A roof edge is one which is used to detect lines across a region and when intensity changes from 1 to 0 to 1 again (or from 0 to 1 to 0 again) also at a distance of one pixel. Gonzalez and Woods [53] undertook an experiment on a ramp edge to illustrate the importance and the sensitivity to noise of the first and second order derivative of ramp edges. Based on the results, smoothing the image to reduce the noise is highly significant.

The basic edge detection method is based on the first order derivative. The first order is also understood as gradient. Usually, a two-dimensional mask is needed in a diagonal edge situation. Roberts [131] firstly used Roberts cross-gradient operators as a two-dimensional mask to deal with edge detection. Prewitt [123] made a modification by using a three by three mask, adding centre point in and filtered for the whole image with Roberts operators. In the same year, Sobel [142] also made modifications based on Roberts operators. He multiplied a number 2 in the centre location as a parameter. The advantage of Prewitt masks is that they are easy and simple, both in expression and in implementation, compared with Sobel masks. Yet, Sobel masks have the ability to smooth the noise based on the formulae. However, the advantage of Prewitt masks is so minimal that it can be ignored. Smoothing and thresholding methods were introduced to reduce the noise and obtain a better result [53].

Senior edge detection methods were based on the second order derivative and considered the image noise. The earliest senior edge detection, proposed by Marr and Hildreth [101], had two significant features. Firstly, a differential operator was needed to calculate the discrete time first order and second order derivative pixel by pixel. Secondly, large and small operators for scale invariant were required. A large operator was used for detecting fuzzy edges and a small one for sharp edges. Based on these two features, Laplacian of Gaussian (also expressed as $\nabla^2 G$) was selected as the most suitable one, where ∇^2 is the Laplacian operator, and G is a Gaussian function. Due to the shape, $\nabla^2 G$ is also called Mexican hat. It is always simple and effective by using the zero-crossing approach. However, if there were not enough zeros crossing positions, another modification proposed by Huertas and Medione [71] would determine the subpixel positions. In 1986, Canny [29] proposed another senior edge detection method called the Canny edge detector. The core idea was to determine the optimal solution under the given conditions. However, there is no such a solution. Therefore, a Gaussian noise was added in the one dimensional step edge to calculate the numerical solution, which is the first order derivative of Gaussian. For a two-dimensional edge, there are four steps to calculate the Canny edge detector. Firstly, reduce the input noise by using

the Gaussian filter. Secondly, calculate the magnitude and angle of the gradient. Thirdly, determine the local maxima by using nonmaxima suppression. Finally, use the double threshold method to obtain the edge.

Threshold

Thresholding is a method to separate the objects from the background based on an image's intensity distribution. A simple iterative threshold method has been widely and effectively used, introduced in Algorithm 2.2, where ΔT is the predetermined parameter which controls the speed of the algorithm. Considering the number of iterations, some modifications were made, such as Newton's method. A commonly used threshold algorithm based on the probability, is Otsu's method [115]. The advantage of this method is that it uses probability to calculate the optimal solution, which divides the image into two parts. Moreover, it transfers the problem into probability on a histogram, which is one dimensional variable. However, according to the information above, the noise cannot be reduced by threshold and it would affect the result of thresholding. Therefore the best option is to smooth the input image first, then to threshold the image. Sometimes, even if the input image is smoothed before thresholding, the result can still not be obtained. For example, in the case of a very tiny cobblestone in a large black background, the result was not satisfactory. Based on the example result, the modification using edge detector methods first was widely adopted. Usually, one single threshold (the methods mentioned above) is suitable, while in some cases, multiple thresholds are needed. Fukunaga proposed multiple threshold methods based on Otsu's method in [51]. However, there was no physical meaning for so many thresholds. Usually, two thresholds or double thresholds are widely used in gray level problems. More than two threshold methods are always used in colour images.

It should be noted that in the Algorithm 2.2, T_1 is the selected threshold; k is the number of iteration; N is a large number and usually it is equal to 5; the size of the image is $m * n$; I_{ji}^{k} is the intensity at the point (j, i) at k^{th} iteration; $M1$ and $M2$ are vectors which are used to save the points; $m1$ and $m2$ are the mean of $M1$ and $M2$; T_k is the threshold at k^{th} iteration; ΔT is the preinstalled error.

2.5 Descriptor

There are different kinds of shape descriptors, mainly divided into two parts. One is Contour-Based Methods, and the other is Region-Based Methods. Both can be

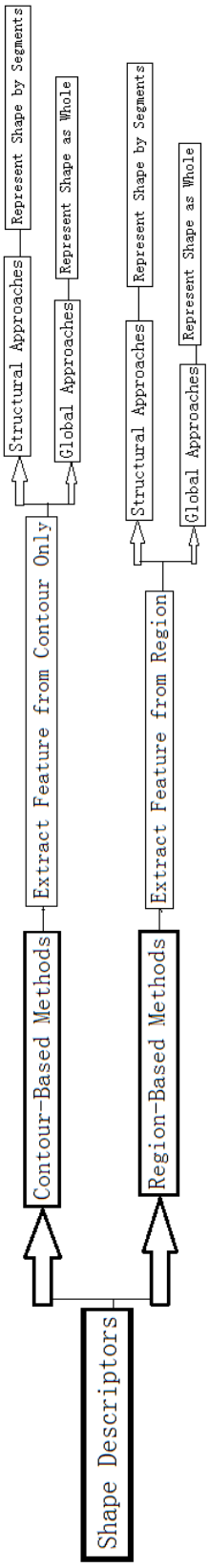

Figure 2.1: **The classification of shape description**

The shape descriptor can be divided into two methods: the contour-based methods and the region-based methods. The main characteristics are listed in the figure. For each method, there are two approaches: the structural approaches and the global approaches. The differences are also explained in the figure.

Algorithm 2.2 Simple Iterative Threshold Algorithm

Select a single threshold T_1
for $k = 2$ to N **do**
  for $j = 1$ to m **do**
    for $i = 1$ to n **do**
      if $I_{ji}^k \geqslant T_{k-1}$ **then**
        Save I_{ji}^k in $M1$
      else
        Save I_{ji}^k in $M2$
      end if
    end for
  end for
  Calculate $m1 = \bar{M}1$, $m2 = \bar{M}2$, and $T_k = \frac{m1+m2}{2}$
  if $T_k - T_{k-1} < \Delta T$ **then**
    Break FOR loop
  end if
end for

further divided into: Structural Methods and Global Methods separately, as noted in Figure 2.1.

2.5.1 Contour-Based Methods

As mentioned above, Contour-Based Methods extract shape features based only on the shape boundary. As shown in Figure 2.1, Global Methods and Structural Methods will be introduced separately in the following paragraphs.

Global Methods

Global methods usually use all the boundary information and shape matching in either spectral space or spatial space. Several global methods are introduced in the following paragraphs.

Simple Shape Descriptors

According to Yong et al. [160], there are several simple shape descriptors, for example, area. Peura and Iivarinen [121] added other simple shape descriptors, such as, elliptic variance. The main drawback of this kind of simple shape descriptors is that only large differences can be found, therefore, they usually filter shape fault separately or together.

Its advantage is simplicity while the disadvantage is inaccuracy. Therefore, this kind of method usually works with other descriptors.

Correspondence-Based Shape Matching

The core idea of correspondence-based shape matching is utilizing the point to point method of different shapes as a measurement of shape similarity. Therefore, all the points on the boundary are feature points. This kind of methods works mainly in the space domain. Correspondence-based shape matching includes numerous methods, for instance, Hausdorff distance.

Hausdorff distance (also called Hausdorff metric or Pompeiu-Hausdorff distance) was proposed by Hausdorff in 1914. The core idea of this method is calculating the distance of two shapes. As such, it is usually used to check the measurement of shape similarity between frames. The mathematics expression of Hausdorff distance is given as

$$D_H(A,B) = max(\sup_{a\in A}\inf_{b\in B} d(a,b), \sup_{b\in B}\inf_{a\in A} d(a,b)) \tag{2.25}$$

where, $D_H(A,B)$ is the Hausdorff distance between A and B; sup is the supremum and inf is the infimum; $d(a,b)$ is the Euclidean distance between any point a in the set A and any point b in the set B.

One shortcoming of this method is that if there is a noise on the boundary or an outlier point, the distance measured is not accurate. Thus, Rucklidge [132] made a modification by changing the expression of Hausdorff distance. The modified formula is as following

$$\begin{aligned} D_H(A,B) &= max(D(A,B), D(B,A)) \\ D(A,B) &= f^{th}_{a\in A} min_{b\in B} d(a,b) \end{aligned} \tag{2.26}$$

where, f^{th} is a th power to the value of f where f is a value in the area $[0,1]$ and usually $f = 0.5$ [30].

Both methods above share the advantage that they can match point-to-point. Nevertheless, their disadvantage is that neither has the property of translation, scale or rotation invariance, which are significant to shape descriptors when matching shapes.

Chamfer distance transform was proposed by Chetverikov and Khenokh [30]. The contribution of [30] is that this method reduced the computation complexity. An example was given in [132] to demonstrate the expensive cost of the computation complexity is despite its good effect.

The new development was a new method called shape contexts, proposed by Belongie et al. to overcome the limitation of Hausdorff distance [15]. The idea of shape contexts is that a vector is set up for any point on the boundary as the starting point, with all other points on the boundary as the terminal points. Based on the length and angle, the histogram for that point is created. All the histograms

of all the points on the boundary make up the shape contexts. Usually, log-polar space is used to distinguish the points around and points far away.

Generally, correspondence-based shape matching is suitable for offline calculation, not for online shape matching.

Shape Signature

Shape signature methods are based on the boundary points and usually have the invariant property of translation and scale. Many methods exist, including centrical profile [37] and centroid distance. These methods are not rotation invariant, therefore, another technique, shift matching, is needed. However, this is too expensive for online matching. Secondly, these methods are not robust. For example, if there is a noise, the shape matching will lead to error.

Boundary Moments

Boundary moments are improvements on shape signature. According to [144], the n^{th} moment M_n and the n^{th} central moment MC_n can be calculated by the following formulae, given shape signature $m(i)$.

$$\begin{aligned} M_n &= \frac{1}{N}\sum_{i=1}^{N}[m(i)]^n \\ MC_n &= \frac{1}{N}\sum_{i=1}^{N}[m(i) - M_1]^n \end{aligned} \tag{2.27}$$

where N is the number of points on the boundary; M_1 is the first order moment that is the centroid. To make the boundary moments translation, rotation, and scale invariant, the formulae above can be normalized as:

$$\begin{aligned} \bar{M}_n &= \frac{M_n}{MC_2^{\frac{n}{2}}} \\ \bar{MC}_n &= \frac{MC_n}{MC_2^{\frac{n}{2}}} \end{aligned} \tag{2.28}$$

Generally, this method is easy to implement and robust, however, it is also extremely difficult to explain the higher order moments in physics.

Elastic Matching

This method was proposed by Bimbo and Pala [17]. They set up a parameter energy function and minimized it by a neural network method.

Owing to the expensive computation and matching complexity, this method is not popular for online implementation. A comparative example was given in [164] to show the expensive computation complexity.

Stochastic Method

The core idea of stochastic method is that for an autoregressive (AR) model,

the parameters are calculated by the least square (LS) method [40]. Since those parameters have the property of translation, rotation, and scale invariance, they are considered as shape descriptors. Disadvantages also exist. For the complex shape boundary, the parameters needed are considerable. Furthermore, most of the parameters do not have physical meanings.

Scale Space Method

Scale space method is also a spatial domain method, like most of the techniques outlined above. Given its insensitivity to noise, this method has been implemented in some applications. The core idea of this method is using a variable width low-pass Gaussian filter to those boundary points. Different widths of Gaussian filter have different effects. A contrastive example [164] was given to illustrate the effect of the width, that is, the Gaussian filter smoothed the shape boundary as width increases. "Fingerprint" or an interval tree was formed as the result of smoothing. The interval tree was first explained by Asada and Brady [7]. They used the Gaussian filter and the second derivatives of Gaussian filter methods to get the interval tree. Mokhtarian and Mackworth [108] made some extension based on [7]. Curvature scale space (CSS) was first called and obtained from the scale space signature by their extension. This method was used because of its compact characteristics, although the implementation and matching is difficult. The group of Mokhtarian et al. made some extensions on the shape retrieval based on the CSS method. The new matching algorithm to compare two peaks is their innovation though it is unstable according to Jeannin [77]. Eberly [43] then proposed a geodesic distance measurement. Geodesic topology as another explanation of scale space method was first proposed by Daoudi and Matusiak [36] who changed the matching process with the geodesic distance measurement. If the shape boundary is complex or contains many points, the result will be a very high interval tree.

Spectral Transform

Spectral transform is widely used because the methods are noise insensitive and do not consider the shape boundary complexity. The FD and the wavelet descriptor (WD) are most well known spectral transform methods.

Traditional FD can only solve a closed curve. Mitchell and Grogan [107] and Lin and Chellappa [97] made modifications, which solved partial shapes using FD. Arbter [6] proposed the affine-invariant FD and K.Arbter et al. [84] showed the application of the method on 3-D objects. A modification called short-time Fourier descriptor (SFD) was proposed by Eichmann et al. [45]. The advantage of SFD is that it can obtain accurate local boundary information. For a disjointed or articulated contour shape, Rauber [126] introduced a new modification, called

UNL FD (UNL is the name of university). Its weak point is that the high dimension vector of UNL FD leads to computation complexity. Rui et al. [133] modified the distance measurement of classifying similarity by using Fourier coefficients. In 2001, Zhang and Lu [165] determined the disadvantage of SFD, namely, that it cannot capture global boundary information. Therefore, Zhang and Lu believed the traditional FD is better for global boundary information.

There are several advantages to FD as follows. Firstly, the computation of FD is very simple. Secondly, specific physical meaning of FD exists. Thirdly, it is very simple to normalise the FD. However, the FD does not need to be normalised for all cases. For example, the LFD $\boldsymbol{F}(\boldsymbol{w} = \mathbf{2})$ uses the unnormalised value. Finally, the FD can deal with both global and local shape boundary information.

Compared with other global descriptors, FD is noise insensitive, easily normalised, together with being well understood and implemented. Therefore, FD is popular in global descriptors.

Compared with FD, WD can deal with not only spectral space but also spatial space. However, Tieng and Boles [154] gave an example to show the disadvantage of WD, which is that it is not as accurate as FD in frequency resolution. Another disadvantage of WD is that its impractical online shape retrieval was caused by matching complexity, as noted by Yang et al. [159].

Structural Methods

Primitives were usually used in these methods. Primitives were defined as shape boundary segments. According to the different selections of primitives, several common decompositions exist. Pavlidis [117] gave some details of the decomposition methods, such as curve fitting.

A general form for this kind of method exists:

$$\boldsymbol{S} = [s_1, s_2, \ldots, s_n] \tag{2.29}$$

where s_i has different meanings based on different methods. For example, in B-spline method, s_i represents spline. $\boldsymbol{S}$ is a vector which is usually used directly or combined with other higher level method.

Chain Code Representation

Freeman [49] first proposed a method using line segments to encode arbitrary geometric configurations in 1961. The key idea of this method is the shape boundary discretization. Using the approximation points instead of the real boundary points and choosing any selected point as a starting point, chain code is generated. The disadvantage of the normalized chain code is scale variant. An example

in the literature [164] was given for two same sized boundary shapes with different shape numbers. The disadvantage of this kind of method is that chain code method is not noise insensitive and has high dimensions. Therefore, it is usually combined with other methods. Iivarinen and Visa [72] made some modifications known as chain code histogram (CCH). This method used probability and solved scale invariance compared with the traditional chain code method. However, it is not rotation invariant. Based on the above reasons, another method was proposed as a modification. The new method is called normalized CCH (NCCH). It recalculated the probability and reduced the dimensions. However, this method is still noise sensitive.

Polygon Decomposition

The core idea of this kind of method is to consider a shape boundary as a polygon, by using line segments [57], with vertices that contain feature information. Usually four elements, which are x and y coordinates, distance and internal angle, are used for every vertex. The literature [164] gave a fixed number for shape vertices based on efficient and robust reasoning. Mehrotra and Gary [104] proposed a method which combined chain code and the polygon method to present an example in the paper. However, the disadvantage is that this kind of method is only positive for man-made objects [164].

Smooth Curve Decomposition

Tokens were defined by Berretti et al.. A horse shape example was given in the literature [16]. The three features are: maximum curvature, orientation and distance between two tokens. Therefore, this is a rotation variant method. However, Berretti et al. cannot avoid the disadvantage. The implementation process was given in the literature [164]. Zhang and Lu also mentioned another disadvantage which is the inefficiency. The efficiency of this method relies on the number of tokens.

Scale Space Method

Dudek and Tsotsos [41] first proposed a new method, which is different from the other structural methods. The traditional structural methods use a feature-by-feature matching, then model-by-model matching. The new method uses a model-by-model matching first, and then a scale space is formed in order to modify robustly and computation complexity is requested. However, this method is not scale invariant. Moreover, parameters as mentioned in [164] are important for the implementation and application. As such, this method is neither popular nor widely used.

Syntactic Analysis

Syntactic analysis is similar to the composition of language. Fu [50] gave an

example to show how syntactic analysis worked in the composition of language. The idea of syntactic analysis is to divide a shape into different sets of parts, just like words, which are built up by alphabets. Sonka et al. [144] gave an example to illustrate how this method works. With this method, a shape can be built up by a "word" which is a set of different "alphabets". Therefore, the matching can be simplified by using "grammar". However, the syntactic analysis method also has drawbacks. It is not used widely because its grammar is not suitable for general cases. For any special case, prior background knowledge is needed, which means different cases have different alphabets, words, and grammar.

Shape Invariants

This kind of method has sought to use something invariant under various transformations. Usually there are three directions: geometric invariants, algebraic invariants, and differential invariants.

References as per the literature [70] [92] [144] [145] gave examples of three shape invariants. The difference is that when the shape can be divided into straight lines or algebraic curves, the first and second method can work well; in other cases, the differential invariants can work.

The disadvantages of shape invariants are as follows. Usually, they are derived from boundary geometric transformations while the real shape changes do not follow the boundary geometric transformations. In addition, like many other methods mentioned above, they are not noise insensitive so small noise turbulence can make the method inaccurate. Furthermore, it is extremely difficult to create new descriptors of shape invariants. Finally, they are, like above methods, extremely difficult in respect of matching shapes.

Recently, as for the previous disadvantage, the literature [92] and [145] made modifications. They used subgraph matching to determine an acceptable result in an acceptable time instead of the best solution. This is also called a "suboptimal solution".

Invariant signature, first proposed by Kliot and Rivlin, was another shape invariant descriptor [88]. The basis of this method is using invariants to derive invariant signatures, which were also named multi-valued signature, and to use signatures to make up a matrix for matching. An example was given to test and illustrate the efficiency in the literature [164]. Squire and Caelli [145] gave another invariant signature. They used probability density function as an invariant signature based on simple transformations, such as scale and rotation. A histogram was also used to match. However, the results of this method did not seem as effective as shape signatures.

2.5.2 Region-Based Methods

These methods are different from the contour-based methods. The difference is that all the pixels are considered in region-based methods, while only shape boundary information is considered in contour-based methods. Moments are usually used in region-based methods. This kind of method can be divided into two areas: global methods and structural methods. Both of these will be outlined separately.

Global Methods

As contour-based methods, global methods also consider the shape as a whole. However, the difference is that these methods are robust. Some global methods will be introduced as follows.

Geometric Moment Invariants

Hu first proposed two-dimensional shape recognition using moment invariants in [69]. The idea stems from the theory of algebraic forms. The formula was given as follows:

$$M_{ij} = \sum_{s}\sum_{t} s^i t^j f(s,t) \tag{2.30}$$

where $i, j = 0, 1, 2, \ldots$

The moments were combined and formed a set called a geometric moment. Due to invariant property (translation, rotation, and scale), this method was also named moment invariants. However, this method can only consider lower order moments. Higher order moments were extremely difficult to derive and not addressed. Kennedy et al. [86] made modifications with a normalised implementation called score normalization. A test was done to confirm the conclusion that geometric moment invariants perform better in simple shapes.

Algebraic Moment Invariants

Taubin and Cooper [147] first proposed this method. The algebraic moment invariants were calculated from the central moments. The advantage of this method is that algebraic moment invariants are useful to lower order moments and higher ones. In addition, algebraic moment invariants are useful to affine transformations. However, there are disadvantages with the main one being unstable performance. References in the literature [136] gave comparable results to show sometimes it worked well but other times it was poor. The conclusion is that the configuration of the outline is highly significant to this method.

Orthogonal Moments

The advantage of this kind of method is that it is easy and robust. The disadvantage is that there are no physical features corresponding to higher order moments.

Based on the idea of Teague [149], Legendre moments and Zernike moments belong to orthogonal moments.

The formula of Legendre moments is as follows:

$$L_{mn} = \frac{(2m+1)(2n+1)}{4} \sum_x \sum_y P_m(x) P_n(y) f(x,y) \tag{2.31}$$

where, $P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2-1)^n$

As the forms of the moments were the same as the orthogonal polynomials, which are called Legendre polynomials, these orthogonal moments were named Legendre moments.

The formula of Zernike moments is as follows:

$$\begin{aligned} Z_{mn} &= \frac{n+1}{pi} \sum_x \sum_y S^*_{nm}(x,y) f(x,y) \\ &\qquad S.T. x^2 + y^2 \leq 1 \end{aligned} \tag{2.32}$$

$$\begin{aligned} S_{nm}(x,y) &= S_{nm}(\rho\cos\theta, \rho\sin\theta) \\ &= R_{nm}(\rho) exp(jm\theta) \\ R_{nm}(\rho) &= \sum_{k=0}^{\frac{n-|m|}{2}} (-1)^k \frac{(n-k)!}{k!(\frac{n+|m|}{2}-k)!(\frac{n-|m|}{2}-k)!} \rho^{n-2k} \end{aligned} \tag{2.33}$$

where ρ is the radius; θ is the angle.

The disadvantages are as follows. Firstly, the calculation of the kernel of Zernike moments is difficult and the shape needs to be normalized. Secondly, the radial features are in spatial domain, while the circular features are in spectral domain. Finally, the circular features sometimes lost important features when calculating each order.

Teh and Chin [150] compared some famous orthogonal moments and obtained results. The most significant contribution by Teh and Chin is that the error of orthogonal moments decreases and then increases. This means that higher order moments usually have larger noise error. They also gave a criterion, called SNR, to select the number of moments. They believed that the Zernike moments and pseudo-Zernike moments are better than the other orthogonal moments. Based on their work, Liao and Pawlak [94] made extensions focussing on accuracy and efficiency. It was determined that accuracy has a relationship with image resolution. For example, a finer image will have less accurate moments. The alternative,

which extended Simpson's rule, was used to increase the speed of calculation as their last contribution.

Generic Fourier Descriptor

This method, which mainly made modifications to compensate for the disadvantages of Zernike moments, was first proposed by Zhang and Lu [163]. It is easy to compute. Moreover, the features are all in spectral domain. Furthermore, for the multi-resolution case, it has good retrieval performance.

Grid Based Method

This method was first proposed by Lu and Sajjanhar [99]. Its idea is to use a binary vector to show the shape feature. The prerequisite is that this method must be normalized first. An example was given in the literature [164]. The advantage is that it is very easy to implement and represent. The disadvantage is that this grid based method is rotation variant. In addition, it is also noise sensitive.

Shape Matrix

Shape Matrix was first proposed by Goshtasby [55]. Taza and Suen [148] made modifications. They used a weighed shape matrix instead, based on the inconstant sampling density in the shape matrix methods. The disadvantage is that shape matrix methods are sensitive to noise.

Structural Methods

As with those mentioned in contour-based methods, these methods also divide the shape into several parts. There are several methods, which will be summarised below.

Convex Hull

When extracting shape boundary by some methods or other segmentations, it always formed some noises? which made the shape irregular such as forming a convex. The convex hull focuses on the structure of the shape and removes the convex instead of the traditional computation. It transfers the shape matching to a string or a graph matching as its most observable advantage.

Medial Axis

This method uses a region skeleton to represent shape. It eliminates the redundant information and keeps the topological information as a structure. It also transfers the shape matching to a graph matching. The disadvantage is that the computation is extremely difficult. Additionally, it is noise sensitive. A solution to its drawbacks was proposed, that is, pre-processing the shape and using polygonal approximation.

Algorithm 2.3 Pre-Processing Algorithm

for $iFrame = 1$ to N **do**
  Read in the three-dimensional raw time lapse images
  Transform those images into matrices
  Sum the matrices together
  Double threshold the obtained result (the two-dimensional matrix)
end for

2.6 Pre-Processing of Time Lapse Image

As mentioned above, the time lapse images were the raw data, and pre-processing was needed to obtain useful information via MATLAB. The process of pre-processing is described as follows. The first step was to read in the time lapse images of the same frame by MATLAB automatically. Then those images had to be transformed into matrices using the program. After, compressing those images of different heights was achieved by adding all the gray level values of different heights together. The final step was to double threshold the matrices. These steps can be summarised as Algorithm 2.3.

It should be noted that the initial data used in this thesis was the data after pre-processing.

The validation was also done. All the initial three-dimensional figures were summed into two-dimension. The two-dimensional figure was compared with the figure plotted from the algorithm result to ascertain whether any neutrophil was missing. The validation result has not been shown in the thesis. In addition, the literature [129] also used the same algorithm to deal with neutrophils in pre-processing section. Therefore, the algorithm is effective.

The old data was implemented in the first year of this PhD. Currently, new data has been implemented. The old data was high resolution data, while the new data was low resolution. All the data was from zebrafish given by the University of Sheffield and legal to UK law in the Animals (Scientific Procedures) Act 1986. Neutrophils were fluorescent for the tracking experiments.

Chapter 3

Neutrophil Model Identification from Centroid Dynamics

The previous chapter introduced signal processing, image processing, the descriptors and the pre-processing. Based on the relevant literature, several methods have been considered as effective tools to solve the problem. This chapter is about tracking the centroid of neutrophils through using some of the methods discussed in the previous chapter. By using the centroid information of neutrophils, some important characteristics can be obtained.

3.1 Introduction

Centroid tracking is a form of target tracking. The central idea of centroid tracking is to consider each neutrophil as a mass point and use tracking algorithms, based on the centroid information, to extract useful behavioural information. Neutrophils show different dynamic behaviours when they are in different states or modes. Tracking the centroid of the neutrophils should provide useful information to identify neutrophil behavioural characteristics.

Farina [46] offered a definition of classical tracking. It is a set of algorithms with the following characteristics after detections that can: recognise the pattern of target; estimate parameters; extrapolate parameters; distinguish different targets; distinguish a false detection; refine the setting of threshold; and manage detections efficiently. Khare and Tiwary [87] gave an updated definition of tracking, which dealt with the transformed target in a sequence of frames. According to Kaawaase et al. [81], tracking can be divided into deterministic and non-deterministic types. The difference between the two types is whether the noise is ignored or not. The non-deterministic type with the noise can be considered as a stochastic process. In

most cases, the noise cannot be ignored and therefore a stochastic process framework is adopted.

In addition to tracking algorithms, the correlation of targets is important. Commonly used algorithms include the NNF and the joint probabilistic data association (JPDA) filter. The latter uses a statistical approach and considers the error distribution of tracking. It assumes that some of the candidates are the targets and the rest are false alarms.

There are several applications of target tracking in biology. Ray et al. [127] proposed a method which combined the KF and the active contour. This method guarantees that the central tracking position of leukocytes *in vivo* is accurate by reducing the location error. Rathi et al. [125] used the PF instead, which provided robust tracking. Li et al. [91] combined the adaptive IMM filter with geometric active contours technique to analyse the stem cell and the proposed method is effective for different behaviours and the changing of behaviours.

3.2 Multiple Model Filter

The MM filter, as mentioned in the previous chapter, is a type of Gaussian sum filter. It adds the systems to a Bayesian framework. It is suitable for a target which changes its behaviours frequently.

The three models (constant acceleration model, constant velocity model, and random walk model) run the KF separately and the probability of all the three models was calculated in each frame. The "best" model (the biggest probability) was selected compared with the probability in each frame. It is a acknowledged that this "best" model can be different in different frames.

The probability of switching models can be calculated as follows.

$\boldsymbol{\mu_0^i} = [\frac{1}{3}\frac{1}{3}\frac{1}{3}]^T$ is the initial guess of the probability, where $i \in \{1, 2, 3\}$ represents the model index. Assume that at the beginning, all three models have the same probability, that is, $\frac{1}{3}$.

Firstly, start with the initial condition probability of all three models $\boldsymbol{\mu_0^i}$ and the normalization constant $\bar{c}_k^j$ can be calculated as follows.

$$\bar{c}_k^j = p\{M_k^j | y_{1:k-1}\} = \sum_{i=1}^{r} p_{ij}\mu_{k-1}^i \tag{3.1}$$

where $r = 3$ is the total number of models; $j \in \{1, 2, 3\}$ represents the model index, too; $\bar{c}_k^j$ is the normalization constant; $\boldsymbol{\mu_{k-1}^i}$ is the likelihood of model i at time $k-1$; p_{ij} is the transition probability for the Markov chain according to which the model switches from i to j. It is believed that the main diagonal values of the

transition matrix should be between 0.8 and 0.98 [10]. Therefore, in this case, it was chosen as follows:

$$\boldsymbol{P_{tr}} = \begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{pmatrix} \tag{3.2}$$

It should be noted that the elements in Equation 3.2 were selected as the following rules. According to Bar-Shalom et al., the main diagonal element should be any value in the range between 0.8 and 0.98, which causes the neutrophil to keep the same model in the next frame with a large probability and receives an acceptable peak error and root mean square error, that is, an acceptable bandwidth. If the value of the main diagonal element were less than 0.8, the peak error would be small, however, the root mean square error was considerable which led to a large bandwidth. In this case, the main diagonal element was 0.9. The other elements should be the same with the value $(1 - 0.9)/2$, unless there is an obvious reason that the neutrophil prefers some model.

Secondly, run the KF separately for all the three models and calculate the likelihood expression Λ_k^j for the neutrophils.

The mathematical model is as follows (no matter whether the chosen model is constant acceleration, constant velocity, or random walk, it has the same form. Nevertheless, the differences among those three models are the size of the matrix $\boldsymbol{F^j}$, $\boldsymbol{G^j}$, $\boldsymbol{H^j}$, and $\boldsymbol{D^j}$):

$$\boldsymbol{x_k^j} = \boldsymbol{F^j x_{k-1}^j} + \boldsymbol{G^j w_{k-1}} \tag{3.3}$$

$$\boldsymbol{y_k^j} = \boldsymbol{H^j x_k^j} + \boldsymbol{D^j v_k} \tag{3.4}$$

where the $\boldsymbol{w_k}$ and $\boldsymbol{v_k}$ are the white noise, $\boldsymbol{x_k^j}$ is the state of j^{th} model and $\boldsymbol{y_k^j}$ is the measurement of the j^{th} model.

So the likelihood expression Λ_k^j can be calculated as follows:

$$\begin{aligned} \Lambda_k^j &= p\{y_k | M_k^j, y_{1:k-1}\} \\ &= p\{y_k | M_k^j, \hat{x}_{k-1}^j, P_{k-1}^j\} \\ &= \frac{1}{\det(2\pi \boldsymbol{S_k^j})^{\frac{1}{2}}} e^{-\frac{1}{2} v_k^{j'} s_k^{j^{-1}} v_k^j} \end{aligned} \tag{3.5}$$

where $\boldsymbol{v_k^j} = \boldsymbol{y_k} - \boldsymbol{H^j F^j \hat{x}_k^j}$ is the residual of the j^{th} observer, and $\boldsymbol{S_k^j} = \boldsymbol{R^j} +$

Algorithm 3.1 Multiple Model Filter Algorithm

Initialise parameters $\boldsymbol{\mu_0^i}$ and the transition probability matrix $\boldsymbol{P_{tr}}$
for $k = 1$ to N **do**
  Calculate the normalization constant $\bar{c}_k^j$
  Run Kalman filter for three models separately and calculate the estimated state $\hat{\boldsymbol{x}}_{\boldsymbol{k}}^{\boldsymbol{j}}$ and the covariance matrix $\boldsymbol{P_k^j}$
  Calculate residual $\boldsymbol{v_k^j}$ and the variance $\boldsymbol{S_k^j}$
  Calculate the likelihood Λ_k^j
  Calculate the normalization constant c
  Calculate the probability μ_k^j
end for

$\boldsymbol{H^j[F^jP_k^jF^{j'} + Q^j]H'}$ is the variance of the j^{th} observer. Λ_k^j is the likelihood with mean $\boldsymbol{H^jF^j\hat{x}_k^j}$ and variance $\boldsymbol{S_k^j}$.

Thirdly, calculate the normalization constant c based on the above steps.

$$c = p\{y_k|y_{1:k-1}\} = \sum_{j=1}^{r} \Lambda_k^j \bar{c}_k^j \tag{3.6}$$

Finally, obtain the probability μ_k^j and repeat the steps.

$$\begin{aligned}
\mu_k^j &= p\{M_k^j|y_{1:k}\} \\
&= \frac{p\{y_k|M_k^j, y_{1:k-1}\}p\{M_k^j|y_{1:k-1}\}}{p\{y_k|y_{1:k-1}\}} \\
&= \frac{1}{c}\Lambda_k^j \sum_{i=1}^{r} p\{M_k^j|M_{k-1}^i, y_{1:k-1}\}p\{M_{k-1}^i|y_{1:k-1}\} \\
&= \frac{1}{c}\Lambda_k^j \sum_{i=1}^{r} p_{ij}\mu_{k-1}^i \\
&= \frac{1}{c}\Lambda_k^j \bar{c}_k^j
\end{aligned} \tag{3.7}$$

where μ_k^j represents the posterior probability of model j at time k.

The initial condition of the probability $\boldsymbol{\mu_0^i}$ and the transition matrix p_{ij} are important to the MM algorithm.

The transition probability describes a Markov chain process. Therefore, the different values of the elements can make some changes in the performance of the MM filter. The algorithm of MM was shown in Algorithm 3.1.

The MM estimator can be divided into the following two types according to [130]: the static MM estimator [2] and the dynamic MM estimator [11].

The IMM filter is one of the most popular dynamic MM estimators [130] based on merging. There are other dynamic MM estimators based on pruning [59].

The idea of the IMM filter was first posed by Blom [21], and became widely known via the literature [22].

The formulae of IMM are nearly the same as MM. The difference is that IMM does not need to calculate the normalization constant. Instead of the normalization constant, IMM calculates the mixing probabilities $\iota_{k-1}^{i|j}$ which was shown as follows.

$$\begin{aligned} \iota_{k-1}^{i|j} &= p\{M_{k-1}^{i}|M_{k}^{j}, y_{1:k-1}\} \\ &= \frac{p_{ij}\mu_{k-1}^{i}}{\sum_{i=1}^{r} p_{ij}\mu_{k-1}^{i}} \end{aligned} \tag{3.8}$$

Therefore, the update equations were shown as follows.

$$\hat{\boldsymbol{x}}_{\boldsymbol{k-1}}^{\boldsymbol{j}} = \sum_{i=1}^{r} \iota_{k-1}^{i|j} \hat{\boldsymbol{x}}_{\boldsymbol{k-1}}^{\boldsymbol{i}} \tag{3.9}$$

$$\boldsymbol{P}_{\boldsymbol{k-1}}^{\boldsymbol{j}} = \sum_{i=1}^{r} \iota_{k-1}^{i|j} (\boldsymbol{P}_{\boldsymbol{k-1}}^{\boldsymbol{i}} + (\hat{\boldsymbol{x}}_{\boldsymbol{k-1}}^{\boldsymbol{i}} - \hat{\boldsymbol{x}}_{\boldsymbol{k-1}}^{\boldsymbol{j}})(\hat{\boldsymbol{x}}_{\boldsymbol{k-1}}^{\boldsymbol{i}} - \hat{\boldsymbol{x}}_{\boldsymbol{k-1}}^{\boldsymbol{j}})^{T}) \tag{3.10}$$

The mode probabilities μ_k^j were, therefore, updated as follows.

$$\mu_{k}^{j} = \frac{\Lambda_{k}^{j} \sum_{i=1}^{r} p_{ij}\mu_{k-1}^{i}}{\sum_{j=1}^{r} \Lambda_{k}^{j} \sum_{i=1}^{r} p_{ij}\mu_{k-1}^{i}} \tag{3.11}$$

The IMM filter is extensively used in tracking maneuvering targets, because of its "self-adjusting variable-bandwidth" ability. Furthermore, it has a better balance of complexity and performance in [10], [22], and [25].

Bar-Shalom and Blom [10] indicated that the advantage of the IMM is that it reduces the computation cost compared with the input estimation [25]. Blom and Bar-Shalom [23] and Helmick et al. [65] showed the IMM smoother. However, IMM, like other filters, loses efficacy when the probabilities of mixture components do not merge [130]. In addition, when the numbers of models increased obviously, the performance of IMM decreased [22]. Based on the IMM algorithm, other algorithms, such as the Interacting Multiple Bias Model (IMBM) and the Interacting Multiple Acceleration Model (IMAM), were proposed and implemented.

3.3 Results

In this chapter, both the high resolution data and the low resolution data are implemented. The former is the data which focuses on a few neutrophils or a small part of zebrafish in high resolution images; while the latter is the data which includes the whole zebrafish and therefore the images are in low resolution. In addition, the high resolution data has a longer record time. Usually, the length of record time in high resolution data is three times that of low resolution data. However, no matter what kind of data is chosen, similar conclusions can be achieved. Therefore, most of the results in this chapter contain high resolution data.

3.3.1 Kalman Filter Implementation

In this part, the KF is implemented with a random walk model based on Algorithm 2.1. The high resolution data is applied to show the tracking effect of the KF. It should be noted that no comparison results are illustrated between the low resolution and high resolution data since both have the same conclusion, that is, the KF can be used in centroid cell tracking and have an acceptable tracking result. The centroid tracking result is shown in Figure 3.1. From the top of Figure 3.1, it is clear that the KF tracked the initial data very well both in the x and y position. A similar tracking effect is also illustrated in the centroid trajectory tracking on the bottom left of Figure 3.1. The centroid trajectory tracking is not perfect, because the mathematical model of neutrophils is not accurate and it is believed that several simple mathematical models working together will effectively express the mode of neutrophils. For this reason the MM filter will be applied later in this chapter. The absolute distance error is shown on the bottom right of Figure 3.1. It is clear that the maximum absolute distance error is less than $3.5\mu m$. The KF, therefore, is implemented in cell centroid tracking.

3.3.2 Kalman Parameters Q and R Test

Firstly, the parameters of the KF will be tested. From the updated formulae, there are two parameters, the signal noise covariance matrix $\boldsymbol{Q}$ and the observed noise covariance matrix $\boldsymbol{R}$, affecting the tracking performance. The different values of these two parameters can affect the estimated results. Therefore, selecting suitable values of these two parameters is highly significant. In the following part, only the random walk model running on the high resolution data result is represented to show the relationship between different values of $\boldsymbol{Q}$ or $\boldsymbol{R}$ and the performance. It should be noted that the results of the other two models and the low resolution

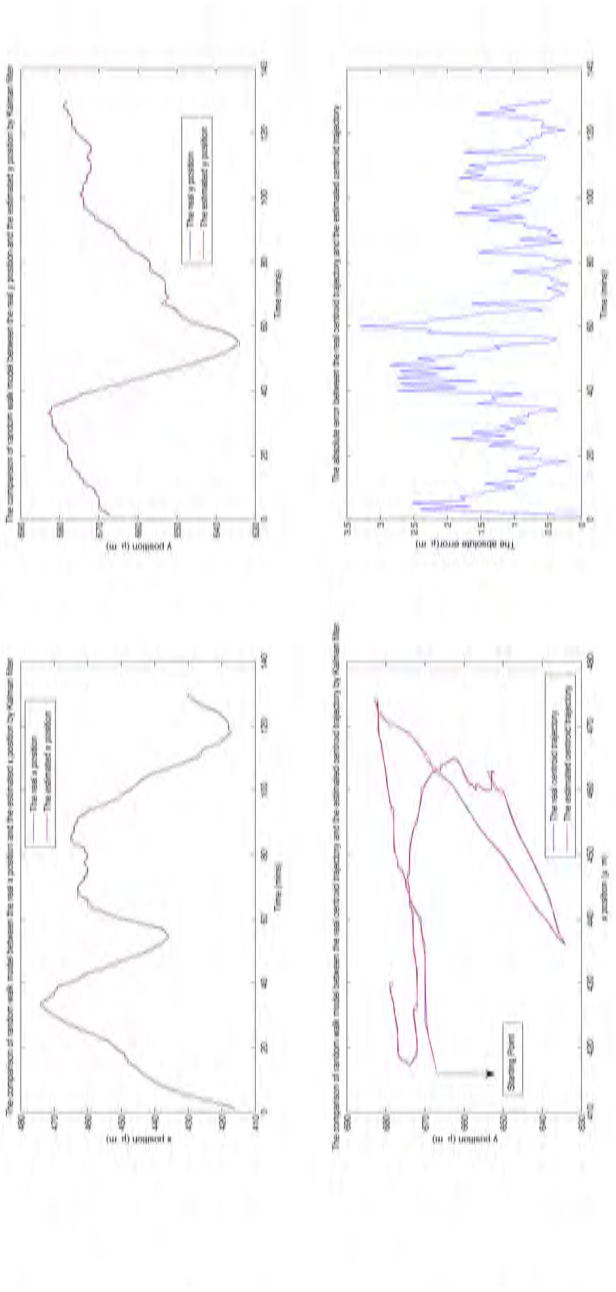

Figure 3.1: **The implementation result of the KF run by random walk model.**

Tracking results based on the x and y direction are illustrated on the top left and on the top right respectively; the centroid trajectory tracking result is illustrated on the bottom left. All the red lines are estimated trajectories by the KF and all the blue lines are real trajectories in the above three figures. The black arrow is the starting point of the trajectory as mentioned on the bottom left figure. The absolute distance error is illustrated on the bottom right figure.

data have nearly the same conclusions as the random walk model and therefore they have not been illustrated in this current work.

Q Test

The aim of Q Test is to test the tracking performance when the signal noise covariance matrix $\boldsymbol{Q}$ is changed and the other parameter, the observed noise covariance matrix $\boldsymbol{R}$, is fixed. According to the white noise assumption in the KF formula 2.20, Kalman parameter $\boldsymbol{Q}$ is a diagonal matrix. All the elements in the diagonal are assumed the same (with the value Q_d) in order to simplify the problem. In this part, different values of Q_d are tested and compared to determine any connection and relationship between the performance and the parameter. The compared tracking results are illustrated in Figure 3.2 and Figure 3.3. From Figure 3.2, it is clear that not all trajectories tracked well since all of them have the same starting point but different finishing points. Most of the tracking trajectories are nearly the same as the real centroid trajectory, except for $Q_d = 0.01$. The trajectory of $Q_d = 0.01$ is obviously inaccurate from Figure 3.2. The reason is that the parameter $\boldsymbol{Q}$ has the ability to compensate for the tracking error without considering the uncertainty of model selection. Furthermore, if Q_d is small, the states cannot adjust themselves to an effective range. For example, if $\boldsymbol{Q}$ is a scalar and $Q_d = 0.01$, the states $\boldsymbol{x_k}$ can only adjust within the square root of Q_d which is ± 0.1 to the new states $\boldsymbol{x_{k+1}}$, however, it is not enough to find the "BEST" solution of $\boldsymbol{x_{k+1}}$. In addition, the mode and module of neutrophils probably do not match the model selected. Therefore the parameter $Q_d = 0.01$ is too small to make the KF track the centroid trajectory well. Furthermore, the parameter $\boldsymbol{Q}$ does have the ability to affect the tracking result. An unsuitable value of Q_d can worsen the tracking. Figure 3.3, to some degree, shows the conclusion clearly. In most of the time range, the $Q_d = 0.01$ trajectory has a larger absolute distance error than the other values. Additionally, the absolute distance error is large and unacceptable. The $Q_d = 0.1$ trajectory has obviously a smaller amplitude of error in most of the time range than that of the $Q_d = 0.01$ trajectory. However, at some times, the error is too large to be considered acceptable. This can be also found in the tracking trajectory in Figure 3.2. Some parts of the $Q_d = 0.1$ trajectory deviate from the real centroid tracking trajectory. The other three values of Q_d are acceptable. Therefore, it is clear that the larger Q_d is, the less error the tracking has. The other two models had the same conclusion and have not been illustrated in this current work. Finally, the conclusion drawn is that for a fixed model run by the KF, a larger number should be selected as the diagonal value Q_d of parameter $\boldsymbol{Q}$ in order to compensate for the error of state tracking.

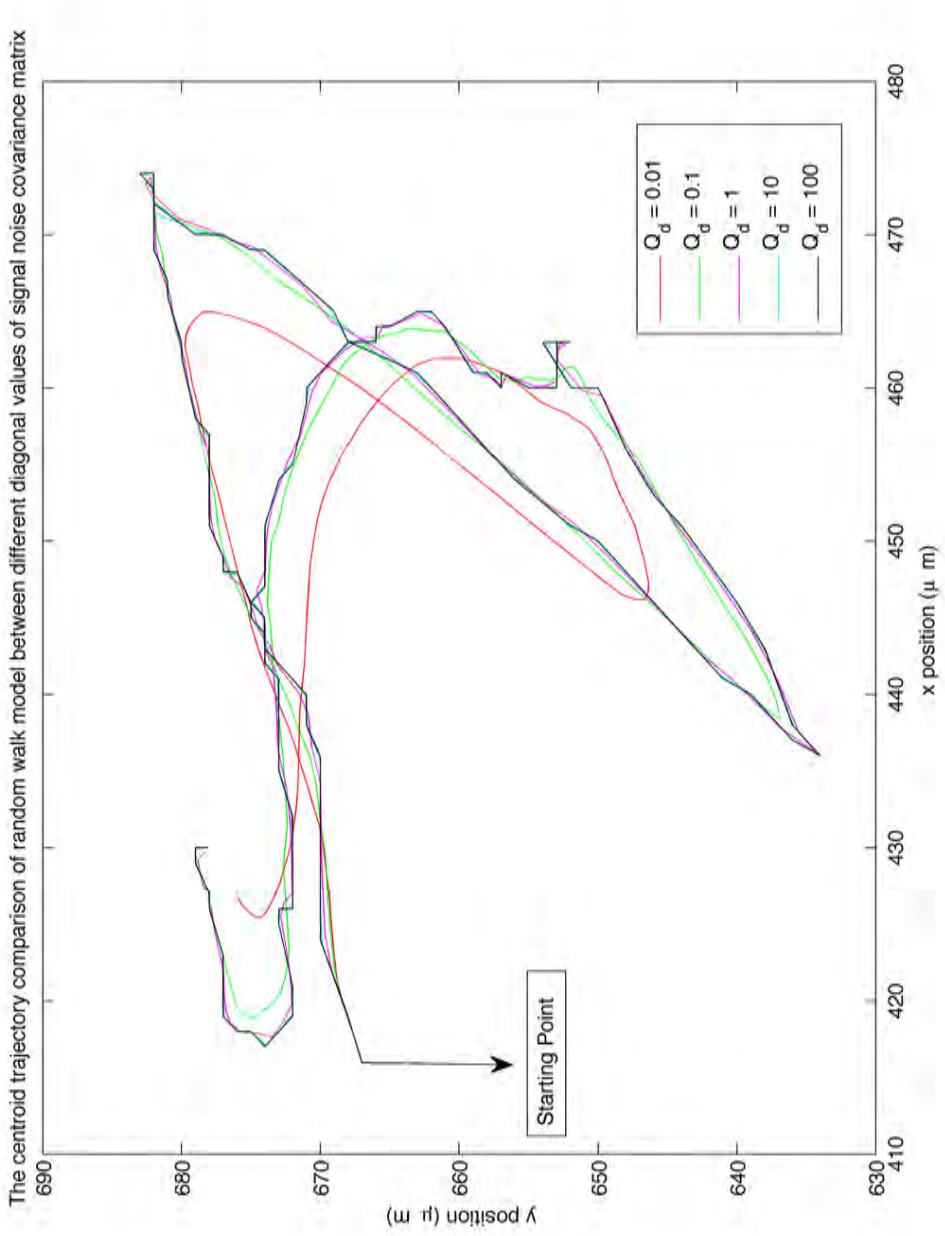

Figure 3.2: The comparison of centroid tracking trajectories between different diagonal values of the signal noise covariance matrix Q.

Q_d is the diagonal value of signal noise covariance matrix $\boldsymbol{Q}$. The signal noise covariance matrix $\boldsymbol{Q}$ is a diagonal matrix based on the white noise assumption. All the trajectories have the same starting point as pointed in the figure but different finishing points. This test is based on the random walk model and the other parameter $\boldsymbol{R}$ is fixed with the diagonal value of $R_d = 10/3$.

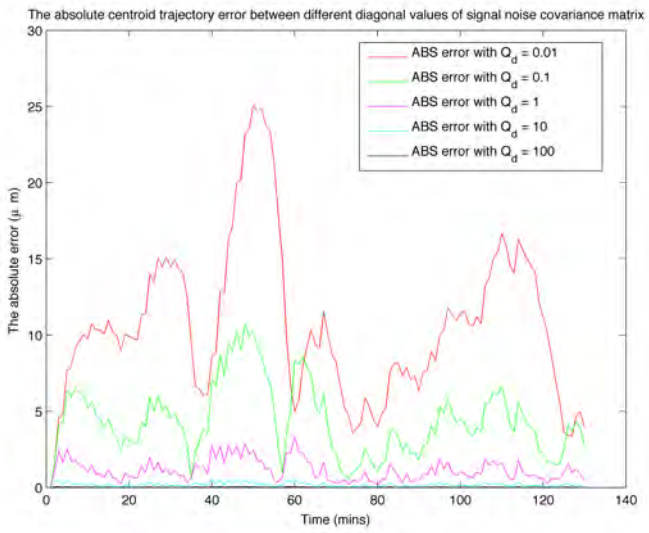

Figure 3.3: **The absolute distance errors between different diagonal values of the signal noise covariance.**

In the legend, Q_d is the diagonal value of signal noise covariance matrix $\boldsymbol{Q}$ and ABS is short for the absolute. The other parameter $\boldsymbol{R}$ and the model are the same as stated in Figure 3.2.

R Test

R Test is nearly the same as Q Test. The aim of R Test is to test the tracking performance when the observed noise covariance matrix $\boldsymbol{R}$ is changed and the other parameter, the signal noise covariance matrix $\boldsymbol{Q}$, is fixed. According to the white noise assumption in the KF formula 2.20, Kalman parameter $\boldsymbol{R}$ is also a diagonal matrix. All the elements in the diagonal are assumed the same (with the value R_d) in order to simplify the problem. In this part, different values of R_d are tested and compared to ascertain any connection and relationship between the performance and the parameter. The compared tracking results are illustrated in Figure 3.4 and Figure 3.5. From Figure 3.4, it seems that almost all the tracking trajectories are acceptable. All have the same starting point, and nearly the same finishing point. Although in some parts of the trajectory they are different, the difference is minimal. It is extremely difficult to distinguish which value of R_d is better than the others. Figure 3.5 solves the above problem and provides the conclusion. From Figure 3.5, it is clear that the absolute distance error of $R_d = 10$ trajectory is larger than the others across most of the time range. The error of $R_d = 10$ is small and acceptable most of the time range, however, in some frames, the error is large which leads to differences in tracking trajectory as per Figure

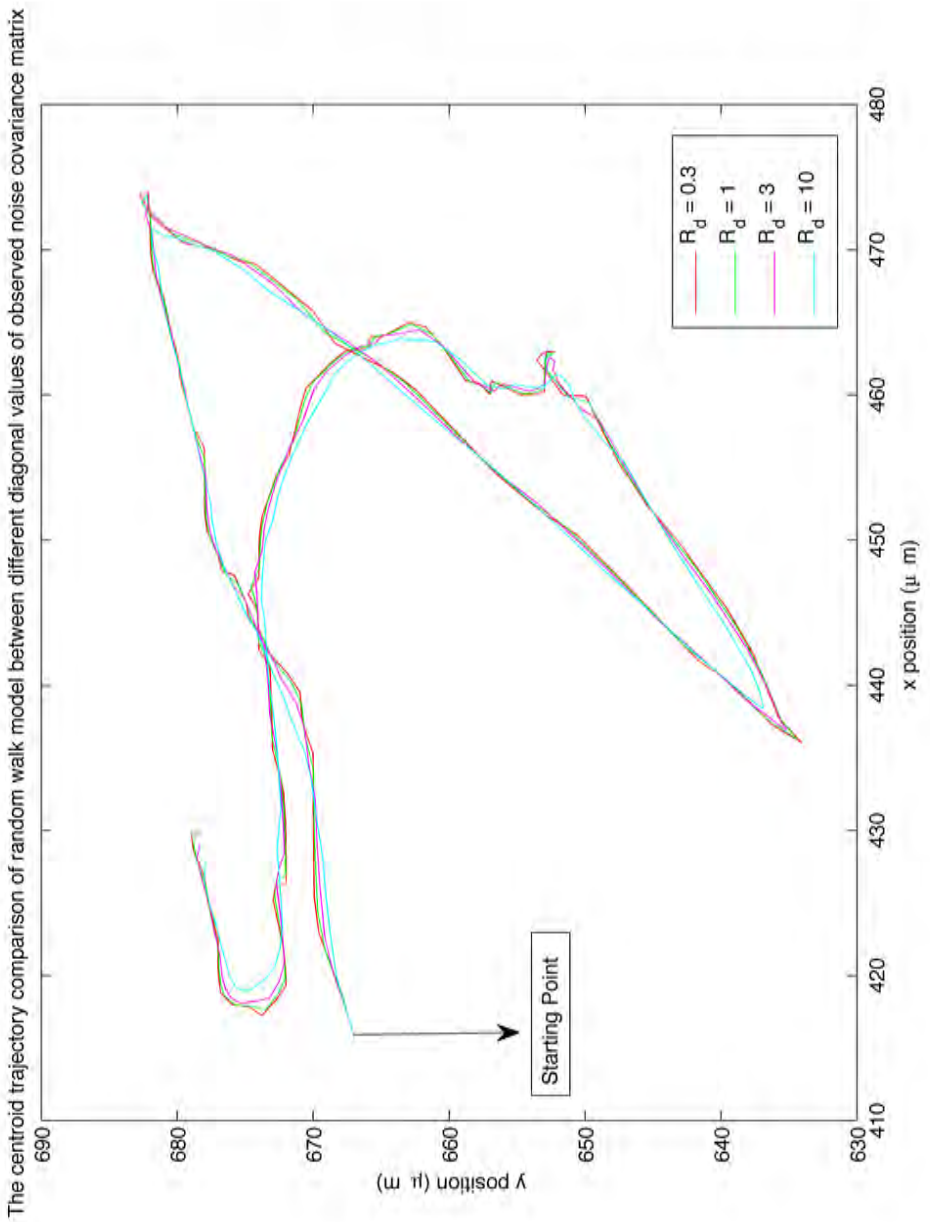

Figure 3.4: The comparison of centroid tracking trajectories between different diagonal values of the observed noise covariance matrix R.

R_d is the diagonal value of the observed noise covariance matrix $\boldsymbol{R}$. The observed noise covariance matrix $\boldsymbol{R}$ is a diagonal matrix based on the white noise assumption. All the trajectories have the same starting point as shown in the figure and nearly the same finishing point. This test is also based on the random walk model and the other parameter $\boldsymbol{Q}$ is fixed with the value of $Q_d = 100$.

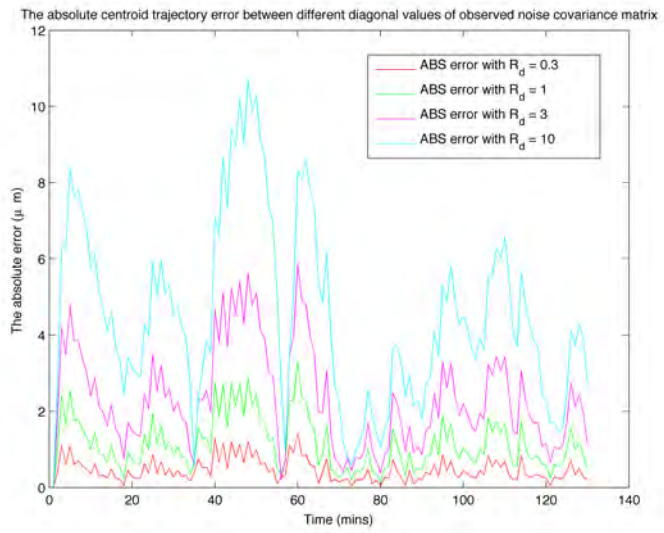

Figure 3.5: The absolute distance error between different diagonal values of signal noise covariance.

In the legend, R_d is the diagonal value of the signal noise covariance matrix $\boldsymbol{R}$ and ABS is short for the absolute. The other parameter $\boldsymbol{Q}$ and the model are the same as stated in Figure 3.4.

3.4. Since the absolute error of all the others is small enough, the parameters of the others can be acceptable and it is clear that the less value the observed noise covariance matrix is, the less error the tracking has. This conclusion can be also explained by the formula. For instance, the difference between the real output and the estimate output is in the range between the negative square root of R_d and the positive square root of R_d. If the value of R_d had been small, the difference would have been small, which would have meant the estimate output was close to the real output and the estimator was acceptable. The other two models have the same conclusion and have not been illustrated in this current work. Finally, the conclusion drawn is that for a fixed model, run by the KF, a smaller number should be selected as the diagonal value R_d of parameter $\boldsymbol{R}$ in order to reduce the tracking error.

In general, the parameters $\boldsymbol{Q}$ and $\boldsymbol{R}$ do have the ability to change the performance of the KF. If a system is to be run only by the KF, the diagonal value Q_d of the parameter $\boldsymbol{Q}$ should be chosen a little larger and the diagonal value R_d of the parameter $\boldsymbol{R}$ should be smaller. However, if the system is to be run by other filters accompanying the KF, such as the MM filter, the diagonal value Q_d of the parameter $\boldsymbol{Q}$ should be chosen smaller to compensate for any loss of efficacy and

the diagonal value R_d of the parameter $\boldsymbol{R}$ should be smaller for accuracy.

It should be noted that the error can be derived straightforwardly by using the Matrix Inversion Lemma but the derivation has not been illustrated in this current work. It is clear that in the formula, when the diagonal values of the matrix $\boldsymbol{Q}$ or $\boldsymbol{R^{-1}}$ increase, the error decreases. However, the formula is not inversely proportional to $\boldsymbol{Q}$, or $\boldsymbol{R^{-1}}$, or $\boldsymbol{QR^{-1}}$.

3.3.3 Three Physical Models Run by Kalman Filter

Three commonly used physical models are the constant acceleration model, constant velocity model, and random walk model. Usually, these models are used separately to describe the characteristics of motion. For example, the process of a plane flying from city A to city B can be considered as a constant velocity model without considering the departure and arrival. The motion of neutrophils is complex and therefore these three commonly used models are tested separately to determine which model is more suitable for the motion mode of neutrophils. The centroid tracking trajectories of all three models are compared in this part.

Figure 3.6 shows the comparison of both the centroid trajectories and the absolute distance error of the three models. At the top of Figure 3.6, all three estimated centroid trajectories are compared with the original one, which is the centroid trajectory from the data. All have the same starting point, but different finishing points. In some parts of the centroid trajectories, several models can track the real centroid well, however, in other parts, some of them could not achieve that. Therefore, it is clear that none of the three models tracks the centroid trajectory perfectly and none is suitable for neutrophil motion mode individually. This is probably because there are several modes and modules of neutrophils, which means that no single model can track perfectly. At the bottom of Figure 3.6, it is clear that none of the three models has small error all the time. In some frames, the constant acceleration model tracks better (with less absolute distance error); in some frames, the constant velocity model tracks better; in other frames, the random walk model tracks better. This is probably because neutrophils run different modes in different frames. For example, in some frames, the neutrophil runs the constant acceleration mode, and therefore, the constant acceleration model works better than the other two models in those frames. Therefore, a method that combines the three models is required and emerged. This method is called the MM filter and the advantage is that it can select the model automatically to solve the problem of the complex motion mode. The relevant results will be represented in next subsection.

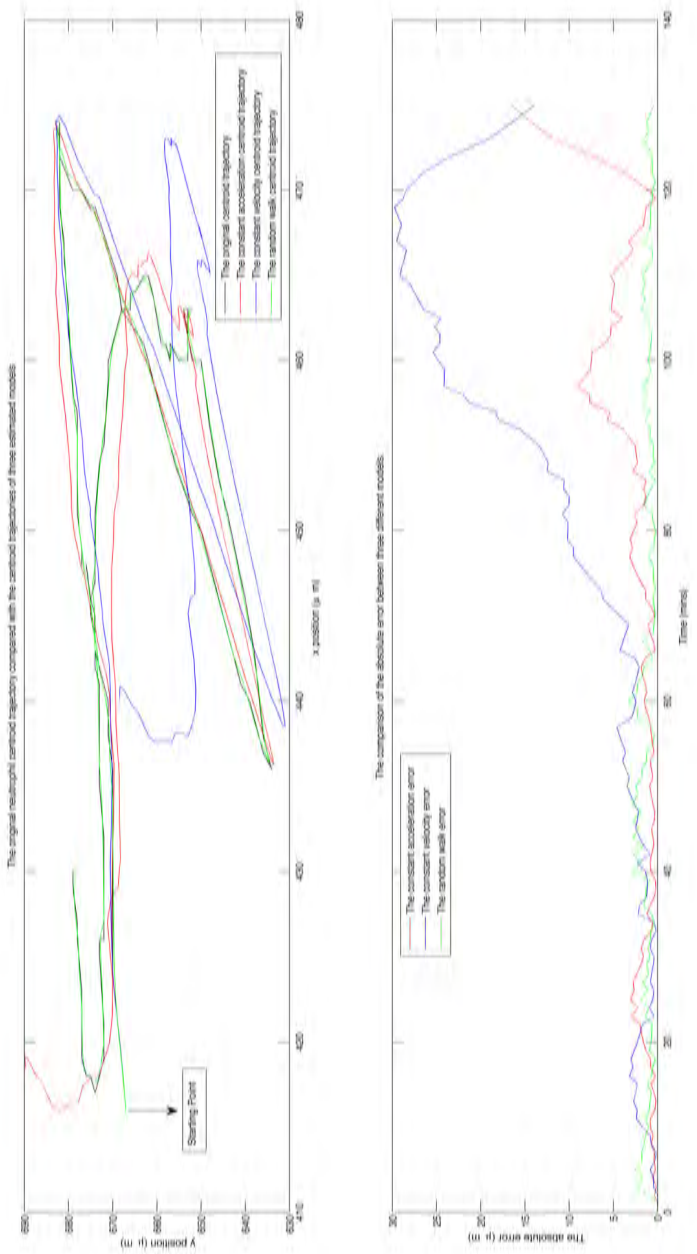

Figure 3.6: **The comparison of three models separately**

On the top figure, all the estimated centroid trajectories of three models are compared with the original one. They have the same starting point, but different trajectories and finishing points. The bottom figure illustrates the absolute distance error of three models. The red, blue and green represent the constant acceleration model, constant velocity model and random walk model respectively.

3.3.4 Multiple Model Filter Implementation

As stated above, the MM filter combines all three commonly used models and chooses the "best" model in each frame based on the probability. The details can be found in Algorithm 3.1. However, two parameters exist, that is, the initial guess $\boldsymbol{\mu_0}$, which is the probability of each model at the beginning, and the transition probability matrix $\boldsymbol{P_{tr}}$, which shows the transition probability from one model to another in the algorithm. These two parameters may affect the tracking results; therefore, they will be tested separately as detailed below.

Initial Guess μ_0

The initial guess $\boldsymbol{\mu_0}$ is the probability of each model in the first frame. According to probability knowledge, the summation of probability of all three models should be equal to 1 in each frame. In this part, four choices of the initial guess $\boldsymbol{\mu_0}$, which are $[1,0,0]$, $[0,1,0]$, $[0,0,1]$, $[\frac{1}{3},\frac{1}{3},\frac{1}{3}]$, will be tested. Figure 3.7 shows the results. In Figure 3.7, the blue, green and red represent the constant acceleration model, constant velocity model, and random walk model respectively. The neutrophils' mode changes in the frame where the black lines appear and therefore, the black lines are called switching points. In Figure 3.7, the switching points are not perfect at the crossing point. This is most likely because the probability trajectory is continuous, while the probability calculated from the program is discrete.

From Figure 3.7, it is clear that after a few frames, the dominant model is the same. The number of those frames is decided by different data, sampling time and other parameters. For instance, in the high resolution data, it is 10^{th} frame, while in the low resolution data, it is 8^{th} frame. This means that if the constant acceleration model is the chosen module in frame 20, no matter what the initial guess $\boldsymbol{\mu_0}$ is, the MM filter will choose the constant acceleration model in frame 20. As such, the initial guess $\boldsymbol{\mu_0}$ only affects the initial conditions and does not affect the tracking results after a few frames. A conclusion drawn is that no matter what values the initial guess $\boldsymbol{\mu_0}$ is, after a few frames, the module of neutrophils is decided by itself. That is, all the neutrophils should have their own module and this module does not depend on the initial guess of $\boldsymbol{\mu_0}$. The initial guess $\boldsymbol{\mu_0}$ can only affect the probability in the first few frames and after that the probability should be decided by the neutrophils' module. Interestingly, the parameter $\boldsymbol{\mu_0}$ is similar to the initial guess $\hat{\boldsymbol{x}}_0$ in the KF.

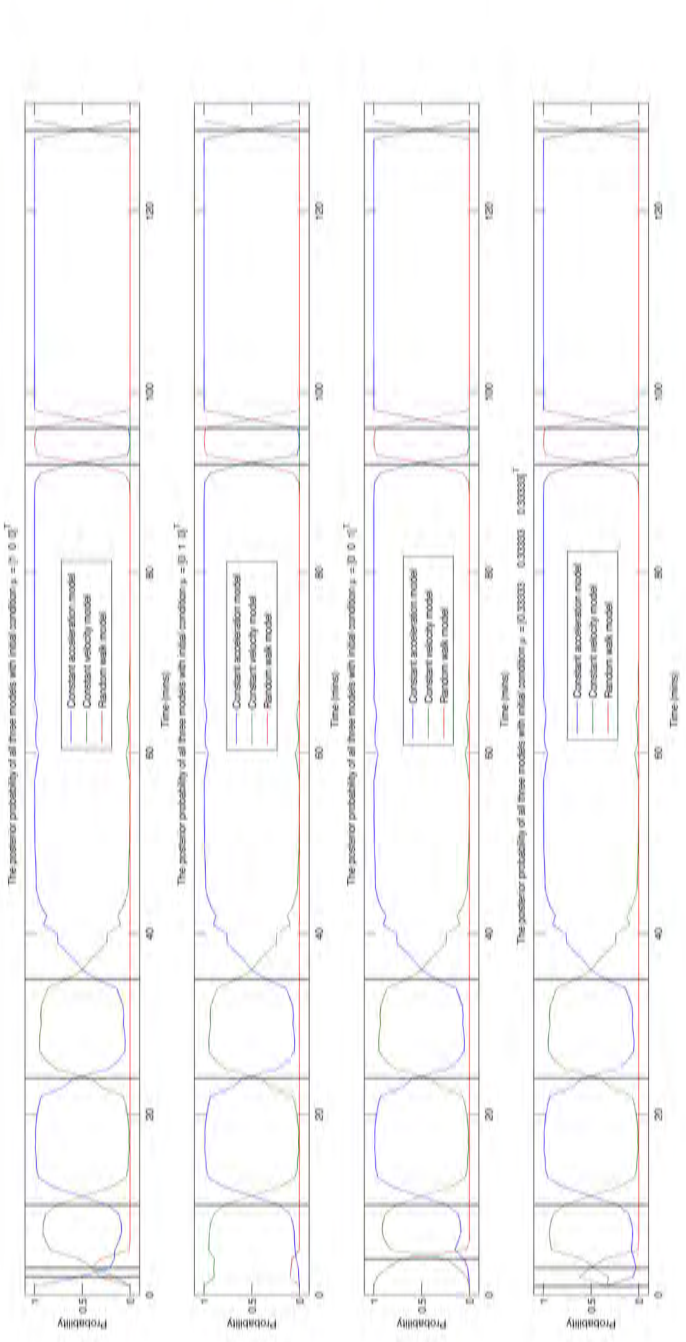

Figure 3.7: The comparison of different initial guesses μ_0

Four different initial guesses μ_0 were tested with the posterior probability of the MM filter while another parameter P_{tr} was fixed with $\begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{pmatrix}$. The black vertical lines are the frames, in which the previous dominant model changes to another. The blue, green, and red are the constant acceleration model, constant velocity model and random walk model respectively.

Transition Probability Matrix P_{tr}

Another important parameter is the transition probability matrix $\boldsymbol{P_{tr}}$. Its main diagonal values represent the probability of one model remaining the same in the next frame. For example, the first element in Formula 3.2 is 0.9, which means that if in frame k, the constant acceleration model were selected, then in frame $k+1$, the constant acceleration model would have a 90% probability of being selected again, and only a 10% probability of changing to another model. According to [10], the values of the main diagonal elements should be between 0.8 and 0.98. In this part, three different transition probability matrices, which are $\begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{pmatrix}$, $\begin{pmatrix} 0.9 & 0.01 & 0.09 \\ 0.01 & 0.9 & 0.09 \\ 0.05 & 0.05 & 0.9 \end{pmatrix}$, and $\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$, are implemented and compared. The result is shown in Figure 3.8.

In Figure 3.8, the blue, green and red are the constant acceleration model, constant velocity model and random walk model respectively. The black lines are the switching points. From Figure 3.8, it is clear that different transition probability matrices $\boldsymbol{P_{tr}}$ have different posterior probability trajectories and different switching points. Therefore, the transition probability matrix $\boldsymbol{P_{tr}}$ can change the neutrophils' modules. The bottom probability trajectory (whose transition probability matrix is $\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$) in Figure 3.8 seems unstable. It is clear that in most of the frames, the probability of the dominant model is not close to 1. This is probably because all the elements in the transition probability matrix $\boldsymbol{P_{tr}}$ are the same ($\frac{1}{3}$), which gives all the models the same probability of changing from one to another, rather than keeping a large probability in the same model. Furthermore, the main diagonal elements follow Bar-Shalom's idea. Therefore, this matrix is not suitable for the transition probability matrix. The probability trajectory in the middle (whose transition probability matrix is $\begin{pmatrix} 0.9 & 0.01 & 0.09 \\ 0.01 & 0.9 & 0.09 \\ 0.05 & 0.05 & 0.9 \end{pmatrix}$) seems stable. However, the elements are not suitable. For example, the second and third elements in the first row are 0.01 and 0.09, which means that in frame k the constant acceleration model is selected. However, if in frame $k+1$ the module changes, either the constant velocity model is selected by the probability of 1% or the random walk model is selected by the probability of 9%. It is unreasonable that the random walk model should have a greater probability of being chosen. In addition, there is no information shown to prove that the random walk model is preferred by neutrophils. Therefore, this matrix is deemed not suitable for the transition probability matrix, either. Given the above reasons, the top probability trajectory (whose transition probability matrix is $\begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{pmatrix}$) is selected as the transition probability matrix $\boldsymbol{P_{tr}}$.

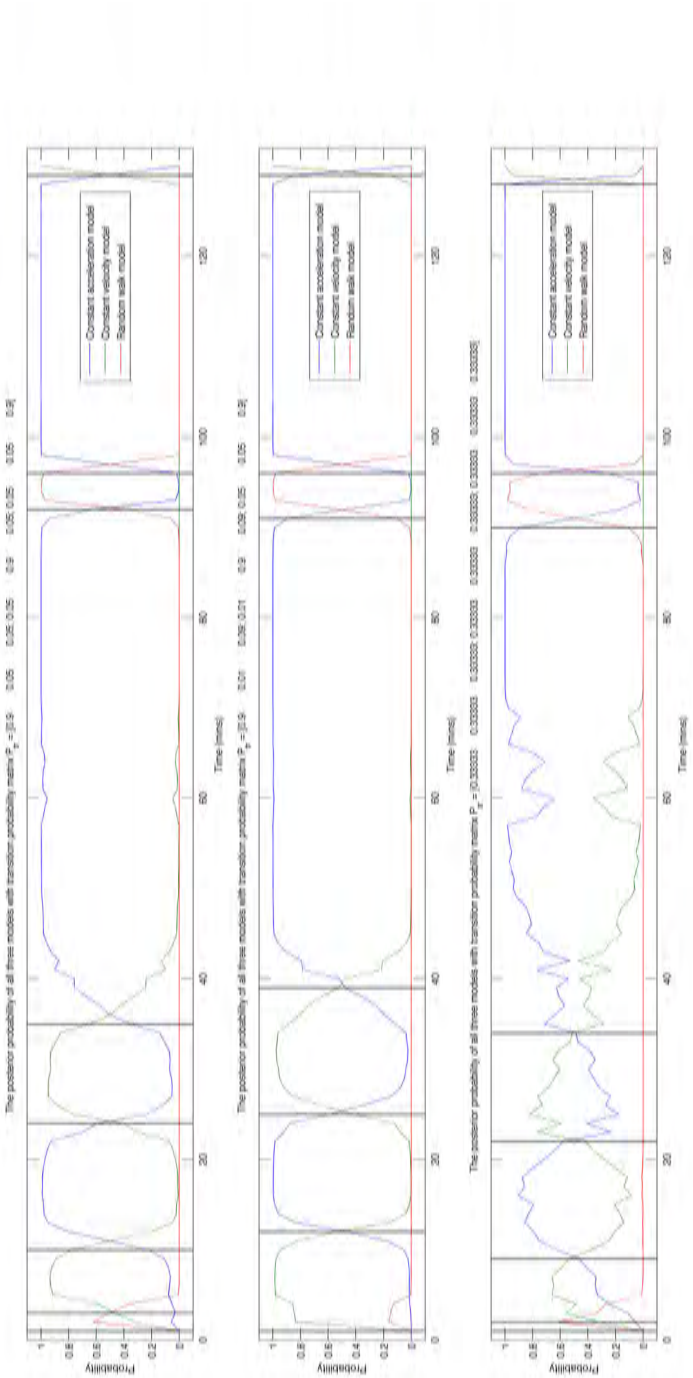

Figure 3.8: The comparison of different transition probability matrices P_{tr}.

Three different transition probability matrices $\boldsymbol{P_{tr}}$ are tested with another parameter $\boldsymbol{\mu_0} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$. The black vertical lines, the blue, green and red are the same as in Figure 3.7.

Tracking Results

According to the analysis and tests above, the parameters are fixed and the tracking result is shown in Figure 3.9. It should be noted that the diagonal value Q_d of signal noise covariance $\boldsymbol{Q}$ in the MM filter is chosen as a small value in order to make the MM sensitive. This parameter violates the rules above, but it makes the MM filter more accurate. Otherwise, each model would have been more robust and the results from the MM filter would have been inaccurate. On the top left of Figure 3.9, the trajectory is separated by different colours, which represent different models. As mentioned above, the blue, green and red lines correspond to the constant acceleration model, constant velocity model and random walk model respectively. The small purple circle indicates the current position. The legend shows the current dominant model with different colours and it changes with time. The centroid trajectory begins at the starting point with the random walk model as the dominant model and ends at the finishing point with the constant velocity model being the dominant one. During the whole tracking process, different colour trajectories are plotted to show the dominant model. The posterior probability of each model is shown in the middle left of Figure 3.9. The blue, green and red correspond to the constant acceleration model, constant velocity model and random walk model, respectively. The purple line indicates the current time point and it shows the posterior probability of all three models. In each frame, the maximum probability model is chosen as the neutrophil module, which can be easily observed from the current purple line. The bottom left of Figure 3.9 shows the absolute distance error of each model. The minimum error model is selected by MM algorithm. It should be noted that the minimum error model is the largest posterior probability model in each frame. The blue, green, red and purple are identical to those specified above. The right part of Figure 3.9 shows the comparison of velocity and acceleration on both x and y directions. The central velocity and acceleration (both plotted in black) are compared with the estimated states from each model. The blue, green and purple are the same as the above colours. The central velocity and acceleration are obtained via the Butterworth filter. It is clear that the estimated velocity and acceleration have very high accuracy considering the dominant model. Therefore, the MM filter performs very well in centroid tracking.

3.3.5 Chemotaxis Field Estimation

This is a new part from paper [82]. As an analysis tool, it is used in centroid tracking.

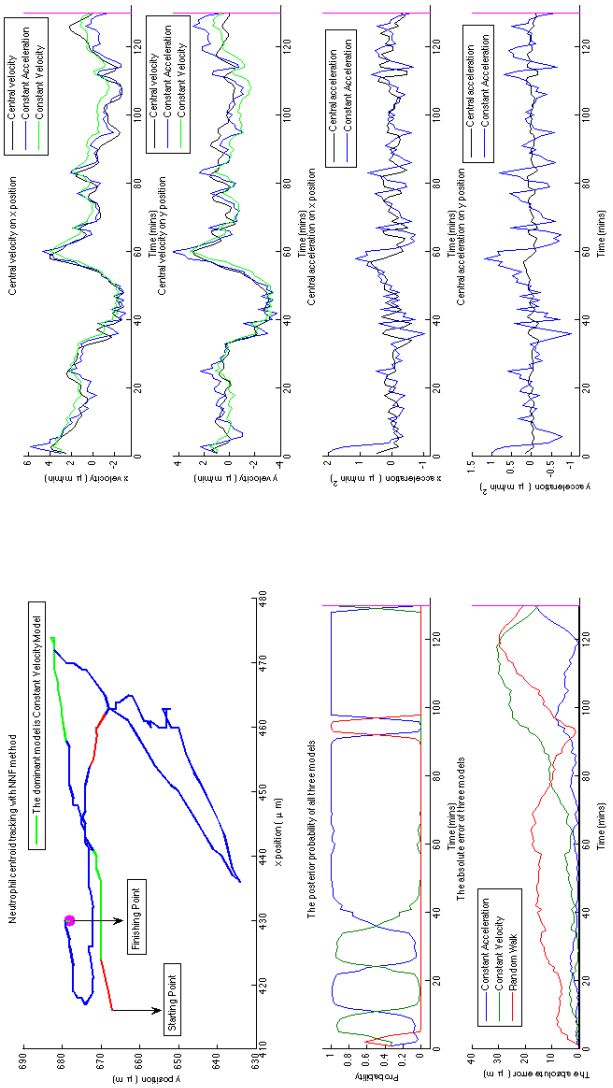

Figure 3.9: **The implementation of the MM filter with high resolution data.**

The top left figure shows the centroid trajectory with the starting point and finishing point. Different colours represent the different dominant models at different frames. The legend is the current dominant model. The blue, green and red are the constant acceleration model, constant velocity model and random walk model respectively. The posterior probability is presented on the middle left figure and the absolute errors of all three models are illustrated on the bottom left figure. The blue, green and red are the same as mentioned on the middle left figure. The central velocity and acceleration on both the x and y direction are plotted and compared with the models on the right figures.

Chemotaxis refers to the movements of cells or organisms provoked by a chemical stimulus. For example, cells move to high concentration to find food. The food is the chemical stimulus and the process is called chemotaxis. In other words, chemotaxis is the process during which cells move from low concentration to high concentration. It is similar to gradient in physics. Therefore, the estimation of the chemotaxis field is highly significant since it will help to better understand the neutrophil mechanism.

There are two assumptions to estimate the chemotaxis field. One is that the field of chemotaxis can be expressed as a functional relationship of basis functions multiplying the corresponsive weights. The other is that the changing rate of the chemotaxis field is a proportional function of velocity.

Defining the field of chemotaxis as Φ and according to the first assumption, the formula of the chemotaxis field is as follows:

$$\Phi_i = B_i w + \Phi_0 \tag{3.12}$$

where $\boldsymbol{B_i}$ is a basis function; $\boldsymbol{w}$ is the corresponsive weight; Φ_0 is the initial field as a DC component (invariant value); i is the frame index.

It should be noted that: both $\boldsymbol{w}$ and Φ_k are unknown; $\boldsymbol{B_i}$ can be different bases, such as Multi-Quadric function, and it is a Gaussian basis function in this dissertation.

Since two parameters are unknown, no method exists to calculate Φ_k. However, according to the second assumption, the following formula exists:

$$\begin{aligned} v_i &= \mu \frac{d(\Phi)}{d(S)} \\ &= \mu \frac{\Phi_{i+1} - \Phi_i}{S_{i+1} - S_i} \end{aligned} \tag{3.13}$$

where v_i is the velocity; μ is a proportional constant; S_i is the position in i^{th} frame.

While, the position S_i is a proportional function of velocity as well.

$$S_{i+1} = S_i + T v_i \tag{3.14}$$

where T is the Taylor expansion coefficient.

Therefore, combining the above formulae, it follows:

Algorithm 3.2 Chemotaxis Field Estimation Algorithm

Run Kalman smoother to calculate the velocity $\boldsymbol{v_i}$
Define parameters and the module of basis function $\boldsymbol{B_i}$
Calculate the weight $\boldsymbol{w}$ by using least square method
Calculate the chemotaxis field $\Phi_i - \Phi_0 = B_i w$ by the basis function multiplying the weight

$$\begin{aligned} \Delta\Phi &= \Phi_{i+1} - \Phi_i \\ &= (B_{i+1} - B_i)w \\ &= \frac{T}{\mu}\|v_i\|^2 \\ &= \alpha\|v_i\|^2 \end{aligned} \tag{3.15}$$

where α is a constant.

From formulae 3.15, the LS method can be used to calculate the weight $\boldsymbol{w}$, and therefore, the chemotaxis field $\Phi_i - \Phi_0 = B_i w$ can be calculated.

The algorithm was shown in Algorithm 3.2.

In implementing the algorithm, there are several parameters, which will be introduced separately.

The Type of Basis Function

The type of basis function is the first parameter to be selected. Only with the fixed type of basis function can the other parameters be fixed. This is because different types of basis function may have different parameters.

Basis function is a mathematical term. For function space, several bases exist and the linear combination of those can form any continuous function in the function space. Fourier basis is one of the most famous. In this current work, the basis function means the radial basis function, first proposed by D.S.Broomhead and Lowe [39] in 1988. The property of the radial basis function is that its values only depend on the distance between the centre and point.

Different types of basis functions affect the results. Usually, there are four types of commonly used basis functions [39], which are as follows.

Type 1 – Gaussian Basis Function

$$f = e^{-\frac{1}{2}(x-\bar{x})'\sigma^{-2}(x-\bar{x})} \tag{3.16}$$

where x is the variable; $\bar{x}$ is the mean of x; σ is the standard deviation of x.

Type 2 – Multi-Quadric Function

$$f = \sqrt{(x-\bar{x})'(x-\bar{x}) + \sigma^2} \tag{3.17}$$

Type 3

$$f = ((x-\bar{x})'(x-\bar{x}) + \sigma^2)^{-\lambda} \tag{3.18}$$

where λ can be any positive value. When $\lambda = 1$, the basis function is called Inverse Quadratic Function, and when $\lambda = 2$, the basis function is called Inverse Multi-Quadratic Function. In this part, $\lambda = 2$ was implemented and compared with other basis functions.

Type 4

$$f = (x-\bar{x})'(x-\bar{x}) log_2 \sqrt{(x-\bar{x})'(x-\bar{x})} \tag{3.19}$$

This basis function is also called Thin Plate Spline Function, which is a special kind of Polyharmonic Spline Function.

The above four basis functions are commonly used, and therefore, all will be compared and the most suitable type for neutrophils will be chosen. It should be noted that the data used in this subsection was high resolution data from the fourth fish and different from the data which was low resolution used in [82].

Each type of basis function has different effects. Therefore, different basis functions are suitable for different kinds of object. It is necessary to determine which type of basis function is more suitable for estimating the chemotaxis field of neutrophils. The comparison result is illustrated in Figure 3.10. It should be noted that in order to undertake a comparison, all the four types of basis functions are changed into the same colour scale.

Figure 3.10 shows the comparison results of the estimated chemotaxis field by different types of basis functions. It appears that the concentration of the chemotaxis field estimated by type 1 basis function changes in a minimal range. This is also the case for the type 3 basis function. Conversely, type 2 and type 4 basis functions seem to estimate the chemotaxis field in a large concentration range. Therefore, neither type 2 nor type 4 is suitable for neutrophils. However, it is still extremely difficult to classify which type more suitable for neutrophils since the

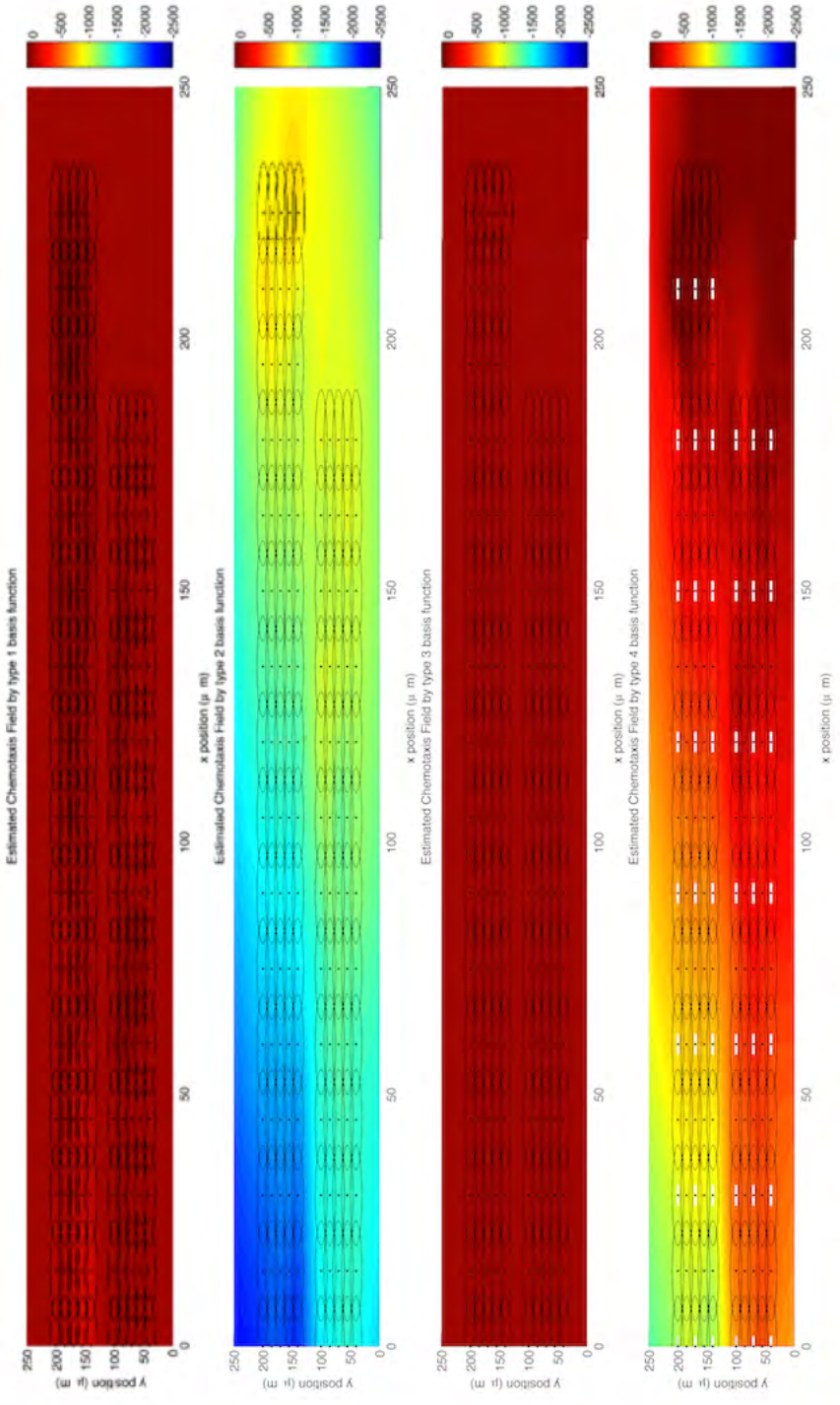

Figure 3.10: **The comparison of the estimated chemotaxis field by different types of basis functions.**

The estimated chemotaxis field of four types of basis functions are adjusted to the same colour scale (from -2500 to 100). The colour scale is too large to obtain a satisfied result.

colour scale is too large (from -2500 to 100) and the concentration range is too small. As such, another comparison result between type 1 and type 3 is needed to determine which type will be better for neutrophils.

Figure 3.11 compares the colour scales between type1 and type 3 basis functions. It is clear that the concentration of the chemotaxis field estimated by the type 3 basis function is in a minimal range (between $-10 * 10^{-7}$ and $8 * 10^{-7}$). Therefore type 3 basis function is not suitable for the chemotaxis field estimation of neutrophils. Based on the above results, the type 1 basis function, the Gaussian basis function, is selected to estimate the chemotaxis field of neutrophils both in this dissertation and in [82].

Standard Deviation of Gaussian Basis Function

Once the type of basis function had been fixed, the standard deviation was the second parameter which could affect the results. It determines the radius of the circles. If it is too small, the circles cannot cover each other; if it is too large, the covered part is too large, which can lead to a redundancy calculation.

Figure 3.12 illustrates the estimated chemotaxis field by type 1 basis function with three different values of the standard deviation σ. The top figure is the $\sigma = 5\mu m$ case; the middle figure is the $\sigma = 10\mu m$ case; and the bottom figure is the $\sigma = 20\mu m$ case. The colour scales of all the three cases are the same. The parameter σ changes the circles' radius. Furthermore, it also changes the performance of the estimation. When $\sigma = 5\mu m$ (the top case), the radius of circle is too small to connect with each other. Therefore the chemotaxis field is meaningless. It seems that the other two cases represent the chemotaxis field. However, when $\sigma = 20\mu m$ (the bottom case), the circles are over covered, and therefore, the chemotaxis field is a redundancy calculation, which exaggerates the chemotaxis field and generates a considerable colour scale. A comparison without the same colour scale between $\sigma = 20\mu m$ case and $\sigma = 10\mu m$ case is shown in Figure 3.13.

From Figure 3.13, it is clear that both the top and the bottom figures present the estimated chemotaxis field and both have nearly the same trend, that is, very high concentration on the top right and low concentration at the middle bottom. The difference is the colour scales. The concentration in the top right area at the bottom figure is nearly three times higher than the top figure and this high concentration chemotaxis is considered as the result of redundancy calculation since the circles are over covered by each other. Therefore, $\sigma = 10\mu m$ case is selected. It should be noted that, in theory, the parameter σ could be any value from half of the step, which is $7.5\mu m$ to the step $15\mu m$.

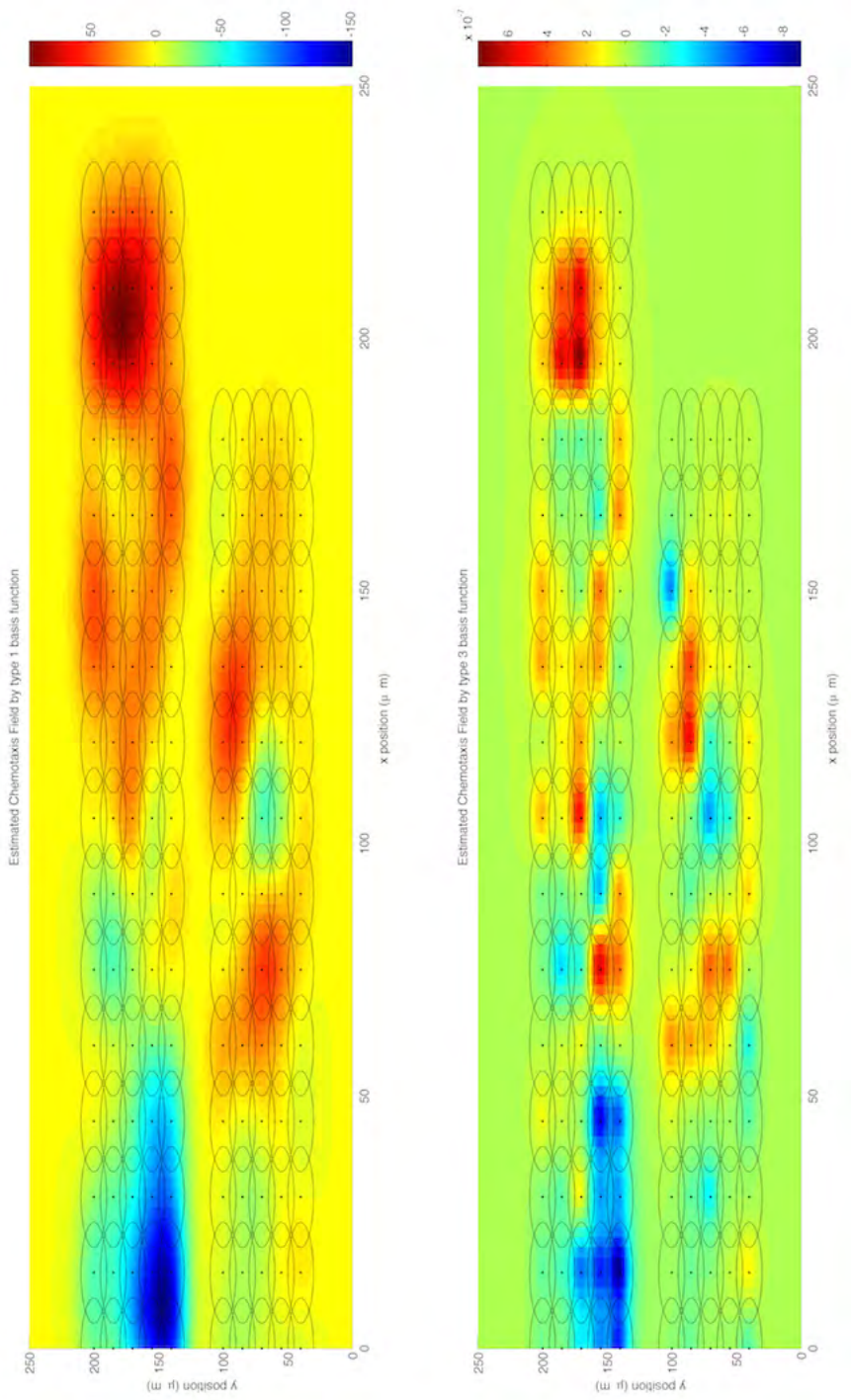

Figure 3.11: The comparison of the estimated chemotaxis field between type 1 and type 3 basis function.

The estimated chemotaxis field of type 1 and type 3 basis functions are compared and the colour scales are different in order to determine the range. They are the same as in Figure 3.10, except for the colour scales. The colour scale of the top figure is between -100 and 100, while it is between $-10 * 10^{-7}$ and $8 * 10^{-7}$ at the bottom one.

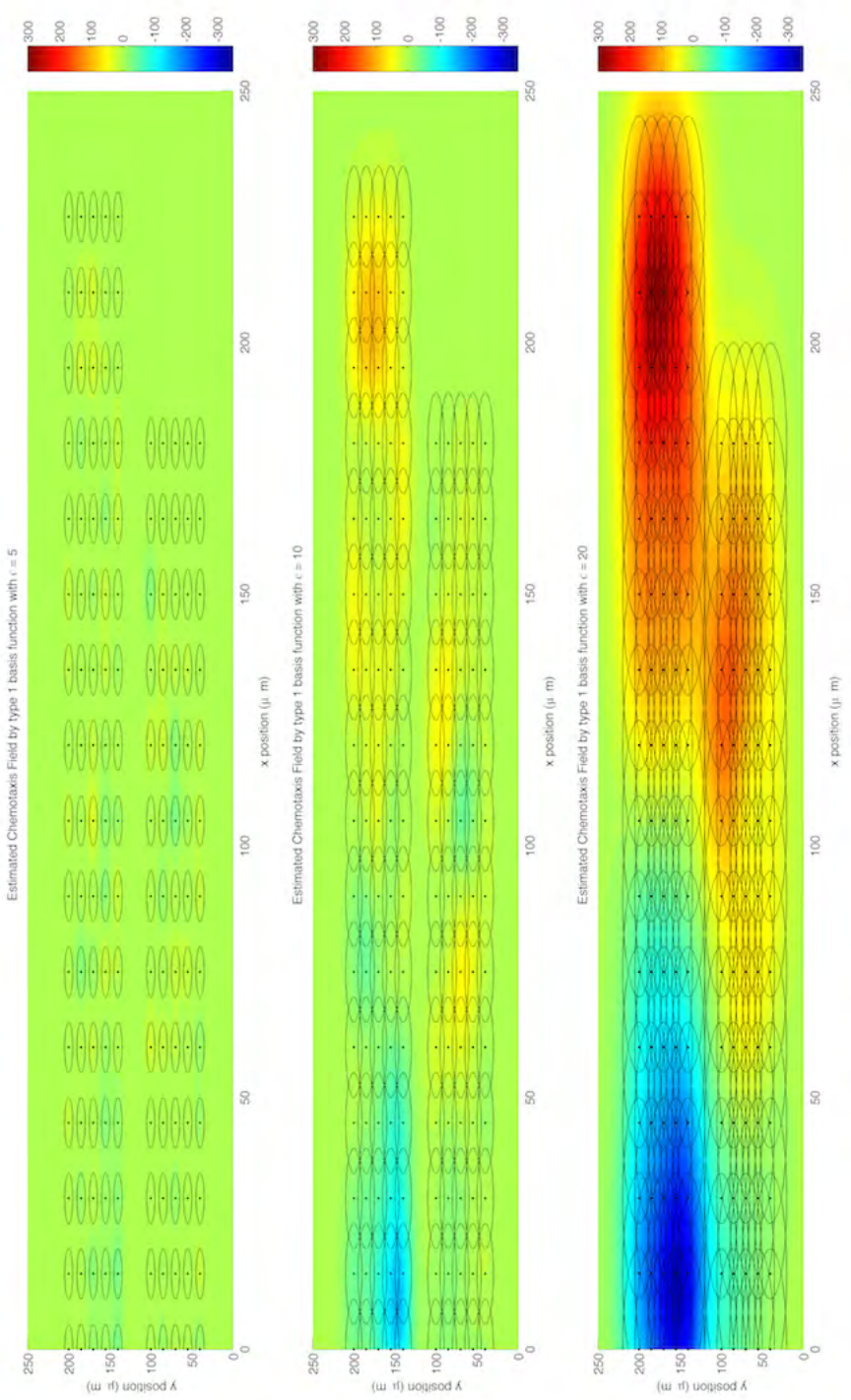

Figure 3.12: **The comparison of the chemotaxis field estimation with different σ.**

The chemotaxis field is estimated by type 1 basis function with three different values of σ. The other parameters are kept the same. The colour scales are adjusted to the same range.

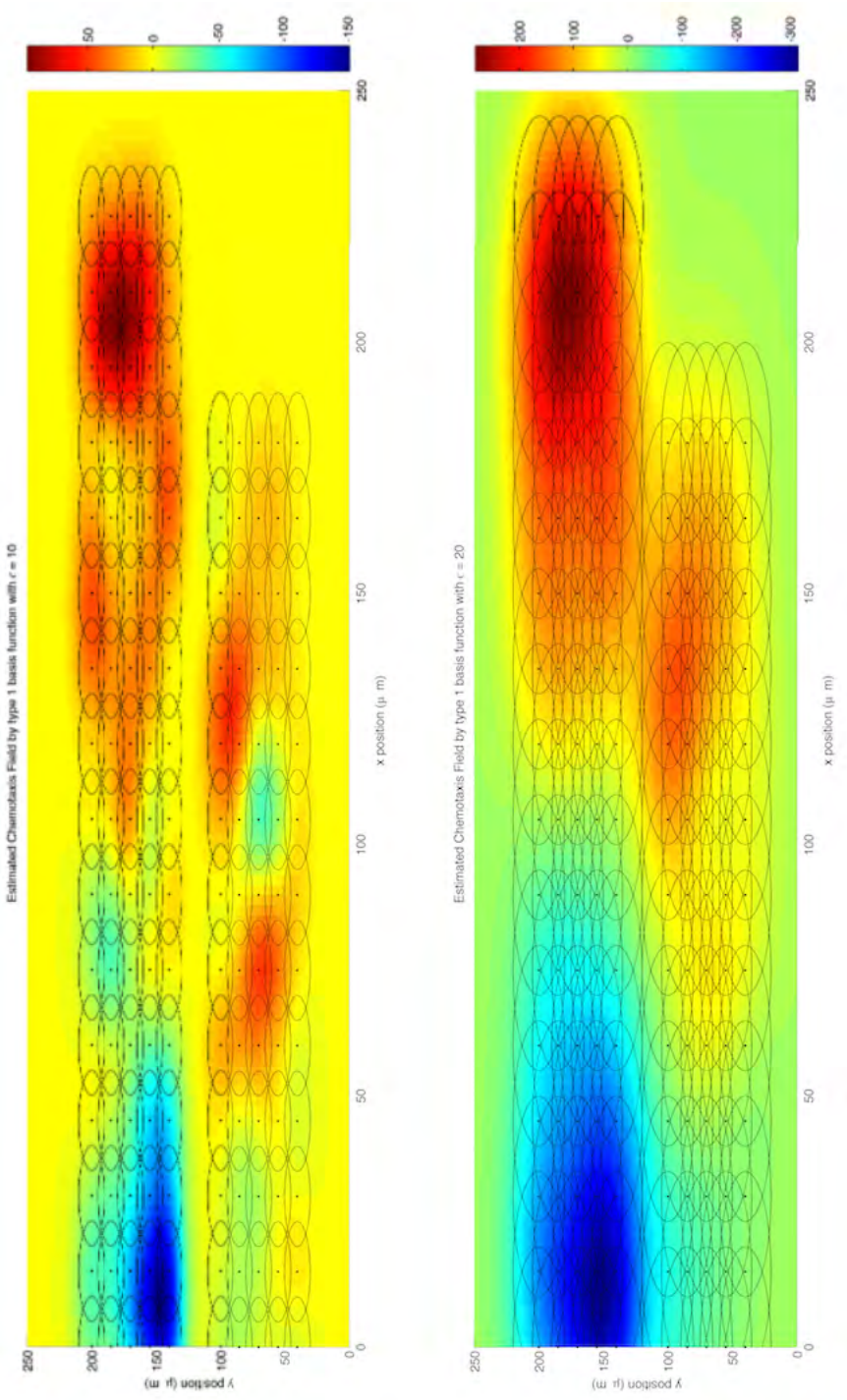

Figure 3.13: **The comparison of different σ**

This figure is the same as Figure 3.12 except for the colour scale. Both of the figures represent the chemotaxis field, which is high concentration on the right and low concentration on the left. The colour range of the over covered σ is considerable (from -350 to 300) (shown at the bottom figure), while the colour range of the other is small (from -150 to 100) (shown in the top figure).

Steps between the Centroids of Gaussian Basis Function

The step was the last parameter and had a very strong connection with standard deviation. It decided the distances between the centroid of the circles.

Figure 3.14 illustrates that different steps have different estimated chemotaxis fields. The results are similar to the different σ results above. All four figures are adjusted to the same colour scale. In this test, all the models use type 1 basis function and the $\sigma = 10\mu m$. The first figure (the top one) is the case of $step = 10\mu m$ and the circles are over covered. Based on the above situation, the chemotaxis field is a redundancy calculation. The third figure (from the top) and the last figure (the bottom one) are not suitable for the chemotaxis estimation because the circles do not cover each other, in either figure. In this case, the chemotaxis field could not be estimated accurately. It should be noted that the third figure is the critical case, which means that the circles just connect to each other but do not cover. Therefore, the basis function could not work perfectly because of lack of the data. It should also be noted that the value of the step parameter should be in the range between σ and $2 * \sigma$, which is between $10\mu m$ and $20\mu m$ in this test. Therefore, it is clear that the step decides the distances between the centroids while the standard deviation σ decides the radius of the circles. The selection of both decide the estimation results of the Gaussian basis function.

From the above results, two conclusions are reached as follows. The first is that the Gaussian basis function can express the chemotaxis field of neutrophils properly, when compared with the three other types of basis functions detailed above. The second is that the selections of both the steps and the standard deviation σ decide the performance of the Gaussian basis function.

3.4 Conclusion

In summary, this chapter has mainly covered the centroid tracking. The introduction section focused on the background and commonly used tracking algorithms. Several tracking algorithms were selected for the neutrophil application, including the KF. As different parameters can affect results, all the parameters of the selected algorithms were tested to determine the optimal ones. Therefore the algorithms could be implemented with the optimal parameters. Finally, another algorithm was implemented. It estimated the chemotaxis field effectively and formed a part of [82]. Innovation and contributions are as follows. The chemotaxis field was proved to exist and was estimated based on the information of centroid velocity. As an effective tool, it was used to analyse neutrophils' behaviour. Therefore, the mechanism was better understood. The MM filter was used to establish the math-

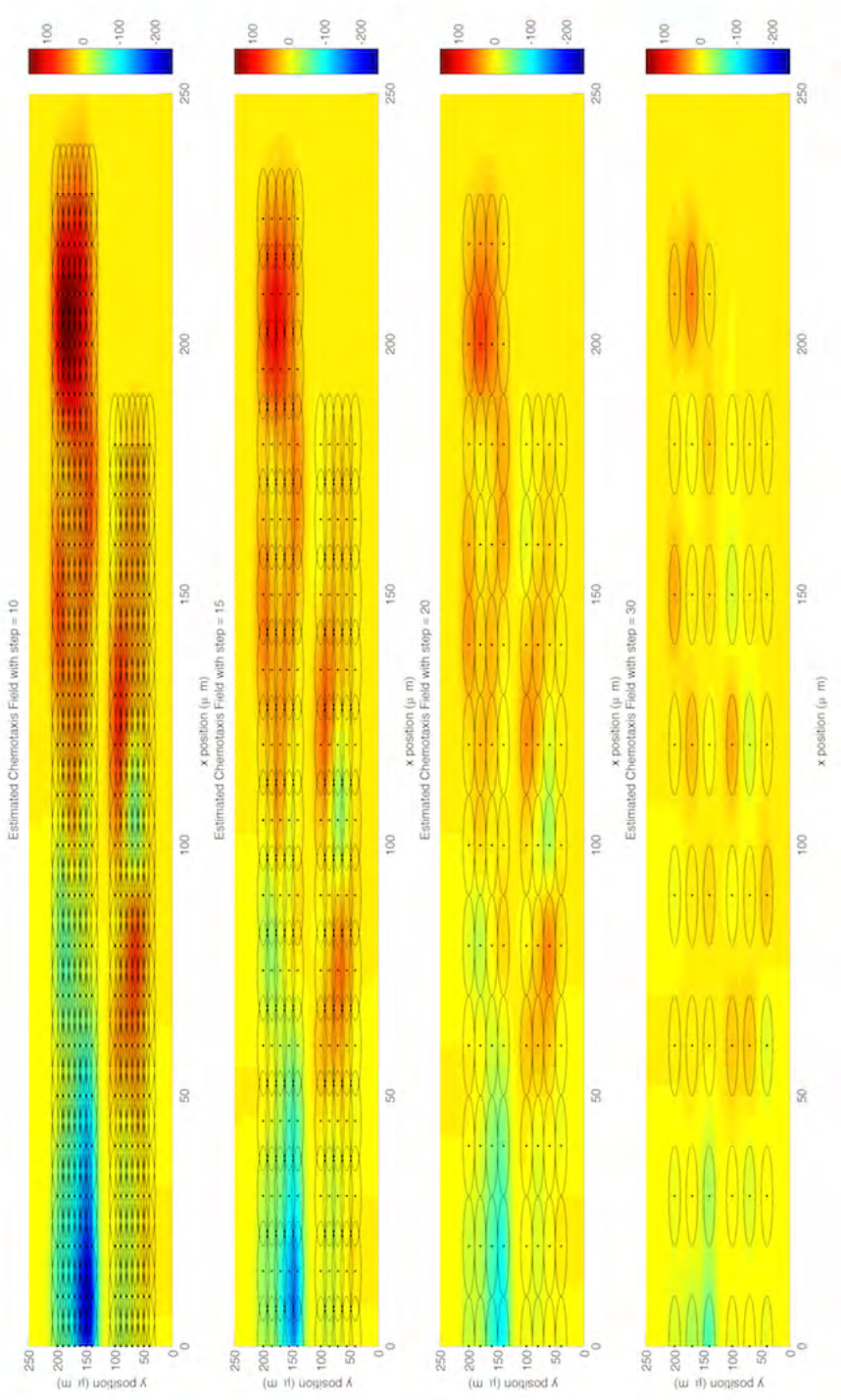

Figure 3.14: The comparison of different steps

The chemotaxis field is estimated and compared with different steps. All the other parameters are kept the same. The colour scales are adjusted to the same.

ematical model for neutrophils and it tracked the changing of neutrophil mode effectively, which is useful to understanding neutrophils' behaviour and modes. In addition, the MM filter implemented in the medical area was another application area. To further understand the mechanism of neutrophils, shape tracking is needed and will be introduced in the next chapter.

Chapter 4

Tracking the Dynamics of Neutrophils' Boundaries

The previous chapter discussed the centroid tracking, including mode behavioural detection by using a MM filter. The chemotaxis field was estimated by the centroid velocity from the KF. In addition to centroid information, shape information is also important. This chapter has, as its focus, the shape tracking of neutrophils to obtain shape information. Usually, shape boundary of neutrophils contains a very huge amount of information, even more than that of the centroid. Therefore, in this chapter, shape tracking will use the information on the shape boundary together with the boundary dynamics to obtain more knowledge.

4.1 Introduction

The central idea of this chapter is to use computer vision technique to follow the change of neutrophils' shape from frame to frame.

A large amount of research on computer vision exists. The famous one is this area is [102] in the early stage, which focused on a low-level process. However, low-level vision was not good, as per [19], and the theory and practice of finding images' important features were rapidly developed thereby opposing the low-level vision. For example, Perona and Malik [119] made efforts on feature detection. Aside from the theory mentioned above, snakes were another important field. Kass et al. [85] first proposed snakes, which was followed by a vast body of relevant research. Scott [138] applied some extensions, using Fourier parameterisation on the initial snake paper. Amini et al. [3] modified the hard constraints of the initial paper. Snakes were first implemented by Cipolla and Blake, using the B-Spline method [31]. Curwen et al. [35] added Lagrangian dynamics to snakes. All

these methods belong to FEM [19]. Deformable templates were another method, which appeared earlier than snakes. However, the development of snakes terminated the rapid development of deformable templates. Numerous variations of deformable templates emerged. For example, Grimson introduced visual object recognition with a three-dimensional approach in [56]. The KF, with the development of active contours, has been widely used in the computer vision area, since the end of 1980s [9]. Terzopoulos and Szeliski [152] established the connection between snakes and the KF. Blake et al. [18] implemented B-Spline with the KF.

Recently, the shape tracking technique has been widely used in biological area. It is believed that cells are polarised and migrated by chemoattractants [158]. Therefore, research in respect of motility and morphology is important to further understand the mechanism. P.A. Negulescu and Cahalan [116] illustrated the sensitivity of shape and motility. Ray et al. [127] applied active contours to track leukocytes. Therefore, this is a method to address the current project, combining an active contour with traditional tracking methods in cell tracking. The techniques used in this chapter will be introduced separately.

4.2 Dynamic Shape Model

To address the dynamic shape changing, tracking techniques are detailed in this section.

Four techniques are introduced: augmented state space based on shape boundary, one frame B-Spline shape tracking and retrieving, time lapse multiple neutrophil NNF tracking, and shape boundary velocity estimation.

4.2.1 Augmented State Space Model

A mathematical model is also needed for neutrophil shape tracking. The basic mathematical model in modern control is the state space model. In this current work, the state space model is the augmented model. The difference is that the augmented state space model considers the shape boundary as a set of a large number of single active individual particles with each particle being one state in the augmented state space model. Only one commonly used model (the constant velocity model) is used in this current work instead of the combination of three models (the constant acceleration model, constant velocity model and random walk model).

It should be noted that the random walk model does not contain the state of velocity. The constant acceleration model has, however, very expensive computation costs, since it has six states. Therefore, the constant velocity model was

selected.

It is assumed that each particle on the boundary satisfied the constant velocity model. In the constant velocity model, the velocity is expected to maintain a fixed constant value. Then the state space model for each particle is formed the same as the discrete time constant velocity model, discussed in chapter two.

As all the particles satisfied the same formulae above, all the particles' formulae can be combined and rewritten in a new augmentation formula:

$$\boldsymbol{x(k)} = \left((\boldsymbol{x_1(k)})^{\boldsymbol{T}}, (\boldsymbol{x_2(k)})^{\boldsymbol{T}}, \cdots, (\boldsymbol{x_n(k)})^{\boldsymbol{T}}\right)^T \tag{4.1}$$

where the elements of the state vector are

$$\begin{aligned}
(\boldsymbol{x_1(k)})^{\boldsymbol{T}} &= \left((s_x^1(k))^T, (v_x^1(k))^T, (s_y^1(k))^T, (v_y^1(k))^T\right) \\
(\boldsymbol{x_2(k)})^{\boldsymbol{T}} &= \left((s_x^2(k))^T, (v_x^2(k))^T, (s_y^2(k))^T, (v_y^2(k))^T\right) \\
&\vdots \\
(\boldsymbol{x_n(k)})^{\boldsymbol{T}} &= \left((s_x^n(k))^T, (v_x^n(k))^T, (s_y^n(k))^T, (v_y^n(k))^T\right)
\end{aligned} \tag{4.2}$$

where $s_x^i(k)$ and $s_y^i(k)$ are the i^{th} position on x and y direction; $v_x^i(k)$ and $v_y^i(k)$ are the i^{th} velocity on x and y direction. All are scalars.

The state vector $\boldsymbol{x}$ describes the position and velocity of the control points which are weighted function in B-spline (B-spline will be introduced later in this chapter) in each spatial direction.

It should be noted that $\boldsymbol{x(k)}$ is a vector and the element $\boldsymbol{x_i(k)}$ is also a vector.

Therefore, the other matrices of the augmented state space model are as follows.

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{F_n} \end{pmatrix} \tag{4.3}$$

$$\boldsymbol{F_1} = \boldsymbol{F_2} = \cdots = \boldsymbol{F_n} = \begin{pmatrix} 1 & h & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & h \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.4}$$

$$\boldsymbol{G} = (\boldsymbol{G_1}, \boldsymbol{G_2}, \cdots, \boldsymbol{G_n})^T \tag{4.5}$$

$$\boldsymbol{G_j} = \left(\frac{h^2}{2}, h, \frac{h^2}{2}, h\right) \text{ for} j = 1, \ldots, n \tag{4.6}$$

$$\tag{4.7}$$

where n is the number of particles on the boundary.

The state transition matrix $\boldsymbol{F}$ describes a dynamic model of the cell boundary movement driven by a constant velocity assumption.

The above formulae form the basic mathematical model for shape tracking, and this is a module in the shape tracking together with the other algorithms.

4.2.2 B-Spline

The B-spline model is a method of parametric Active Contours and was first mentioned as snake in [85] and [151]. Usually, there are several important parameters, which are introduced below.

The basis function $\boldsymbol{B_{L,d}(s)}$ is also called span matrices. According to [19], span matrices can be solved in the formula off-line as follows:

$$\boldsymbol{B_{L,d}(s)} = \frac{(s-L)\boldsymbol{B_{L,d-1}(s)} + (L+d-s)\boldsymbol{B_{L+1,d-1}(s)}}{d-1} \tag{4.8}$$

with the ground instance

$$B_{L,1}(s) = \begin{cases} 1, if\ L \leqslant s \leqslant L+1 \\ 0, otherwise \end{cases} \tag{4.9}$$

where $\boldsymbol{B_{L,d}}$ is the L^{th} basis function for a spline of order d;

The placement matrices $\boldsymbol{G_\sigma}$ is:

$$(G_\sigma)_{ij} = \begin{cases} 1, if\ i - b_\sigma = j \\ 0, otherwise \end{cases} \tag{4.10}$$

where $b_\sigma = \sum_0^\sigma m_i - d$ is the index and m_i is knot multiplicities.

For $0 \leqslant s < 1$ it follows:

$$\boldsymbol{x(s+\sigma)} = (1, s, \cdots, s^{d-1})\boldsymbol{B_{\sigma,d}(s)}\boldsymbol{G_\sigma}\boldsymbol{Q_c} \tag{4.11}$$

where $\boldsymbol{Q_c}$ is a vector of the control points.

The metric matrix $\boldsymbol{\beta}$ was used in the program instead of using $\boldsymbol{B_{L,d}(s)}$ directly. The formula was given as follows.

$$\boldsymbol{\beta} = \frac{1}{L}\int_0^L \boldsymbol{B_{L,d}(s)B_{L,d}^T}(s)ds \tag{4.12}$$

Therefore, based on the formulae above, the control points $\boldsymbol{Q_c}$ can be solved using the LS method. The details can be found in Algorithm 4.1. It should be noted that the metric matrix for curves $\boldsymbol{U}$ is the function of the metric matrix $\boldsymbol{\beta}$. Using $\boldsymbol{\beta}$ and $\boldsymbol{U}$ in the program has reduced the computation complexity.

Algorithm 4.1 B-Spline Algorithm

Calculate the placement matrices $\boldsymbol{G_\sigma}$ and the span matrices $\boldsymbol{B_{L,d}(s)}$ off-line given L and d
for $i = 0$ to $L - 1$ **do**
  Find s between i and $i + 1$
  for $j = 1$ to the length of s **do**
    Calculate $\boldsymbol{U} = (1, s, \cdots, s^{d-1})\boldsymbol{B_{\sigma,d}(s)G_\sigma}$
    Calculate the control points $\boldsymbol{Q_c} = \boldsymbol{(U'U)^{-1}U'Z}$, where $\boldsymbol{Z}$ is the measurement data, by using least square method
  end for
end for

In the program, the different interval lengths of the low resolution data were verified, with the results shown later in this chapter. As the error for different interval lengths of the low resolution data is small, finally the interval length as $L = 10$ and the order of the polynomial as $d = 4$ were decided.

Combined with the augmented state space model in the previous subsection, the mathematical model was established as follows:

$$\begin{aligned} \boldsymbol{x_{k+1}} &= \boldsymbol{Fx_k + Gw} \\ \boldsymbol{z_k} &= \boldsymbol{Hx_k + v} \end{aligned} \tag{4.13}$$

where, $\boldsymbol{x_k}$ represents the control points based on the B-spline model, containing four elements: the control points position in x direction, the velocity of the control points in x direction, the control points position in y direction, and the velocity of the control points in y direction ; $\boldsymbol{w}$ and $\boldsymbol{v}$ are white noise with the same definition as in chapter two; $\boldsymbol{z_k}$ is the output, that is, the measurement at time k, only having two elements: position measurement in x direction and the position measurement in y direction.

B-Spline contains not only the process of calculating the control point, but also the process of fitting the curve and other processes, such as pose recovery. However, those processes were not used in the thesis and program, and they will not be presented. The process of fitting curves is generally used in a real time system which requires very fast calculation and low computational cost, which were validated in [12]. Usually, the process of fitting curves uses a recursive LS estimation algorithm [12].

The B-Spline algorithm is usually combined with other tracking algorithms, such as the KF. The application of combined algorithm will be introduced in the velocity estimation section.

4.2.3 Nearest Neighbourhood Method

The NNF is a filter using the "nearest neighbourhood" technique. It can be widely used in a single algorithm or as a part of a complex algorithm [93]. The nearest neighbourhood technique in the target tracking is the "closest" to the validated measurement compared with the predicted measurement from the previous time period. However, a problem exists. According to [12], Bar-Shalom and Fortmann believed that sometimes the chosen nearest neighbourhood is, probably, not the right measurement. This might lead to incorrect tracking. For example, the neutrophil disappeared in frame k. Then the nearest neighbourhood measurement either did not exist or had the wrong value. Based on these reasons, Li and Bar-Shalom [93], using the data association method, solved the problem having calculated the probabilities of different events.

The mathematical expression of the nearest neighbourhood measurement is as follows:

$$min(D^2(z_{k+1})) \tag{4.14}$$

where $D^2(z_{k+1})$ is the square of the normalized distance, and the formula of $D^2(z_{k+1})$ is defined as follows:

$$D^2(z_{k+1}) = [\boldsymbol{z_{k+1}} - \hat{\boldsymbol{z}}_{\boldsymbol{k+1|k}}]' \boldsymbol{S}_{\boldsymbol{k+1}}^{-1} [\boldsymbol{z_{k+1}} - \hat{\boldsymbol{z}}_{\boldsymbol{k+1|k}}] \tag{4.15}$$

where $\boldsymbol{z_{k+1}}$ is the validated measurement at time $k+1$; $\hat{\boldsymbol{z}}_{\boldsymbol{k+1|k}}$ is the predicted measurement at time $k+1$ given the estimated measurement $\hat{\boldsymbol{z}}_{\boldsymbol{k}}$ at time k; $\boldsymbol{S_k}$ is the covariance matrix of the measurement noise and usually $\boldsymbol{S_k}$ is assumed as time-invariance, that is, $\boldsymbol{S_1} = \boldsymbol{S_2} = \ldots = \boldsymbol{S}$

The observation equation was introduced below:

$$z_k = \boldsymbol{H}\boldsymbol{x}_k + v_k \tag{4.16}$$

where $\boldsymbol{H}$ is the discrete time observation matrix; $v_k \sim N(0, \boldsymbol{S}_k)$ is the observer noise. Usually the covariance matrix $\boldsymbol{S}_k$ of the white noise v_k is assumed time-invariance, that is, $\boldsymbol{S}_k = \boldsymbol{S}$, $\forall k \in I^+$

It should be noted that whether the matrix $\boldsymbol{H}$ contains zero columns depends on the model. If the model is the random walk, $\boldsymbol{H}$ does not contain zero columns. However, if the model is the constant velocity, $\boldsymbol{H}$ contains two zero columns. If the model is the constant acceleration, $\boldsymbol{H}$ contains four zero columns. Therefore, $\boldsymbol{H}$ can be written as $\boldsymbol{H} = [\boldsymbol{H_1}, \boldsymbol{0}]$ in general, where $\boldsymbol{H_1}$ is the corresponding matrix of position elements and $\boldsymbol{0}$ is the corresponding matrix of velocity elements (or the combination of velocity and acceleration elements or an empty matrix).

Taking the expectation of (4.16), the equation was obtained as follows:

$$\hat{z}_k = \boldsymbol{H}\hat{\boldsymbol{x}}_k \tag{4.17}$$

Combined with the KF (the KF is introduced in chapter two) formulae, the predicted measurement $\hat{z}_{k+1|k}$ can be calculated by:

$$\hat{z}_{k+1|k} = \boldsymbol{H}\hat{\boldsymbol{x}}_{k+1|k} \tag{4.18}$$

However, $\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{F}\hat{\boldsymbol{x}}_k$ (from the KF formula). Therefore the predicted measurement at time $k+1$ can be rewritten as a function of the estimated state $\hat{\boldsymbol{x}}$ at time k. The formula is shown below:

$$\hat{z}_{k+1|k} = \boldsymbol{H}\boldsymbol{F}\hat{\boldsymbol{x}}_k \tag{4.19}$$

where $\hat{\boldsymbol{x}}_k$ is the state estimator, and $\boldsymbol{F}$ is the discrete time system matrix. In this part, the constant velocity model and B-Spline were used in the previous work. Therefore, $\hat{\boldsymbol{x}}_k$ and $\boldsymbol{F}$ have the same values as the above section in this chapter. The nearest neighbourhood algorithm can be found in Algorithm 4.2.

The relevant results are shown later in this chapter.

The nearest neighbourhood method solved the relevant neutrophils in the continuous time index.

According to the nearest neighbourhood method, the distances, between the estimated state of a single neutrophil at time $k+1$, given the state at time k, and the state of all the neutrophils at time $k+1$, are calculated, and the minimum distance should be the relevant neutrophil at time k and $k+1$. All the neutrophils are found in the correct order and index by using the nearest neighbourhood

Algorithm 4.2 Nearest Neighbourhood Algorithm

for $iFrame = 1$ to $N - 1$ **do**
  for $m = 1$ to n_{iFrame} **do**
    Use B-spline to calculate the state space of control points $\hat{\boldsymbol{x}}^{\boldsymbol{m}}_{\boldsymbol{iFrame}}$
    Do a one-step prediction $\hat{\boldsymbol{z}}^{\boldsymbol{m}}_{\boldsymbol{iFrame+1|iFrame}}$ of an output of the state space model as a nearest neighbourhood measurement and calculate the mean of the output $\hat{\bar{\boldsymbol{z}}}^{\boldsymbol{m}}_{\boldsymbol{iFrame+1|iFrame}}$ compared with the next time mean $\bar{\boldsymbol{z}}^{\boldsymbol{m}}_{\boldsymbol{iFrame+1}}$ of all the particle on the boundary as a target measurement
  end for
  Find the minimum distance and decide whether the minimum distance is out of the threshold
  Check whether more than one cell corresponds to the same cell or some cells disappear
  Decide the new coming cells and disappearing cells
end for

method. Therefore, from each frame, the augmentation state space model gave the correct information of control points based on the B-spline model, and the nearest neighbourhood algorithm gave the correct index information from frame to frame. The velocity of the control points was included in the information. It was meaningless in physics and engineering. However, by using the B-Spline retrieved technique, the velocity of particles on the boundary can be calculated. Although the velocity does not exist, it can show the trend of neutrophils moving to the chemoattractant field.

4.2.4 Velocity Estimation

The velocity estimation section will obtain the significative results. The core idea of the velocity estimation is to use the KF to estimate the boundary velocity.

The velocity estimation is the extension of the state observer. Usually, the observer is used to observe states which are not "read" or "seen" from instruments. It is a mathematical method to solve the problem. For this study, although the velocity of particles on the neutrophil boundary does not exist, it is highly significant for analysing features.

The same technique as implementing B-Spline is used to calculate the velocity on the boundary. Formula 4.17 gave the estimation of output equation, where $\boldsymbol{H}$ is the observation matrix. As mentioned above, $\boldsymbol{H}$ represents the transformation from the control points to the estimation and has the following formula.

$$\boldsymbol{H} = [\boldsymbol{H_1}, \boldsymbol{0}] \tag{4.20}$$

The $\boldsymbol{H}$ matrix changed every frame and every neutrophil. Even for the same neutrophil, the $\boldsymbol{H}$ matrix had different sizes in different frames.

Based on the above statements, two methods to calculate the velocity from the control points exist. One uses mathematics to change the output equations; the other is the retrieved technique based on B-Spline. For the first method, the output equation can be transformed as follows: $\boldsymbol{Z} = \boldsymbol{H}\boldsymbol{Q}_c = [\boldsymbol{H}_1\boldsymbol{0}][\boldsymbol{S}_Q\boldsymbol{V}_Q]'$, where $\boldsymbol{Z}$ is the measurement of shape boundary; $\boldsymbol{H}$ and $\boldsymbol{H}_1$ are the same as mentioned above; $\boldsymbol{Q}_c$ is the state vector and also known as the control points; $\boldsymbol{S}_Q$ is the position state and $\boldsymbol{V}_Q$ is the velocity state. Therefore, the boundary velocity can be calculated by $\boldsymbol{V}_b = \boldsymbol{H}_1\boldsymbol{V}_Q$, where $\boldsymbol{V}_b$ is the velocity on the boundary. However, in this method, the size of velocity changes every frame and the number of velocity is equal to the number of real boundary points. For the second method, the number of the boundary velocity remains the same and it is easy to compare and analyse.

The comparison results are shown later in this chapter.

4.3 Results

In this section, the results are shown from the B-Spline section.

It should be noted that the results in this chapter were different from those in the previous chapter. In this chapter, most of the results were based on multiple targets shape tracking while the results in the previous chapter were based on single target centroid tracking. Therefore, most of the tracking results in this chapter were zoomed in on one neutrophil, either in a specific figure or in another figure for a clear comparison.

It should also be noted that all the data in this chapter was low resolution data from Fish 7.

4.3.1 B-Spline Implementation

B-Spline is a technique based on the shape tracking. It can track and follow the shape of neutrophils very well. The following experiment verified the effect of using B-Spline in shape tracking.

B-Spline Implementation

According to Algorithm 4.1, the control points $\boldsymbol{Q}_c$ can be calculated. To verify whether B-Spline is a good shape tracking estimator, the estimated output equation 4.17 is calculated. The estimated output $\hat{\boldsymbol{z}}_k$ is compared with the measurement $\boldsymbol{z}_k$. Figure 4.1 illustrates the comparison result.

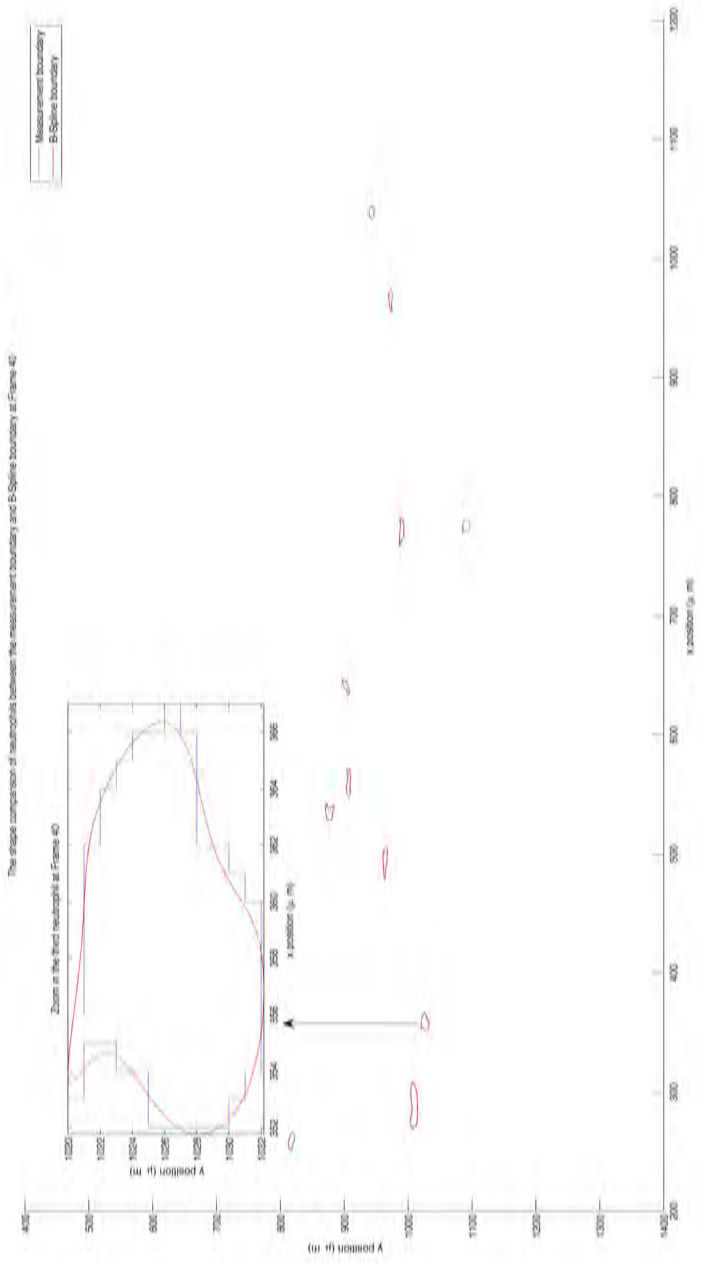

Figure 4.1: **B-Spline method implemented on neutrophils.**

B-Spline is implemented and its retrieved boundary is compared with the measurement boundary. The third neutrophil is zoomed in at the top left area of Figure 4.1. The measurement boundary is the neutrophil shape boundary from the data. Therefore, the shapes are not smooth. Those shapes are represented in broken blue line. B-Spline retrieved boundary is the estimated neutrophils shape boundary and therefore, they are smooth. B-Spline retrieved boundary is shown in solid red line.

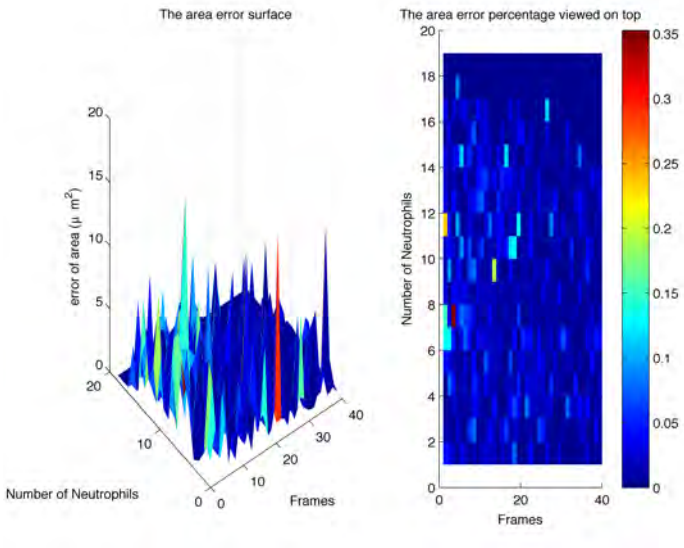

Figure 4.2: **The area error comparison between the measurement boundary and the B-Spline retrieved boundary.**

The left figure shows the area error of all the neutrophils in all the frames as a surface. The right figure presents the error rate of area error. The colour bar was the error rate.

In Figure 4.1, all the neutrophils in Frame 40 are estimated by B-Spline in a solid red line and all the measurement boundaries are plotted in a broken blue line. All the neutrophils are tracked, with none missing. This is also true for the neutrophils in other frames. On the top left area, the third neutrophil is zoomed in. From this zoomed in figure, it is clear that B-Spline estimates the neutrophil shape boundary well. In addition, B-Spline smooths the measurement boundary. This also happens for the other neutrophils. To show the effect of the B-Spline method, the area error has been shown in Figure 4.2.

The left of Figure 4.2 is the surface of area error. All the measured neutrophil areas are recorded and compared with the calculated B-Spline neutrophil areas. From the left of Figure 4.2, it is clear that most of area errors have minimal values. From the program calculation, the average area of a neutrophil is $118\mu m^2$ in the data. Therefore, the area error is acceptable. However, it is still unknown the accuracy of B-Spline. In addition, the areas of different neutrophils are different. Based on the above reasons, the area error percentage is plotted in the right of Figure 4.2. It is clear that the area error percentage of only a few neutrophils in a few frames is large. From the program calculation, assuming the accurate rate is 95%, only 9% of neutrophils have a larger error in 600 neutrophils. Assuming

that the accurate rate is 75%, only 0.17% of neutrophils have larger error in 600 neutrophils, which means that only one neutrophil in one frame has a larger error (the 8^{th} neutrophil in the 3^{rd} frame). The above calculation result proves that B-Spline is an effective tool in shape tracking.

Interval Length Parameter of B-Spline

The parameters, the interval length L and the spline matrix order d, decide the performance of B-Spline. d is the spline matrix order and it is a positive integer. Usually, d is equal to 2, 3, and 4. The parameter, $d = 2, 3, or 4$, represents using the combination of lines, ramp, or binomial curve to approximate the measurement shape respectively. It should be noted that the B-Spline parameter d is fixed during the program and it could not be changed. In this study, the parameter d was equal to 4. L is the number of control points in B-Spline. For different resolution cases, the best parameter L is different. In the low resolution case, a comparison experiment was done for different parameters $L = 5, 10, 20$. The results are as follows.

Figure 4.3 shows the performance of different numbers of control points. The third neutrophil is zoomed in at the top left of the figure. The measurement shapes are plotted in black lines. The red, green and blue broken lines are $L = 5$, 10, and 20 respectively. It is clear that when $L = 5$, the B-Spline shape could not represent the measurement shape clearly, especially for the pseudopod. While for $L = 10$ and 20, the pseudopod on the top left is presented clearly. Moreover, the B-Spline shapes of both cases are similar to the measurement shape. To select the best parameter, the area errors are plotted, as shown in Figure 4.4.

Figure 4.4 illustrates the area error comparison of different numbers of control points on the left. The area error percentage is also compared on the right. Both the coordinate scales and the colour scales are adjusted to be the same, for easy comparison. It is clear that the area errors of the $L = 5$ case are greater than those of the other two on the left of Figure 4.4. However, it is extremely difficult to distinguish whether the $L = 10$ or $L = 20$ case is better. On the right hand side, the area error percentage represents the error rate. The small square with a deep colour, such as red, is the bad estimator of B-Spline. For instance, the area error percentage is deep red for the 8^{th} neutrophil in the third frame of $L = 5$ case, which means that the shape boundary from B-Spline has a considerable area error percentage. It is clear that the $L = 5$ case has a large number of those small squares with deep red and other colours with values greater than 0.25. It appears that $L = 20$ case has fewer numbers of area error percentage than those of the other two. However, the calculation cost of $L = 20$ case is also larger than those

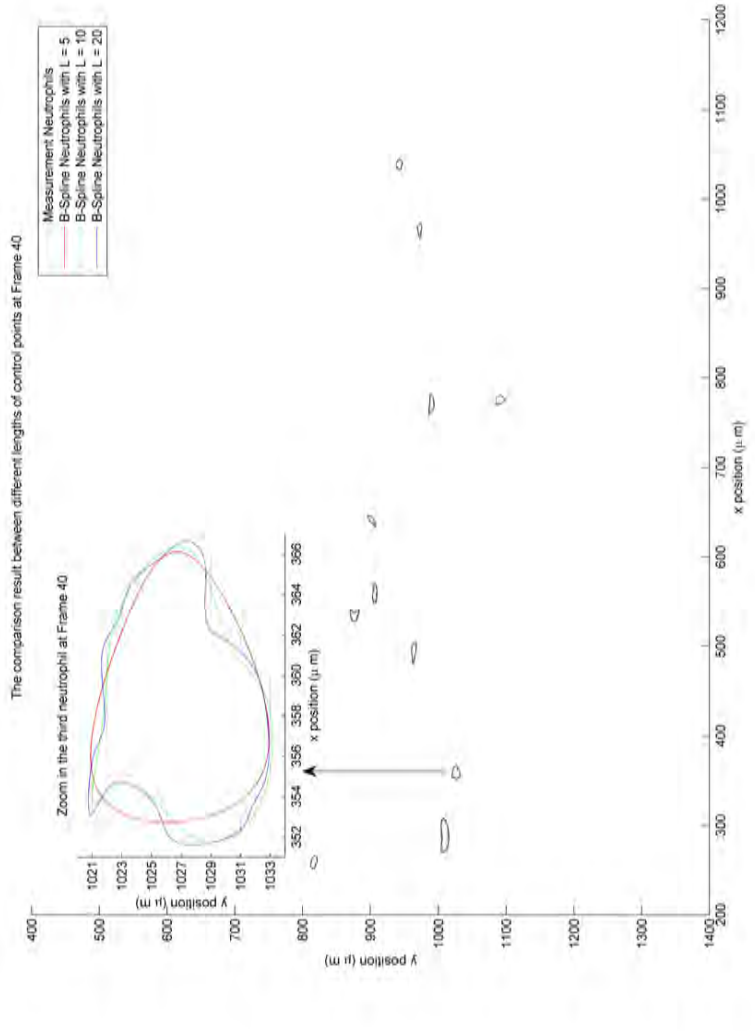

Figure 4.3: **The comparison of different control points**

The measurement shapes are compared with different numbers of control points. The third neutrophil is zoomed in. The red, green and blue broken lines represent L = 5, 10, and 20 B-Spline shape respectively, where L is the number of control points. The measurement shape is plotted in black line.

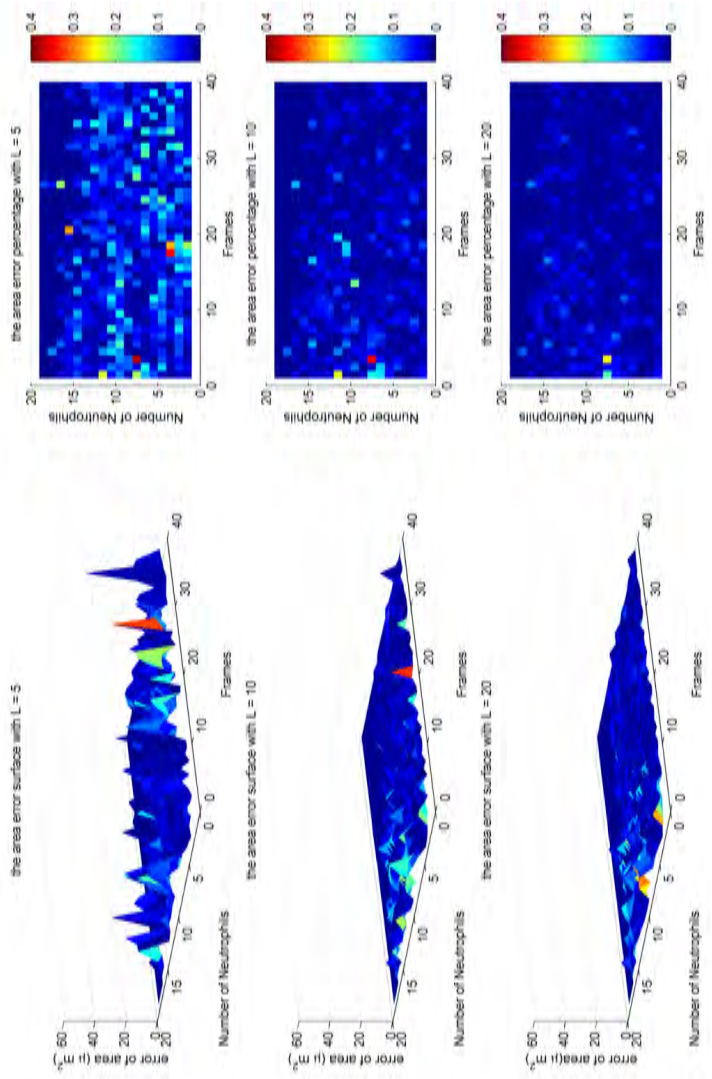

Figure 4.4: **The area error comparison of different numbers of control points**

The area error surfaces of different numbers of control points are shown on the left. The area error percentage is on the right. The colour scales and the coordinate scales are changed to the same. The top, middle and bottom figures are the case L = 5, 10, and 20 respectively.

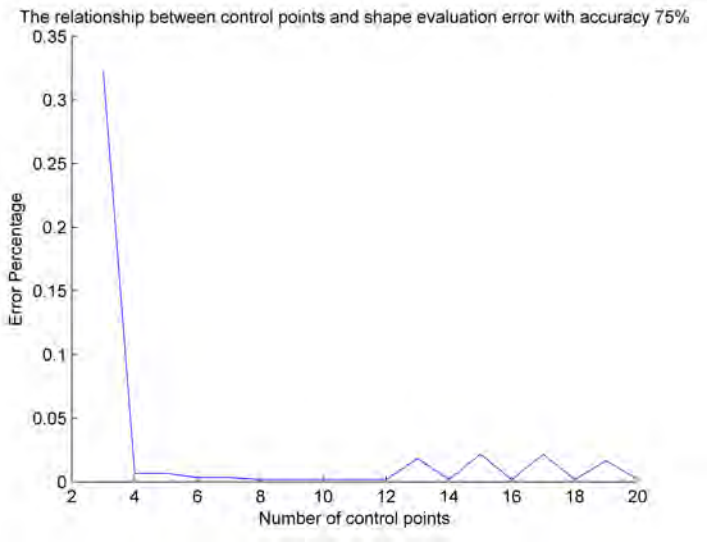

Figure 4.5: The relationship between the error percentage and the number of control points with 75% accurate rate

The error percentage is calculated by dividing the number of neutrophils whose error rate is larger than 25% by the total numbers of neutrophils. The smaller the error percentage is in the figure, the better the parameter of B-Spline estimator is. Therefore, the best choice of parameter L is any value in 8, 9, 10, 11, 12, 14, 16, 18 or 20 and the error percentage is near to zero.

of the other two. In addition, the number of area error percentage for the $L = 10$ case is acceptable. Therefore, based on the above reasons, $L = 10$ is selected as the parameter of B-Spline in both this chapter and other chapters in this thesis.

Based on the above results, another experiment is implemented to calculate the "best" number of control points. The number of control points from 3 to 20 is tested and the area error percentage is calculated and plotted, as shown in Figure 4.5 and Figure 4.6.

Figure 4.5 and Figure 4.6 show the relationship between the error percentage and the number of control points. The core idea of error percentage is dividing the number of neutrophils whose area error percentage is greater than acceptable by the total numbers. The different acceptable area error percentages lead to the different error percentages. In addition, the different error percentages lead to the different retrieved performances of B-Spline. Therefore, different area error percentages affect the performance of B-Spline. In Figure 4.5, the accurate rate is 75% which means the area error percentage is 25% and it is clear that the B-Spline with the number of control points from 4 to 20 have minimal error percentage. Therefore, all the numbers from 4 to 20 can be considered as the number of control

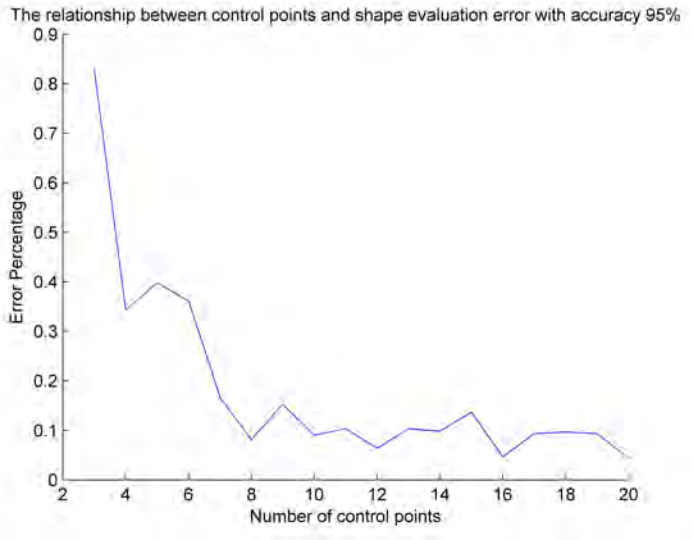

Figure 4.6: **The relationship between error percentage and the number of control points with 95% accurate rate**

The error percentage is calculated the same way as in Figure 4.5. The best choice of parameter L is either 16 or 20 since the accurate rate was up to 95%.

points. The best choice is 8, 9, 10, 11, 12, 14, 16, 18, or 20, since all have the smallest error percentage. Changing the area error percentage to 5%, the result is shown in Figure 4.6. It seems that the B-Spline with the number of control points from 8 to 20 has a small error percentage. The best choice is 16 or 20. The other numbers of control point are also acceptable. Based on both figures, the number of control point should be selected as 20. However, a larger control point number will lead to larger calculation cost. Therefore, the control point number is decided as 10, having considered the balance of both the calculation cost and the performance of B-Spline.

Although the B-Spline method is a good shape retrieved technique, the corresponding neutrophils in the process of frames are still unknown. Therefore, another technique, the nearest neighbourhood method, is needed.

4.3.2 Nearest Neighbourhood Filter Implementation

The NNF is used to find the same neutrophil between frames. This method is based not on the shape tracking but on the centroid tracking technique. The core idea is that the NNF predicts the centroid position in Frame $k+1$ by using the centroid information in Frame k, where k can be any positive integer. Then, the NNF

compares the predicted position with the measurement position in Frame $k+1$ and determines the minimum distance. If the minimum distance were less than the acceptable value, called threshold, the neutrophil would be considered as the same one in Frame $k+1$. The details can be found in Algorithm 4.2. Therefore, the NNF can effectively determine the index of neutrophils during the experiment. The NNF is implemented and neutrophils are indexed with the measurement shape in Figure 4.7.

In Figure 4.7, all the neutrophils are indexed by different colours and the measurement shapes are plotted with dots of the same colours. The No. 61 Neutrophil is zoomed in on the bottom right of the figure. It is clear that the NNF effectively indexes all the neutrophils. However, there is a drawback in neutrophil index tracking by using NNF. Neutrophils are always in the figure until frame k, and disappear in next frame $k+1$. Then they reappear in the following frame $k+2$. This is probably because of the pre-processing part. In pre-processing, medical photographs of neutrophils were dealt with via the double threshold method and during this process neutrophils might disappear then reappear. For example, in a frame k, one neutrophil was in the figure, but in the next frame $k+1$, it changed shape from horizontal to vertical. This might lead to its disappearance in that frame. However, in the following frame, $k+2$, it changed shape to the opposite direction, which led to the neutrophil's reappearance. The NNF cannot combine the indexes of the same neutrophil in the above situation, which makes the index number huge. Based on the above reason, the traditional NNF lost its effectiveness in neutrophil index tracking. Therefore, the INNF was proposed to solve the traditional NNF's consistent problem in neutrophil application. The INNF is implemented in Figure 4.8.

All the neutrophils are indexed with different colours in Figure 4.8. The measurement shape boundaries are plotted with dots and are the same colour as the index. No. 36 Neutrophil is zoomed in on the bottom right of the figure. Figure 4.8 proves that the INNF also has the ability to index all the neutrophils. However, by comparing Figure 4.7 with Figure 4.8, it is clear that No. 65, 39, 52, 72, 8, 9, 49 and 50 Neutrophils have exactly the same index. All the other neutrophils have different indexes. In addition, those neutrophils of different indexes have the same position information, meaning that those neutrophils are the same but with different indexes. For instance, No. 61 Neutrophil in Figure 4.7 and No. 36 Neutrophil in Figure 4.8 are the same neutrophil with the same measurement shape in the same frame and the same position information. It should be noted that No. 36 Neutrophil appeared from Frame 17 to Frame 28 just in the right area. Then it disappeared in Frame 29. In the next frame (Frame 30), at nearly the same

position and nearly the same shape neutrophil, which was No. 61 Neutrophil in the NNF, appeared. That neutrophil was still active in the right area until the end (Frame 40). Therefore, it is reasonable to believe that No. 61 Neutrophil in the NNF was the same neutrophil indexed 36 in INNF. It is necessary to combine the indexes to keep the same neutrophil tracking process completely. Moreover, the number of indexes in Figure 4.8 is less than that in Figure 4.7. Based on the above phenomenon, the INNF solves the traditional NNF problem and reduces the indexes. The most important contribution of the INNF is keeping the continuity of neutrophils. However, it still has some problems. It could not solve the problem when neutrophils are missing in 2 frames or more. Neither could it solve the problem of two neutrophils merging and separating.

4.3.3 Kalman Filter in Shape Tracking

Without the KF, both the B-Spline method and INNF method together can effectively track the shape changing. However, the process is complex. Therefore, the KF in shape tracking is considered as an effective tool to reduce noise and much of the computation. Firstly, KF is the best estimator in linear Gaussian cases to reduce noise and all the mathematical models in this chapter were linear Gaussian cases. Secondly, KF with B-Spline and INNF together formed a new framework, which reduced much of the computation; since by using the control points' information from the first frame in B-Spline, KF calculated the estimated control points instead of those control points from B-Spline from the second frame. The estimated control points from KF should be exactly the same as the control points from B-Spline in theory. Therefore, KF was a good substitution of B-Spline from the second frame. Furthermore, the velocity elements on the control points can also be read from the KF results directly, and therefore, the velocity on the boundary can be retrieved directly by using the new framework in the next subsection.

Figure 4.9 represents the absolute boundary point error comparison result between using KF in shape tracking and without using KF in shape tracking, to demonstrate the KF's effectiveness in reducing noise in shape tracking. All the absolute boundary point errors are calculated and compared in both methods in all the frames and for all the neutrophils. For each neutrophil, the absolute boundary point error is the summation of the absolute error from each boundary point. All the scales are adjusted to be the same to facilitate the comparison. The left figures and the right ones are the same but from different viewed angles. The top figures are derived from the method with KF and the bottom ones from the method without KF. It is clear that the surface of the estimated boundary error without KF on the left hand side is obviously sharper than that of the method

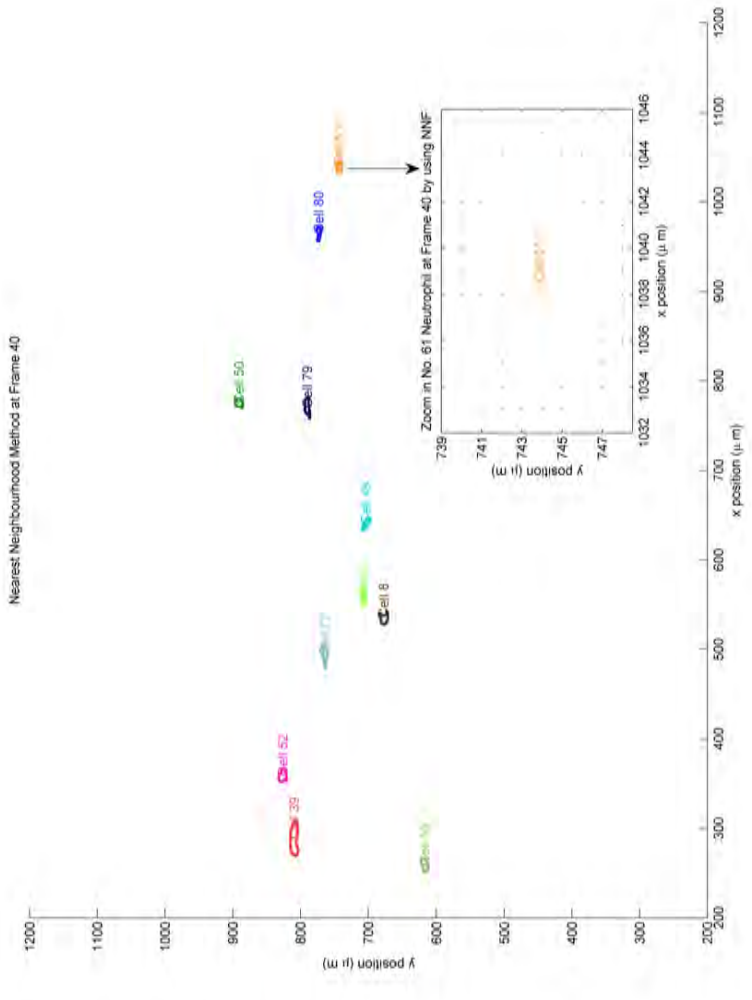

Figure 4.7: **The NNF is implemented and all the neutrophils are indexed with the measurement shape.**

All the neutrophils in each frame are indexed with different colours. No. 61 Neutrophil is zoomed in on the bottom right of the figure. The measurement shape of No. 61 Neutrophil is plotted with dots.

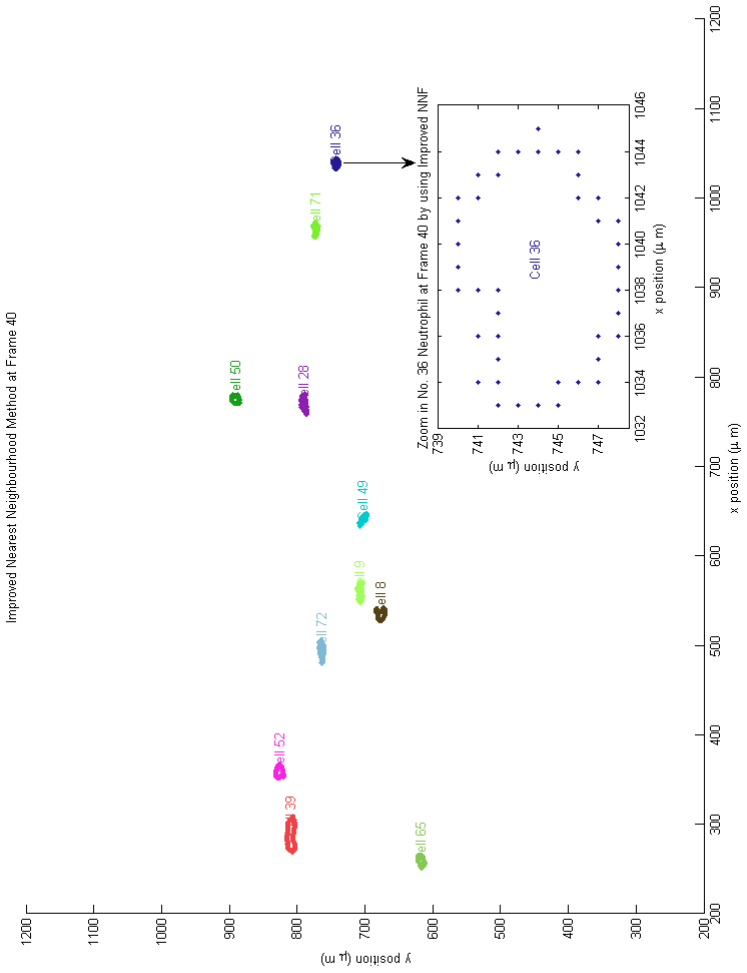

Figure 4.8: The improved NNF is implemented and all the neutrophils are indexed with the measurement shape.

All the neutrophils in each frame are indexed. If the index of neutrophils is the same with Figure 4.7, then the colour of that neutrophil must be the same, such as No. 39 Neutrophil. No. 36 Neutrophil is zoomed in on the bottom right of the figure. The measurement shape is plotted with dots.

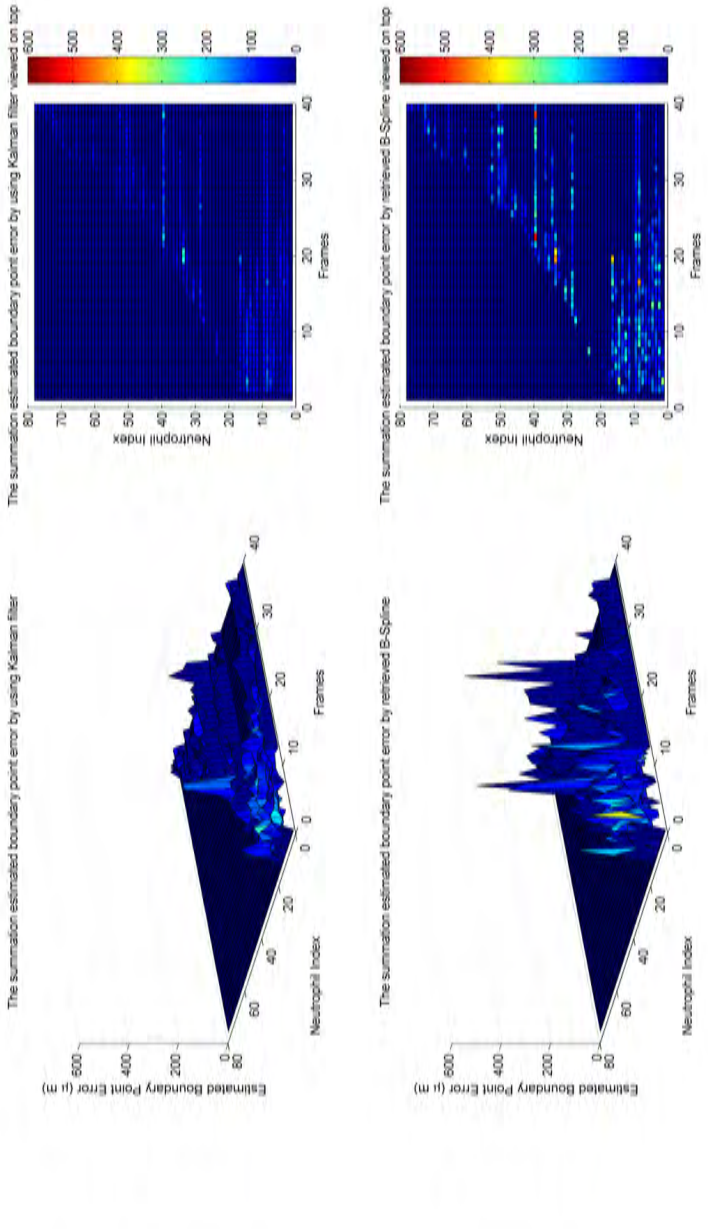

Figure 4.9: **The comparison of the absolute boundary point error between using KF and without using KF**

The KF combining with B-Spline and INNF is implemented and the absolute estimated boundary point error is plotted on the top. The error surface of B-Spline combining with INNF is plotted at the bottom. Both the left hand side are the surface and the right hand side are the same but viewed on top. All the figures are adjusted to the same scales. The absolute estimated boundary point error is the summation of the absolute value of the difference between the measurement boundary point position and the estimated boundary point position.

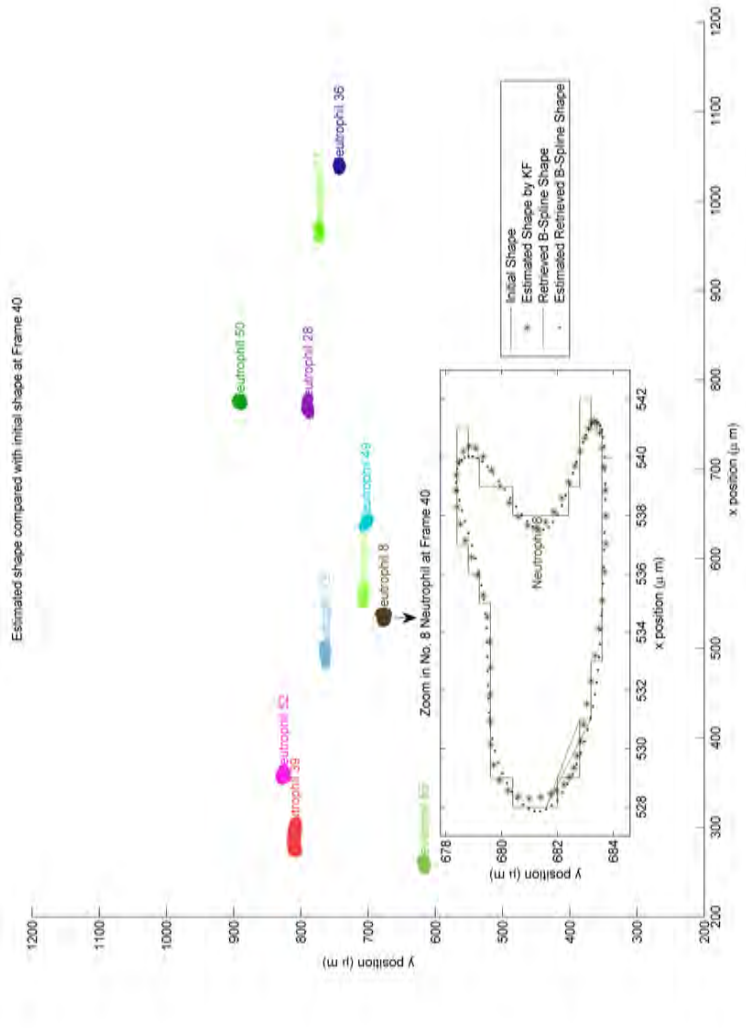

Figure 4.10: **The comparison between different retrieved shapes and the measurement shapes**

All the neutrophils are plotted in different colours. No. 8 Neutrophil is zoomed in on the bottom left area. The measurement shape (the initial shape) is plotted with solid line. The estimated shape by KF is plotted with star. The retrieved shape by B-Spline is plotted with broken line. The estimated retrieved shape by B-Spline is plotted with dots. It should be noted that the retrieved shape is the shape from B-Spline retrieved technique, while the estimated retrieved shape is the shape calculated from B-Spline output equation.

with KF. From the right hand side, the method with KF (the top right figure) has an obviously smaller number of errors than that of the method without KF (the bottom right figure), since the colour indicates the low colour scale in the top right figure. Both the left and the right figures reach the same conclusion, that the new framework with KF in shape tracking can effectively reduce the noise. It should be noted that there are some neutrophils whose estimated boundary point error is always zero, in all frames. For example, No. 18 Neutrophil was one with zero error for all frames. This did not mean that the tracking was perfect. No. 18 Neutrophil never appeared due to the INNF method. In INNF, it was just No. 5 Neutrophil disappearing in Frame 7 and appearing again in Frame 8. Therefore, the INNF combined both trajectories and considered No. 18 Neutrophil not to exist. As such, the new framework with KF in shape tracking does reduce the noise enormously. The comparison of shape is illustrated in Figure 4.10.

Figure 4.10 compares the measurement shape boundary (initial shape) with the other shapes from different methods. All the neutrophils are tracked with different colours. No. 8 Neutrophil is zoomed in at the bottom left. The solid line, star, broken line, and dot present the measurement shape (initial shape), the estimated shape by KF, the retrieved B-Spline shape, and the estimated retrieved B-Spline shape, respectively. It should be noted that the retrieved shape and the estimated retrieved shape are different techniques. The retrieved shape is the shape from the B-Spline retrieved technique and the estimated retrieved shape is the shape calculated from the B-Spline. In other words, the retrieved shape uses a variable s changing from 0 to 1, and then the shape position is calculated as a function of the same variable s. Conversely, the estimated retrieved shape is calculated by the formula $\boldsymbol{z}_k = \boldsymbol{H}\boldsymbol{x}_k$. It seems that various methods have nearly the same tracking effect. However, if all the absolute errors on the boundary are added together, the difference is huge, as shown in Figure 4.9. Based on both Figure 4.9 and Figure 4.10, the KF in shape tracking can effectively reduce the noise and form a new framework to reduce the calculation. In addition, the KF with INNF in shape tracking can keep the neutrophils continuous by adding the estimated control points and positions when neutrophils disappear. Furthermore, the velocity, either on the boundary or on the control points, can be retrieved or observed directly from the estimated Kalman state. Therefore, the new framework is useful and effective in shape tracking compared with the traditional method.

It should be noted that using the Kalman smoother instead of the KF, the tracking result should be more effective, since the Kalman smoother reduces the shape tracking noise obviously, in theory, by using the backward method when neutrophils disappear.

4.3.4 Velocity Estimation

The core idea of the velocity estimation is using the KF or the Kalman smoother to observe the velocity on the control points. Using the velocity on the control points estimates the velocity on the boundary. As mentioned above, there are two methods. The comparison between these two is shown in Figure 4.11. Given that the velocity sizes on the boundary in these two methods are different, Figure 4.11 represents the shape of velocity in order to compare them.

It is clear from Figure 4.11 that both methods have the same velocity shape. This means both the velocity estimators obtain the same results if the size is not considered. The velocity shape comparison is similar to the comparison between the sampling and the continuous time signal. For instance, the sine wave is a continuous time signal, and the sampling of the sine wave can retrieve the shape but with different sizes (only having values on the sampling time). Therefore, the second method, B-Spline retrieved technique, is used in this current work and in the following chapters, since the velocity has the same size and is easy to compare and analyse. The velocity estimation is implemented and the result is in Figure 4.12.

Figure 4.12 illustrates the velocity estimation based on the neutrophil shape boundary in Frame k and Frame $k+1$. The solid line is the shape boundary in Frame k, and the dot-dash line is the shape boundary in Frame $k+1$. All the neutrophils are coloured and tracked with velocity on the boundary pointing to the next frame. No. 28 Neutrophil is zoomed in at the middle bottom of the figure. It should be noted that in the first few frames, the velocity estimation was not accurate. This is probably because that the KF needs some frames to modify the velocity estimator. After that, the estimated velocity pointed from the boundary in Frame k to the boundary in Frame $k+1$ well, such as for the zoomed in neutrophil. The estimated velocity, to some degree, represents the shape changing and proves that the new framework with KF has the ability to track the neutrophils well. Therefore, the shape boundary changing can be transferred to the velocity shape, although the relationship is not bijection. The comparison between boundary shape and velocity shape is represented in Figure 4.13.

It is clear that when most of the velocity shapes are located in the first quadrant, the shape moves toward the upper right. As the movement on x direct is obviously greater than on y direct, the neutrophil mainly moves to the right and with small shape changing on the top. However, there are some velocities on the fourth quadrant, which means some parts of the neutrophil move to the lower right. Generally, the velocity shape can represent the trend of boundary shape changing. In addition, it proves that the chemotaxis field, which is a time variable

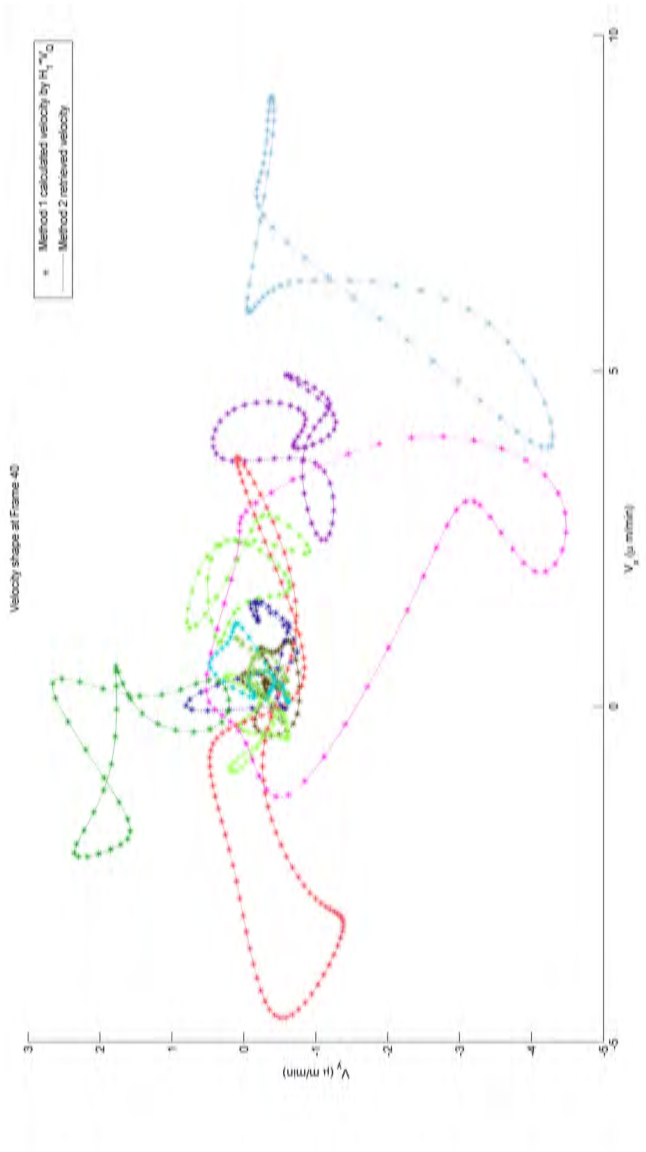

Figure 4.11: **The comparison of boundary velocity shapes based on two methods**

The star represents the first method, which is calculated by the transformed matrix ($\boldsymbol{H_1}$) of the observation matrix multiplying the part of the control points ($\boldsymbol{V_Q}$). The dot-dash line is the second method which is calculated by the retrieved technique of B-Spline. The different colours represent the different neutrophils. They are the same colour as mentioned above. For example, the red is No. 39 Neutrophil.

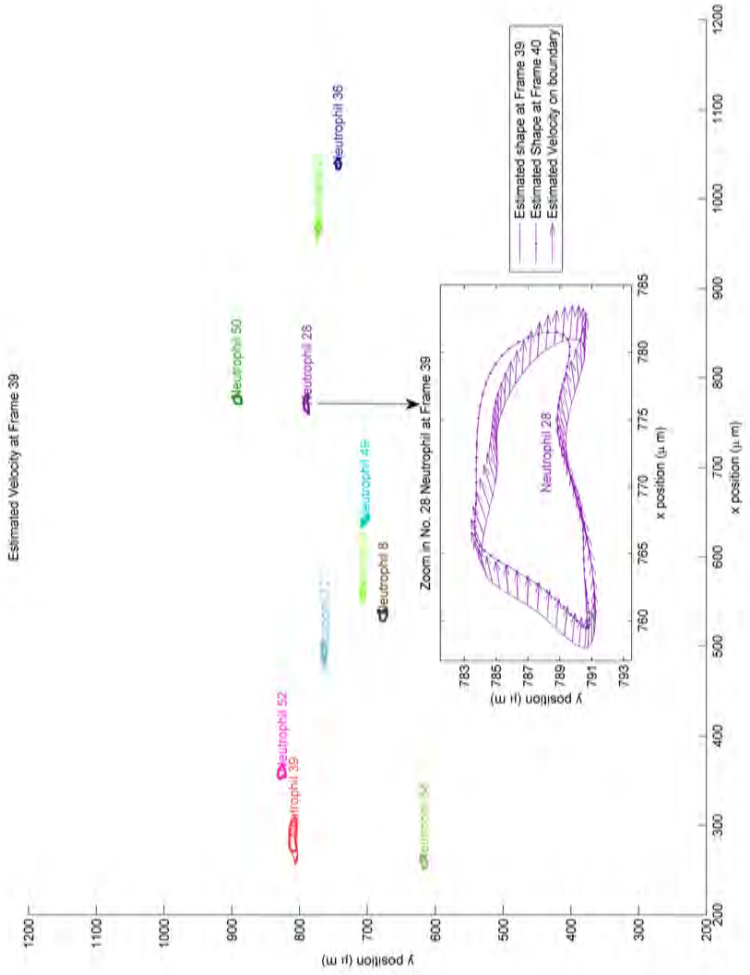

Figure 4.12: Neutrophils' shape boundary velocity estimation

All the neutrophils are estimated with boundary velocity in different colours. No. 28 Neutrophil is zoomed in at the middle bottom. The solid line is the estimated shape in Frame 39 and the dot-dash line is the estimated shape in Frame 40. The boundary velocities are plotted with the same colour and with arrows starting from Frame 39.

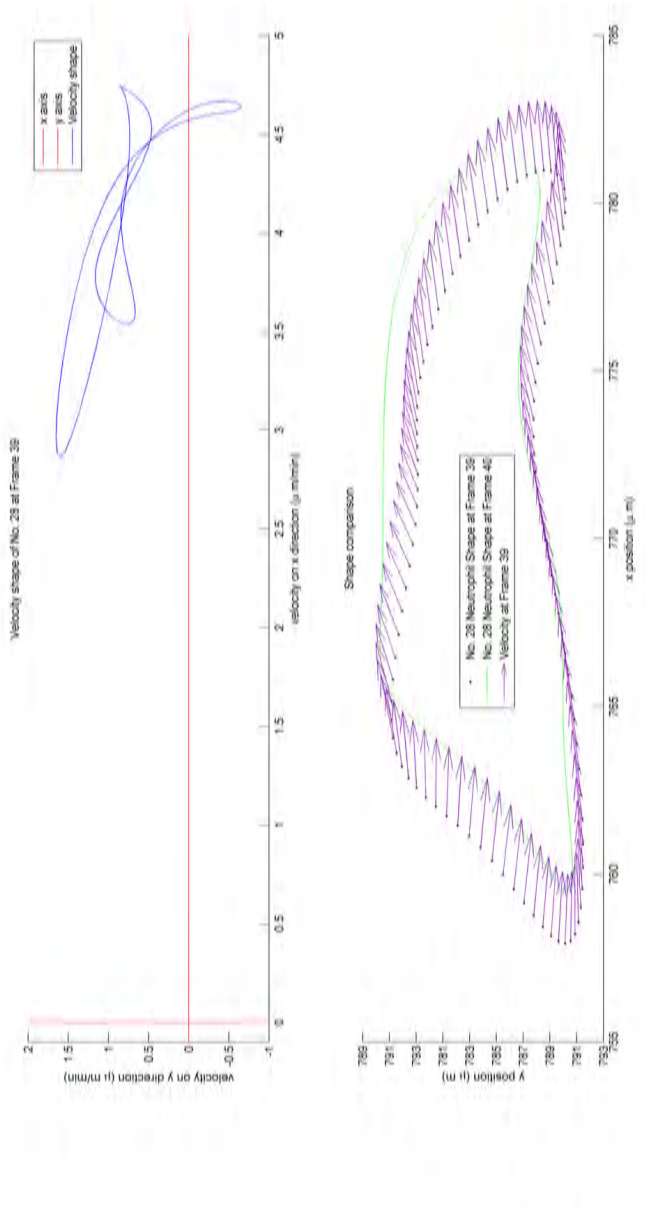

Figure 4.13: The comparison between the velocity shape and the estimated boundary shape

The velocity shape is plotted on the top figure. The x and y coordinate are plotted in red. The shape comparison is plotted at the bottom figure. The black dotted line is the shape in Frame 39 and the green broken line is the shape in Frame 40. The velocity on the boundary is plotted as arrows starting at the shape boundary in Frame 39.

function, exists. Furthermore, it also proves the conclusion of the previous chapter, that the chemotaxis distributes asymmetrically. It should be noted that the chemotaxis field is obviously effective to the active neutrophils. No. 28 Neutrophil is an active neutrophil, which meant that it is alive and moves to different positions with different times by the chemotaxis. With the help of velocity on the boundary, the chemotaxis field can be determined more clearly than the field only considering the centroid velocity. Compared with the centroid velocity, the shape contains more information. For example, when the pseudopod is formed, the shape velocity is used for analysis. For this reason, the shape velocity is significant in shape tracking.

4.4 Conclusion

This chapter has focused on shape tracking. At the beginning, background knowledge was introduced. Several commonly used algorithms were introduced, such as B-Spline. The results were listed after the algorithms. Parameters of algorithms can affect the results. Therefore, a test was performed to determine the optimal parameters. The low resolution data was implemented in all the algorithms used in this chapter. The boundary velocities were observed and were more useful in shape boundary dynamics. Innovation and contributions are as follows. Firstly, the INNF solved the problem when neutrophils disappeared and reappeared in the next frame. The combination of the KF, the INNF, and the B-Spline formed a simple framework to track the dynamics of neutrophils and other cells. In addition, the framework reduced a huge number of calculations by using the estimated control points instead of the control points from B-Spline and simplified the process of shape tracking. Finally, the velocities of the shape boundary were estimated. Compared with the estimation of positions, the velocities contained much more information and it is believed that this information would help people to better understand the mechanism.

Chapter 5

Characterisation of Neutrophil Modes by Shape Descriptors

The previous chapter mainly focused on shape tracking. A new framework was established to solve the similar problem of cell tracking. This chapter is primarily concerned with shape analysis, which analyses the data information of shape tracking. The commonly used method is the FD, the major method used in this chapter. Some descriptors are also specified in this chapter, either in combination with the FD or using other parameters.

5.1 Introduction

Shape analysis is a method that automatically analyses shape characteristics. For example, neutrophils can change their shape with time, however, using the shape analysis method, several kinds of features can be extracted, such as dead neutrophils and active neutrophils. The process of extracting the characteristics is also called the shape descriptor, which can, usually, represent the original objects' characteristics completely. There are several methods to describe the characteristics of a neutrophil shape.

The most important aspect of shape representation is the effective shape feature information on the boundary and the interior content [164]. A good shape descriptor usually means a method whereby the retrieval has a rotated, translated and scaled shape invariance with a low computation complexity. However, sometimes, a shape descriptor with only scaled shape invariance is still useful and effective.

Generally, there are two classes of methods for shape descriptors, namely Contour-Based Methods and Region-Based Methods. The difference between these

two is from where the shape features are extracted. Contour-based methods usually extract shape features from the contour only, while region-based methods extract them from the region. According to Zhang and Lu, the shape descriptor can be divided into structural and global approaches in both contour-based methods and region-based methods. The difference is that structural approaches represent the shape by segments while the global approaches do it as a whole. Several methods used in this thesis will be introduced separately in the following sections.

5.1.1 Fourier Descriptor

As noted above, a large number of descriptors exists, however, in this section, the FD was used in order to analyse the mobility, as per the advantages determined in Chapter Two.

The FD uses the core idea of Fourier transform with the following formula:

$$F(y) = \int_{-\infty}^{\infty} f(x) e^{-2\pi j x y} dx, \forall y \in R \tag{5.1}$$

where $F(y)$ is the Fourier transform, and $f(x)$ is the initial function.

Usually, Fourier transform is used in one-dimensional areas, however, in this case, two-dimensional Fourier transform was required. According to the linear properties of Fourier transform, the two-dimensional position $\boldsymbol{z} = [x, y]'$ can be rewritten as a complex number $z = x + iy$, where $i^2 = -1$, and in this case, the two-dimensional position can be addressed using Fourier transform, too. The formula is as follows.

$$F(z) = F(x) + iF(y) \tag{5.2}$$

The algorithm of two-dimensional FD is shown in Algorithm 5.1

Algorithm 5.1 Two Dimensional Fourier Descriptor Algorithm

for $iFrame = 1$ to N **do**
  Read in the neutrophils boundary shape $\boldsymbol{u = x + iy}$
  Calculate FFT $\boldsymbol{uu = fft(u)}$
  Delete the first element value $\boldsymbol{uuT = [0; uu(2:end)]}$ and normalise the result from the second element $\boldsymbol{uuTN = uuT./abs(uuT(2))}$
  Get the real and image part of the result $\boldsymbol{uuTNR = [real(uuTN); imag(uuTN)]}$
end for

5.1.2 Height and Radius Descriptor

In addition to the FD, the height descriptor and radius descriptor were also used. Based on the test result of adding a noise point (the details of the test have not been included in this current work) and the retrieved shape characteristic, there is another way to transform the formula with the radius and angle. The details are as follows.

$$\begin{aligned} \boldsymbol{Z} = \boldsymbol{H}\boldsymbol{Q_c} = R(t,s)[\cos(wt) + i\sin(wt)] = R(t,s)e^{iwt} \\ R(t,s) = \sqrt{[x(t,s) - x_c(s)]^2 + [y(t,s) - y_c(s)]^2} \end{aligned} \tag{5.3}$$

where $\boldsymbol{Z}$ is the retrieved shape boundary; $\boldsymbol{Q_c}$ is the control points from the previous chapter; $\boldsymbol{H}$ is the transfer matrix; t is the points vector (can be also retreated as time vector) with the values $[1, 2, \ldots, 101]$; s is the frame with the values $[1, 2, \ldots, 40]$; $R(t,s)$ is the radius, and it is a time and space variant parameter; $\boldsymbol{wt}$ is the speed of a point running around the shape and it is a vector with the values $[360^\circ/100, 360^\circ * 2/100, \ldots, 360^\circ * 101/100]$; therefore, w can be calculated as 3.6° or $\pi/50$; $x(t,s)$ is the x-axis boundary position; $y(t,s)$ is the y-axis boundary position; $x_c(s)$ and $y_c(s)$ are the "centroid" position of the shape and they can be the real centroid positions in frame s, or not. This depends on which method is used for analysis and this method will be discussed in the future work section. Furthermore, they are time variant parameters, that is, they may be different with different frames.

An interesting result is produced. As $\boldsymbol{Z}$ is the function of t and s, there are two ways to reduce the formula. One is fixing t and the other is fixing s. They are as follows.

When s is fixed, which means in the same frame, all the retrieved shape boundary points of the FD have the following formula.

$$Z(t) = \sqrt{[x(t) - x_c]^2 + [y(t) - y_c]^2}e^{iwt} \tag{5.4}$$

When t is fixed, which means at the same retrieved shape boundary point but in different frames, the FD sequence has the following formula.

$$Z(s) = \sqrt{[x(s) - x_c(s)]^2 + [y(s) - y_c(s)]^2}e^{iwt} \tag{5.5}$$

The first case shows the relationship between all the retrieved shape boundary points in the same frame, while the second indicates the relationship between the same point in different frames. If the FD is used, the first case is just the FD in each frame and the second one is the FD at the same frequency but in different frames.

The first case is a common FD and as an analysis tool is used widely. The second one is highly significant, because when the same point on the boundary has the same position and the same centroid position, the FD should remain the same. In this case, this method can be used to detect whether the neutrophil changes the shape or not.

According to the properties of Fourier transform, the formula above can be changed in the following way.

$$\begin{aligned} F(\boldsymbol{Z}) &= F(R(t,s)e^{iwt}) \\ &= F(R(t-w_0,s)) \end{aligned} \tag{5.6}$$

where $w_0 = \frac{w}{2\pi} = \frac{1}{100}$.

For the first case, the above formula can be simplified as $F(\boldsymbol{Z}) = F(R(t-w_0))$, and $F(\boldsymbol{Z}) = F(R(s)e^{iwt})$ for the second one, which is a shift of $F(R(s))$.

The formula above has the meaning that Fourier transform of the retrieved shape equals to Fourier transform of the radius with only the shift of shape boundary points. However, the radius is a discrete time variable, which means $R(t-w_0,s)$ has no physical meanings. In addition, because the shift part w_0 is too small, usually, $R(t-w_0,s) \approx R(t,s)$. Therefore, it follows $F(\boldsymbol{Z}) \approx F(R(t,s))$. The advantage of this derivation is that it reduces the complexity of computation. That is, changing two-dimensional FD into one dimension. Further testing and verification based on the radius description will form part of future work.

In addition to the radius descriptor, the height descriptor is introduced. The core idea of the height descriptor is considering the minimum of the logarithmic FD as a feature. From the result of the test, it is clear that each FD of neutrophils has a minimum and those minima are different for different neutrophils. The round has the lowest minimum in the logarithm of the FD figure. In the figure, the minimum appeared as the height of the FD, as such, it was named as the height descriptor.

As a descriptor, the height descriptor describes the height feature, which is different for different neutrophils. Although it is not clear whether the height descriptor has any other function, it describes the feature of neutrophils and it has been used in comparison with other features, such as the area, in the results section.

It should be noted that, both the radius and height descriptor are independent on control points. They cannot establish a link with the above chapters. Therefore, both were considered as an assistant analysis tool compared with the FD, although both used Fourier transform.

5.2 Results

It should be noted that all the methods in this chapter have been used for shape analysis. Therefore, this chapter has a close connection to the previous chapter one.

5.2.1 Fourier Descriptor Implementation

It should be noted that the data is the low resolution data, which is exactly the same data as that from the INNF in the previous chapter. In addition, FFT is used to reduce the calculation. This is because the FFT result is the same as that of Fourier transform for the real and image part. Based on the above two points, the FD of No. 9 Neutrophil is implemented in Figure 5.1 and Figure 5.2.

Figure 5.1 shows the implementation of FD for a two-dimensional neutrophil shape. The shape boundary of the same neutrophil in different frames is compared on the left hand side, and the logarithms of FD are plotted and compared on the right hand side. It is clear that the difference between shapes in Frame 39 and Frame 40 is minimal, however, the FD demonstrates an enormous difference. Furthermore, the shape boundary sizes of the same neutrophil in different frames are different. Therefore, it would be unreasonable to compare the FD, unless the size had been the same.

Figure 5.2 gives a clear view that the sizes of the same neutrophil in different frames are different. It is clear that No. 9 Neutrophil in all 40 frames has different frequency sizes. The colour bar is the logarithmic values of FD. It is unreasonable to compare the FD because of the different sizes. The reason for the sizes being different is probably as follows. The FD uses the initial shape boundary information, and in this case, the size of the shape cannot be guaranteed to be the same. Furthermore, noises exist in the initial data. Based on these reasons, further research on reducing noise, comparing FD with the same size, and combining the results from the previous chapter is needed.

5.2.2 Testing and Verifying

Before testing the hypotheses, a small and simple example, called a round test, is introduced. It is a round shape in two-dimension dealt with the FD.

Figure 5.3 gives the result of the round test. The round shape is plotted in red on the top figure. The corresponding FD is plotted in the middle figure. There is an impulse signal near zero and it is not an empty figure. As the impulse is extremely difficult to distinguish and analyse, the logarithmic FD is plotted at the

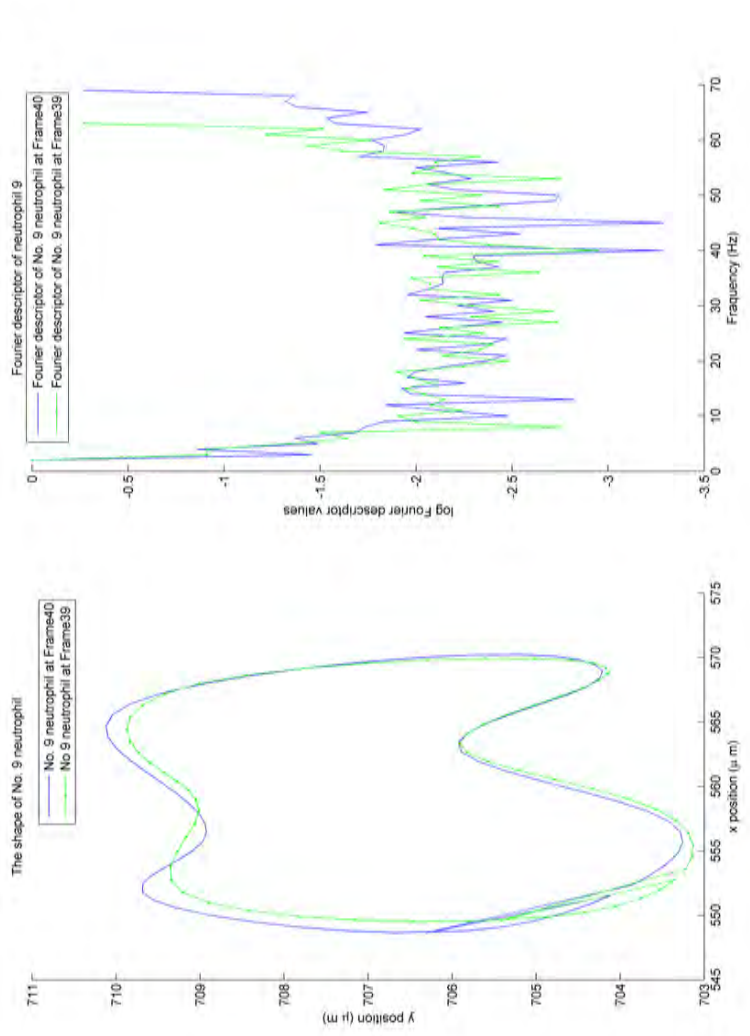

Figure 5.1: **The FD is implemented and compared in different frames.**

The shape of No. 9 Neutrophil is plotted on the left hand side. The blue solid line is the shape in Frame 40, and the green dot-dash line is the shape in Frame 39. Both of the FDs are plotted on the right hand side. Because the sizes of the same neutrophil at different frames are different, the sizes of FDs are different. The logarithms of FD are calculated in the y direction.

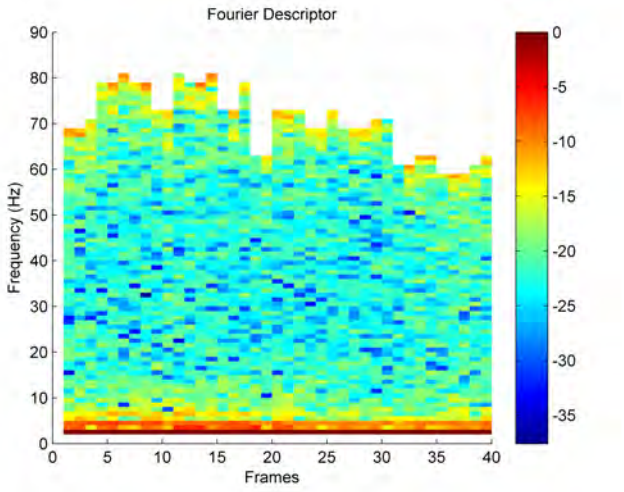

Figure 5.2: The FD distribution of No. 9 Neutrophil

The FD distribution of No. 9 Neutrophil is plotted for all 40 frames. The colour bar is the logarithmic values of FD. Because the same neutrophil in different frames has different numbers of boundary points, the maximum frequency of the same neutrophil is different.

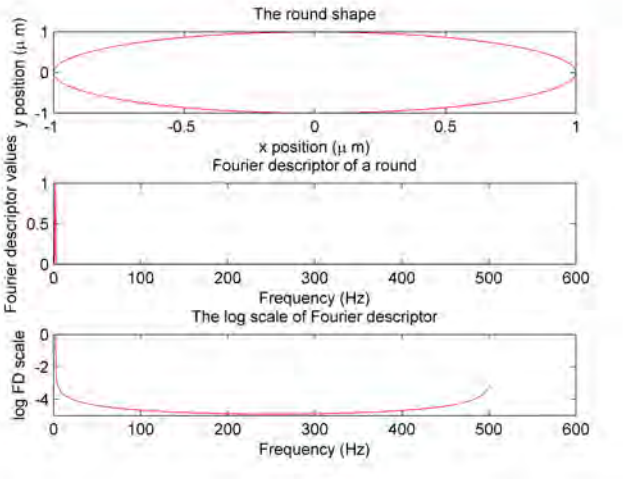

Figure 5.3: Round test

The top figure is the round shape. The middle figure is the FD of the round. The bottom one is the logarithmic values of FD. It is clear that the FD is an impulse signal around zero. It is extremely difficult to distinguish. Therefore, the logarithm is used.

bottom. It is clear that the logarithmic FD of the round without noise is a highly smooth curve. There is no noise on the smooth curve, which means that there is no high frequency noise on the shape. Therefore, this round logarithmic FD is assumed as the basic FD, and all the others will compare both the shape and the FD with it.

Effects of Noise

The aim of this part was to check whether the high frequency noise on the shape boundary affected the FD or not. Usually, when a noise occurs, it always distributes all around the shape not on a point and the magnitude is always minimal. It is usually considered as high frequency noise on the entire shape boundary. In this test, the amplitude of noise sequence was decided as $[0.01\%, 0.03\%, 0.1\%, 0.3\%]$. The shape comparison is shown at the top of Figure 5.4. Since the noise is minimal, it is difficult to distinguish the difference in shape. However, a very huge difference exists on the FD shown at the bottom of Figure 5.4. It is clear that when the noise magnitude is kept in the range between 0.01% and 0.03%, the logarithmic FD has nearly the same height as the round FD, only with a very high frequency noise on the FD. The 0.03% noise has a larger amplitude high frequency noise compared with the 0.01% noise. However, when the noise magnitude increases to the range between 0.1% and 0.3%, it is clear that the FD height is higher than that of the round. The height and magnitude of the 0.3% noise are even higher and larger than that of the 0.1% noise. Therefore, the greater the noise is, the greater the height of the logarithmic values of FD.

However, the test case was done under auspicious conditions. Real neutrophils have irregular shapes, and small irregular parts, called pseudopods, were considered as noises added on the local shape, which would cause the FD to make obvious changes and never become smooth as shown in the round test. Therefore, undertaking the test on the local shape is needed.

Effects of Local Shape Changing

Except for the effects of noise, neutrophil pseudopods can also change the FD. When a pseudopod occurs, it could be considered as several points, convex or concave, on the initial shape. Therefore, in this part, a one-point-pseudopod, three-point-pseudopod, five-point-pseudopod, and seven-point-pseudopod are tested to determine whether the pseudopod influences the FD.

Figure 5.5 illustrates the results of pseudopod effects. FDs of one-point-pseudopod, three-point-pseudopod, five-point-pseudopod, and seven-point-pseudopod are pre-

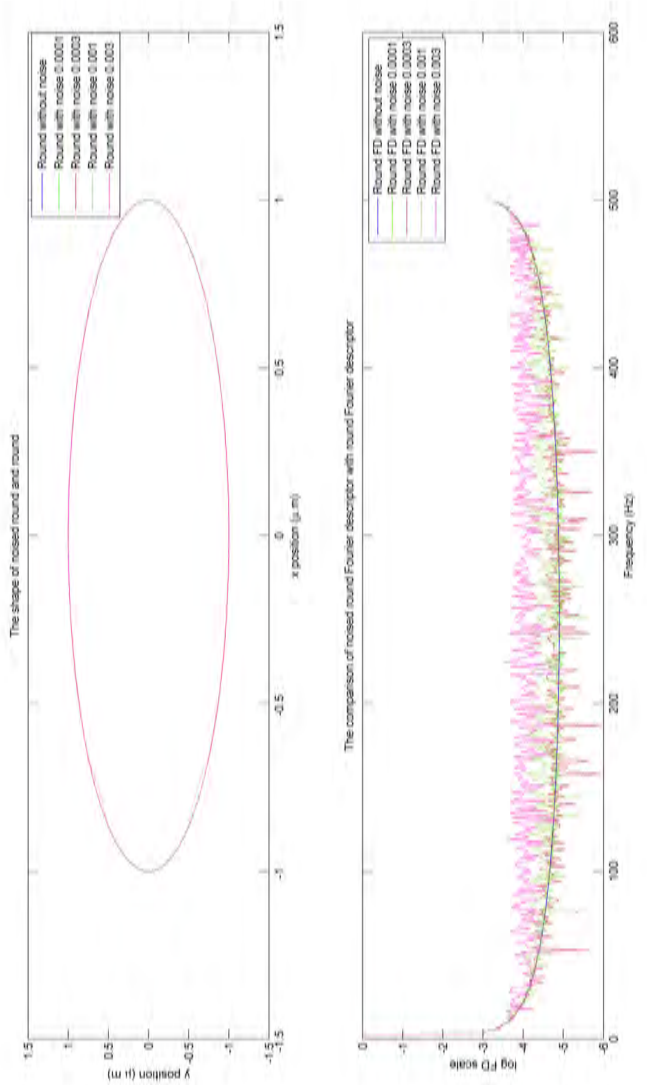

Figure 5.4: The test of noise effects

A sequence of noise is added on the round and the logarithmic values of those FDs are plotted and compared. On the top of the figure, all the shapes of both the noised round and the round without noise are compared and plotted with different colours. Because of the small amplitude of the noise, there are minimal differences. However, the differences of the logarithmic values are large as shown at the bottom. The logarithmic value of a round FD is a smooth curve, while the noised one is not smooth even if the noise is minimal. In addition, the heights of the logarithmic value of FD are different. It seems that the greater the noise is, the greater the height of the logarithmic values is.

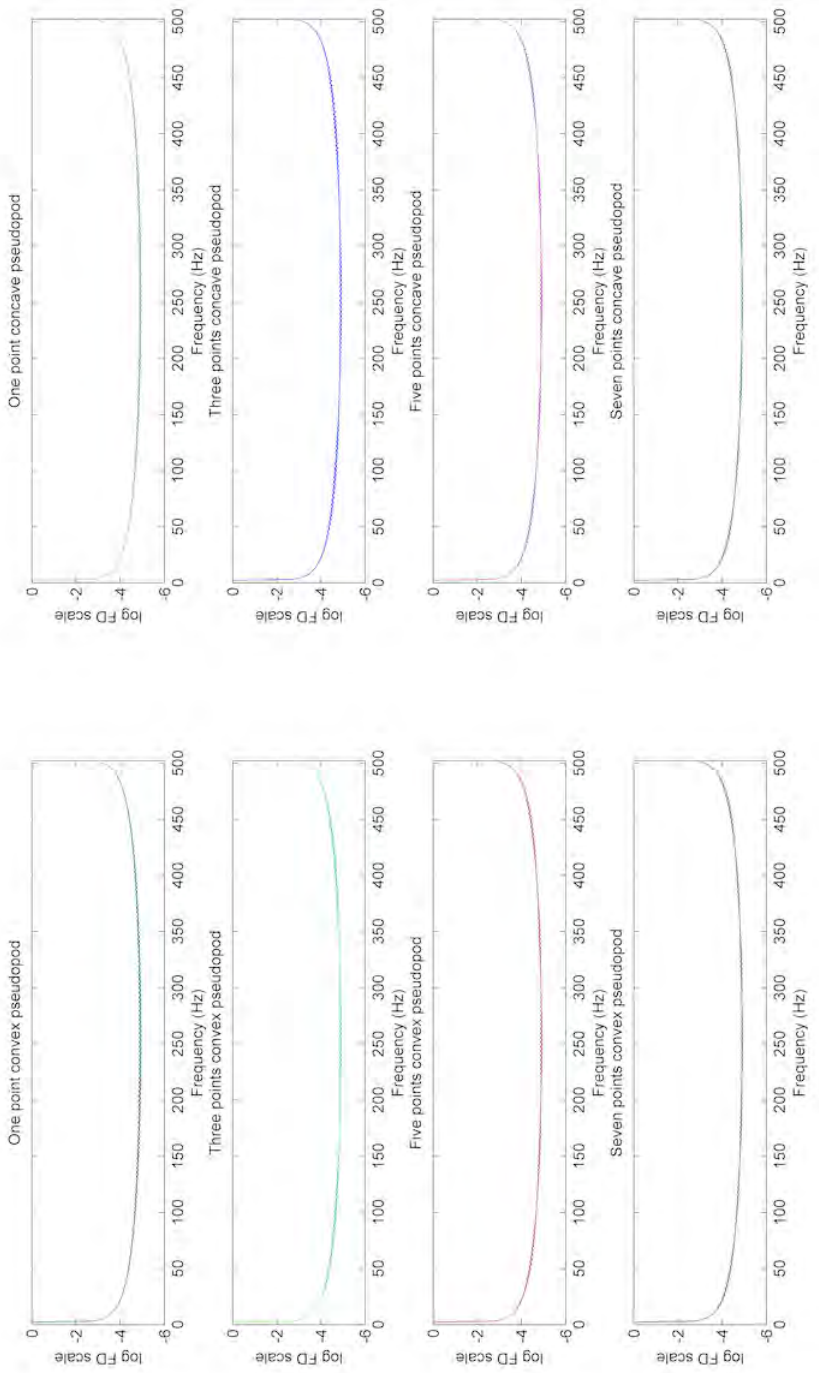

Figure 5.5: The test of local shape changing effect

When a pseudopod occurs, it can be considered as several points convex or concave on the initial shape. Therefore, the comparison results are shown in figures. All the figures on the left hand side are the comparison of the logarithmic values of FD between the convex pseudopod and the initial round, and all the figures on the right are the comparison of the logarithmic FD between the concave pseudopod and the initial round. The first row on the top is the case of the pseudopod with only one point. The second row on the top is the case of three-point-pseudopod. The third row on the top is the case of five-point-pseudopod, and the bottom row is the case of seven-point-pseudopod.

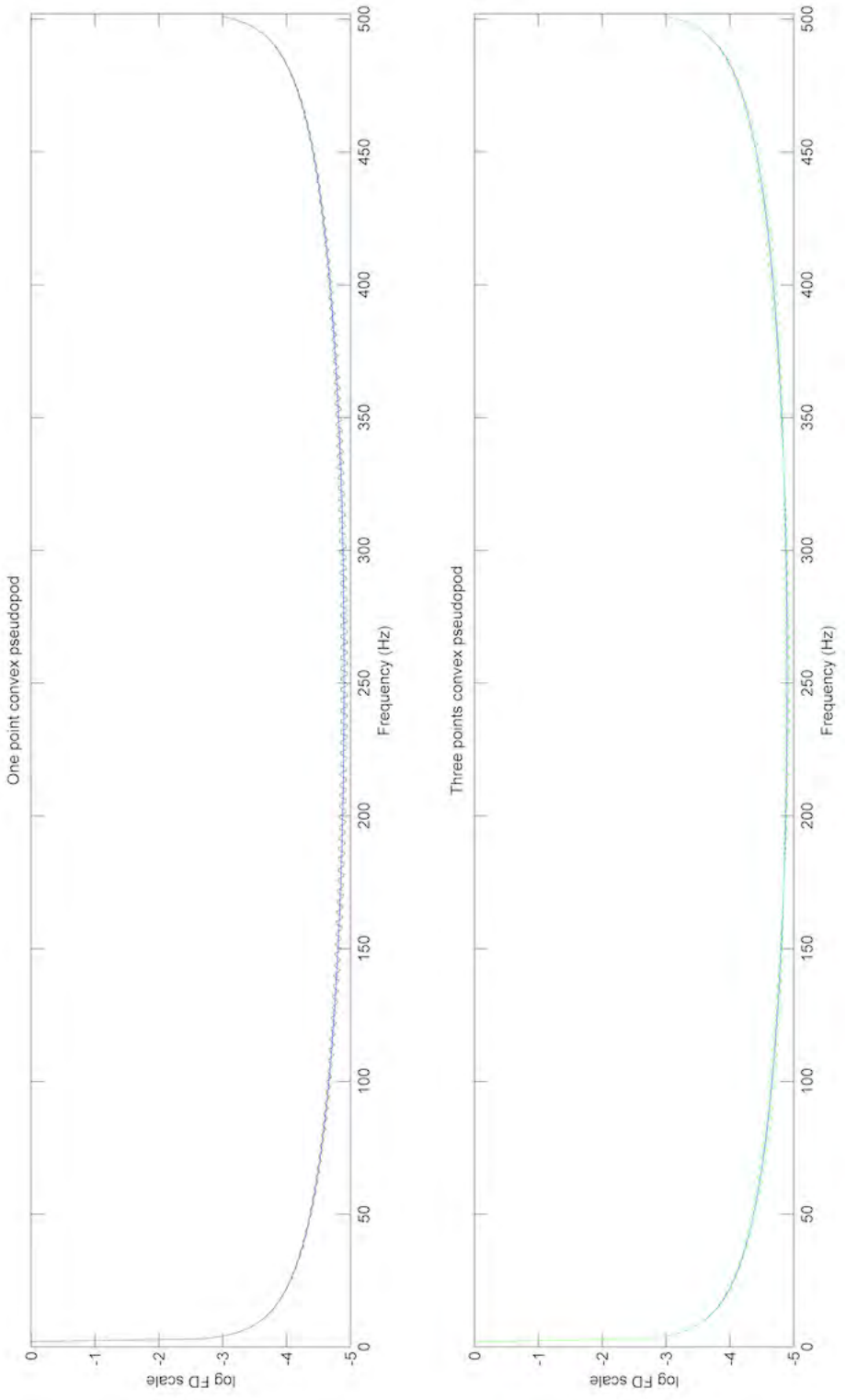

Figure 5.6: The drawing of partial enlargement of Figure 5.5

The part of Figure 5.5 is zoomed in to illustrate the comparison clearly. It only compares with the different numbers of points, because the convex pseudopod and concave pseudopod have the same logarithmic FD.

sented in the first row, second row, third row, and the last row of Figure 5.5 respectively. The left hand side of Figure 5.5 is the convex pseudopod and the right hand side is the concave pseudopod. The logarithmic FD of pseudopods is compared with the round FD values in all the figures. From Figure 5.5, it is easy to draw several conclusions as follows.

Firstly, no matter what kind pseudopod (convex or concave), the logarithmic FD of the same points pseudopod remains the same. This conclusion can be extended to infer that when the radius of neutrophil changes, either increasing or decreasing at the same rate, the FD remains the same. This is related to the radius descriptor and will be presented in the section on future work. Secondly, it is clear that for a one-point-pseudopod, there is only one vibration wave area. The shape of this vibration wave seems to be a damped vibration whose largest amplitude is at the mid frequency of FD. However, there are three vibration wave regions in the three-point-pseudopod and each appears to be a damped vibration whose largest amplitude is at the mid frequency of that wave region. The one-point-pseudopod and the three-point-pseudopod are zoomed in, in Figure 5.6, to compare them clearly. A similar conclusion can be drawn from the five-point-pseudopod and seven-point-pseudopod. Therefore, it can be summarised that the number of pseudopod points corresponds to the number of damped vibrations. Another test, which has not been detailed in the thesis, was undertaken to demonstrate that when the number of pseudopod points increased, the vibration region increased and became increasingly harder to distinguish. Additionally, when the magnitude level of the pseudopod increases, the FD height increases and the conclusion above seems more obvious. This conclusion can be also extended as the number of points in the pseudopod determined the number of damped vibration regions. This is also referred to in the future work section.

From the results of the previous chapter and the conclusion in the first part in this chapter, the FD has to be transformed to the same size with the help of B-Spline and it solves the comparison problem in the logarithmic FD. Therefore, in the following results' parts, the logarithmic FD will use the results and data from the previous chapter.

5.2.3 Combining Fourier Descriptor with Shape Tracking

The key idea of this subsection is that the retrieved shape information is used on the FD, and it can be obtained from B-spline, which uses the initial shape information. The KF and INNF method can observe the estimated control points' state and the estimated state can substitute the control points from B-Spline. This was implemented using the initial shape boundary information in the initial frame

working with the KF and INNF to obtain the estimated control points. Then B-Spline technique was used to change the estimated control points into retrieved shape boundary as intermediate result. Finally, the above intermediate result was combined with the FD to analyse. In this case, the different sizes' problem was solved by using the initial shape boundary information. Additionally, it reduced the computation complexity.

According to Fourier Transform and the knowledge from the previous chapter, a two-dimensional retrieved shape boundary and the FD of the retrieved shape boundary can be changed into the following forms:

$$\begin{aligned} \boldsymbol{Z} &= \boldsymbol{H}\boldsymbol{Q_c} \\ F(\boldsymbol{Z}) &= F(\boldsymbol{H}\boldsymbol{Q_c}) \\ &= F(\boldsymbol{H}\hat{\boldsymbol{Q}}_c) \\ &= \boldsymbol{Z_4}\boldsymbol{H}\hat{\boldsymbol{Q}}_c \end{aligned} \tag{5.7}$$

where $\boldsymbol{Z}$ is the retrieved shape boundary; $\boldsymbol{Q_c}$ is the control points from the previous chapter; $\hat{\boldsymbol{Q}}_c$ is the estimated control points from the previous chapter; $\boldsymbol{H}$ is the transfer matrix; $F(\boldsymbol{Z})$ is the Fourier transform on $\boldsymbol{Z}$; $\boldsymbol{Z_4}$ is a matrix with formula $\boldsymbol{Z_4} = [(\boldsymbol{Z_2})', (\boldsymbol{Z_2^2})', ..., (\boldsymbol{Z_2^N})']'$, where N is the total number of frequency; $\boldsymbol{Z_2}$ is a vector with formula $\boldsymbol{Z_2} = [Z_1, Z_1^2, ..., Z_1^N]$ and Z_1 is the angle with value $Z_1 = e^{-j2\pi/N}$. Since Z_1, $\boldsymbol{Z_2}$, and $\boldsymbol{Z_4}$ can be calculated offline, the FD of shape can be reduced to the function of $\hat{\boldsymbol{Q}}_c$, which is the estimated control point from the KF. The comparison result is given in Figure 5.7.

The green dot-dash line is the first method calculated by the traditional FD. The red dotted line is the second method calculated using the formulae above. From Figure 5.7, it is clear that the reduced method (the second method) has the same logarithmic values of FD as the one calculated from the traditional method (the first method). An important point should be noted that the retrieved estimated neutrophil shape boundary is better than the retrieved neutrophil shape boundary. This is because the estimated control points have self-modification ability from using the KF. In addition, a comparison is made to prove that the retrieved estimated neutrophil shape boundary is better. The result is that the retrieved estimated neutrophil shape boundary is closer to the initial shape boundary in some frames; in other frames, both the methods are close to the initial shape boundary. The results' figure is not provided in this current work. Therefore, the FD being combined with the methods in the previous chapter can reduce a huge number of computations; it also reduces the complexity of the analysis methods. Therefore, the Fourier analysis is a linear function of the estimated control points, which

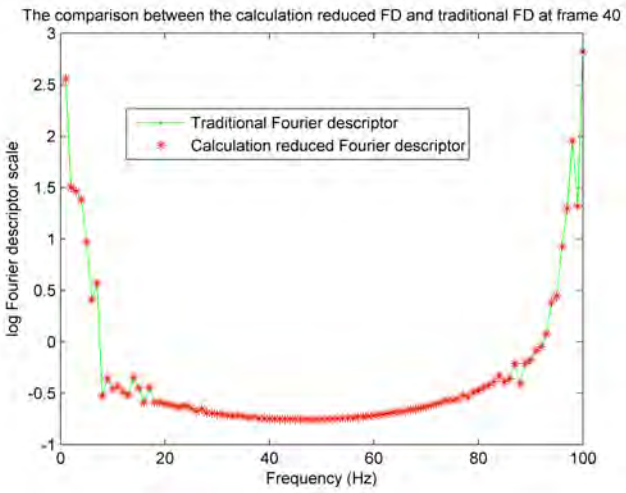

Figure 5.7: The logarithmic FD comparison of two methods

The green dot-dash line is the logarithmic FD calculated from the traditional method, which uses the shape boundary information. The red star line is the logarithmic FD calculated by the reduced method, which transfers the Fourier calculation to a factor multiplying the estimated control points.

combines the methods in the previous chapter.

5.2.4 Height Descriptor Implementation

The results of the radius descriptor and the relevant two cases mentioned above are not provided in this current work. They are discussed in the future work section. In this part, only the height descriptor result will be presented.

The FD is a highly significant analysis tool. However, according to the real data analysis, the same neutrophil in different frames has different FDs, both the height of FD and the noise on the FD. The neutrophil probably changes modules in different frames. Therefore, the FD height probably decides the neutrophils' module. From the conclusion in concentric circles test, which is not discussed in this thesis, the FD height is not related to the radius. The area of neutrophils, therefore, as the most important parameter, was considered firstly. Roundness, another important parameter, was also considered.

The FD height of No. 9 Neutrophil is compared with the area and the roundness illustrated in Figure 5.8. Frankly, there is no real physical meaning for the y axis. According to Figure 5.8, the roundness seems closely related to the negative height of FD. Although they are not perfectly matched, most of the tendency or changing rate is nearly the same. However, the area sometimes has the same

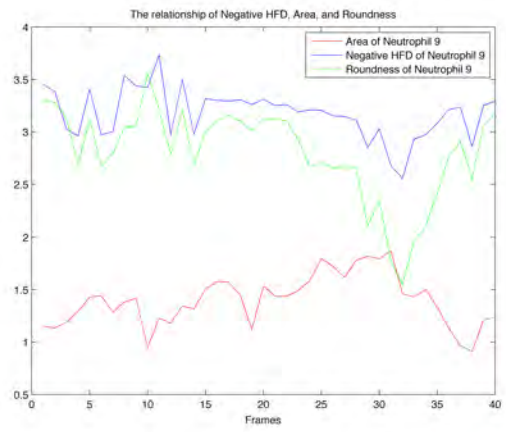

Figure 5.8: **The relationship of three parameters**

The red curve is the area of No. 9 Neutrophil. The blue is the negative height of logarithmic FD. The green is the roundness of No. 9 Neutrophil. It should be noted that there is no y axis label or unit. It is because that none of three parameters have the same label or unit. For instance, the unit of area is μm^2, while both negative HFD and the roundness are dimensionless. Therefore, no label or unit is on the y axis.

tendency, while at other times, it has the opposite. Based on the results above, a conclusion is drawn that there is no obvious relationship between the area and the height of FD.

The roundness must be noted. The formula of roundness used in this current work is different from the traditional formula. It is given as follows in Equation 5.8

$$RN = \frac{l}{d} \tag{5.8}$$

where, RN represents the roundness; l is the length of neutrophil; d is the width of neutrophil.

In future work, perhaps, roundness could be used to define an active neutrophil instead of the different FD heights. Although the relationship between the height of FD and the roundness is not 100% certain, it, at least, establishes a link in analysis between space domain and frequency domain for future work.

In addition to the height descriptor, another descriptor will be represented in the following subsection.

5.2.5 Low Frequency Descriptor F(w=2) Implementation

The LFD means the logarithmic Fourier values at frequency $w = 2$. There is no relevant literature about the LFD and its physical meaning. However, it is believed that a relationship between $\boldsymbol{F(w = 2)}$ and the roundness exists. No matter what the exact physical meaning $\boldsymbol{F(w = 2)}$ is, it can be used to express some features of neutrophils.

To show the relationship between the LFD and the roundness, several comparison tests were undertaken and are illustrated below.

Figure 5.9 shows the distributions of all the neutrophils in all the frames in the low resolution data between $\boldsymbol{F(w = 2)}$ and the area, and between $\boldsymbol{F(w = 2)}$ and the roundness. In the distribution, there is no obvious conclusion. However, considering the different modules of neutrophils in different frames, it is reasonable to believe that the distribution is the linear combination of the different modules in different frames. Therefore, some special modules of neutrophils are extracted and the distributions are illustrated as follows. It should be noted that many similar modules with similar distributions in different neutrophils exist and the examples have only been listed for those different modules.

Figure 5.10 illustrates the distribution of No. 13 Neutrophil. The top figure represents the relationship between the LFD and the area. It seems diverged. The bottom figure is the relationship between the LFD and the roundness. The relationship seems linear with the gradient negative. However, the No. 13 Neutrophil is very special, because it crossed No. 12 Neutrophil and during the crossing process, No. 13 Neutrophil and No. 12 Neutrophil covered each other and then separated. This process is called merging and separating. The star points in the first several frames belonged to No. 13 Neutrophil while the other frames were No. 12 Neutrophil and this conclusion can be explored in future work. Given its complexity, Figure 5.10 is highlighted as a special case. Other crossing neutrophils exist but none will be indicated as analysis examples.

Figure 5.11 shows the distribution of No. 49 Neutrophil. It seems that both the top and bottom figure represent a linear relationship with the gradient vertical. It represents a type of neutrophil module at the bottom figure. However, as for the other neutrophils, the area and $\boldsymbol{F(w = 2)}$ do not seem to be related.

Figure 5.12 represents the distribution of No. 28 Neutrophil. The bottom of Figure 5.12 shows a diverging relationship between roundness and F(w=2); while the top of Figure 5.12 has a diverging relationship pointing to the right. This diverging relationship (the bottom figure) also expresses another module of neutrophils.

Figure 5.13 shows the distribution of No. 9 Neutrophil. The bottom of Figure

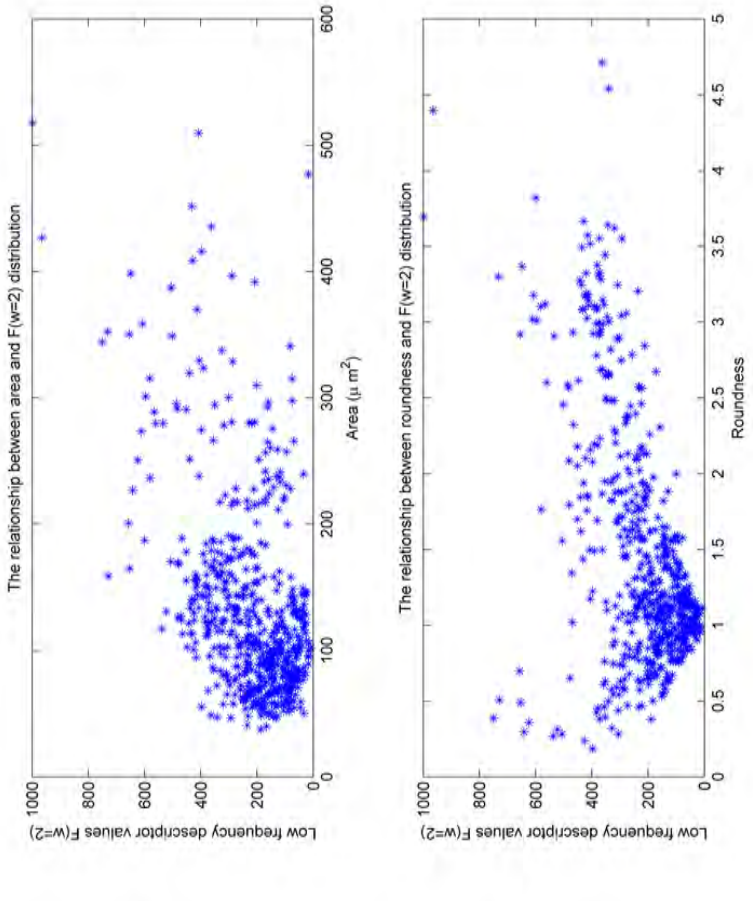

Figure 5.9: The LFD $F(w = 2)$ distribution

The top figure is the LFD $F(w = 2)$ distribution of the neutrophils' area. The bottom figure is the LFD $F(w = 2)$ distribution of the neutrophils' roundness. The blue stars in both are the neutrophils corresponding values. For instance, a star on the top figure is the area value corresponding to a neutrophil. All the neutrophils in all the frames are counted in both of figures.

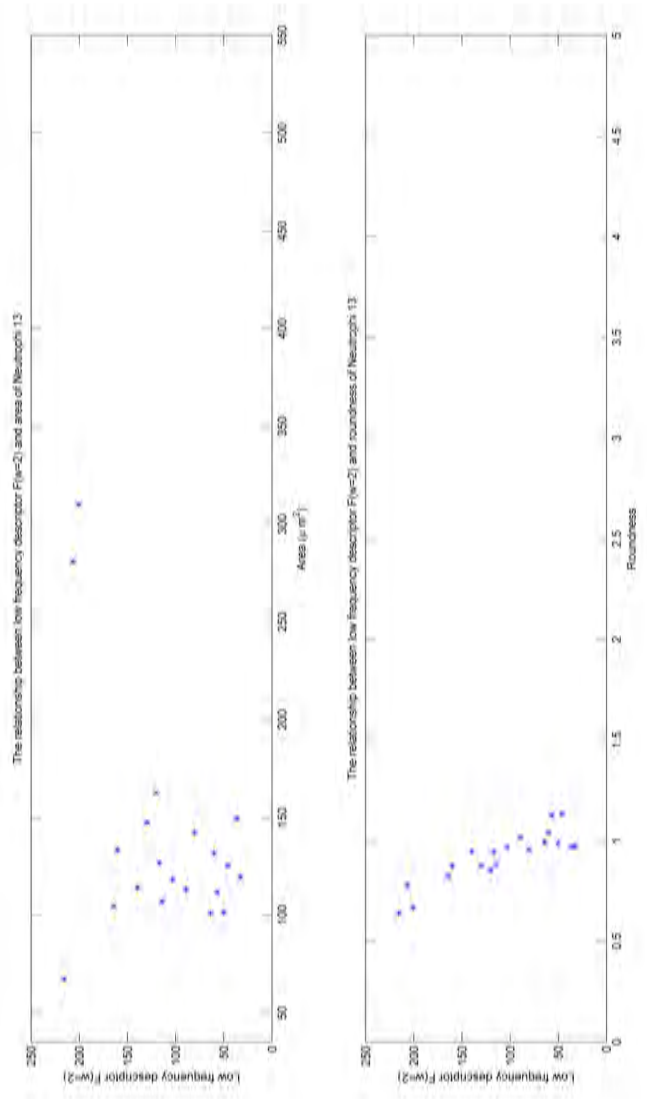

Figure 5.10: **The No. 13 Neutrophil LFD $F(w = 2)$ distribution**

The top figure seems diverged and the bottom one seems to have a linear relationship with the gradient negative. The distribution of No. 13 Neutrophil is a special case and it is not correct since after a few frames, No. 12 Neutrophil and No. 13 Neutrophil merged and then separated. It is an example of merging and separating neutrophil.

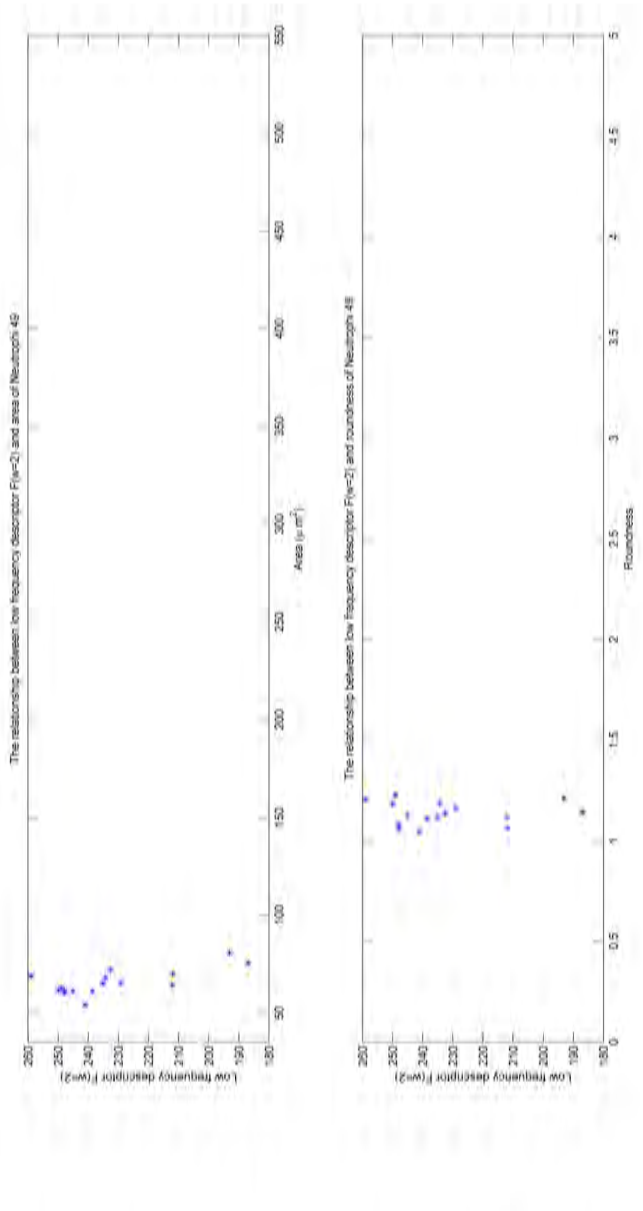

Figure 5.11: **The No. 49 Neutrophil LFD $F(w = 2)$ distribution**

Both the top figure and the bottom one indicate that a linear relationship exists. The gradient of both seems vertical. It represents a type of neutrophil.

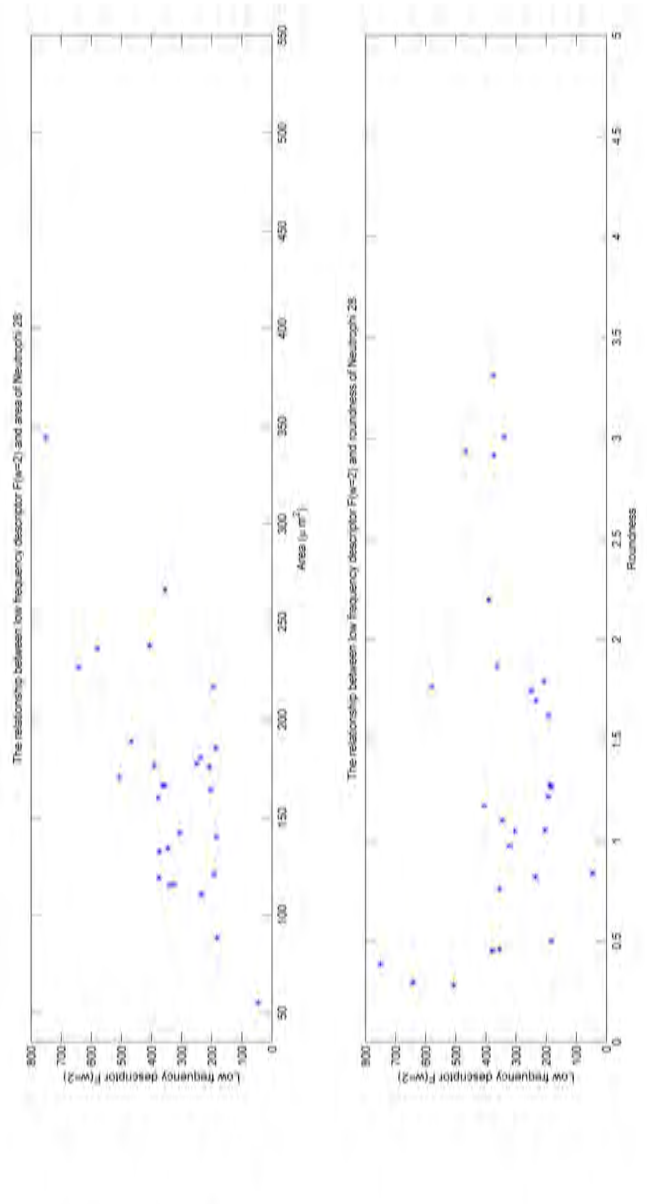

Figure 5.12: The No. 28 Neutrophil LFD $F(w = 2)$ distribution

The top figure seems to diverge toward the right hand side, while the bottom one diverges irregularly. It also represents a type of neutrophil.

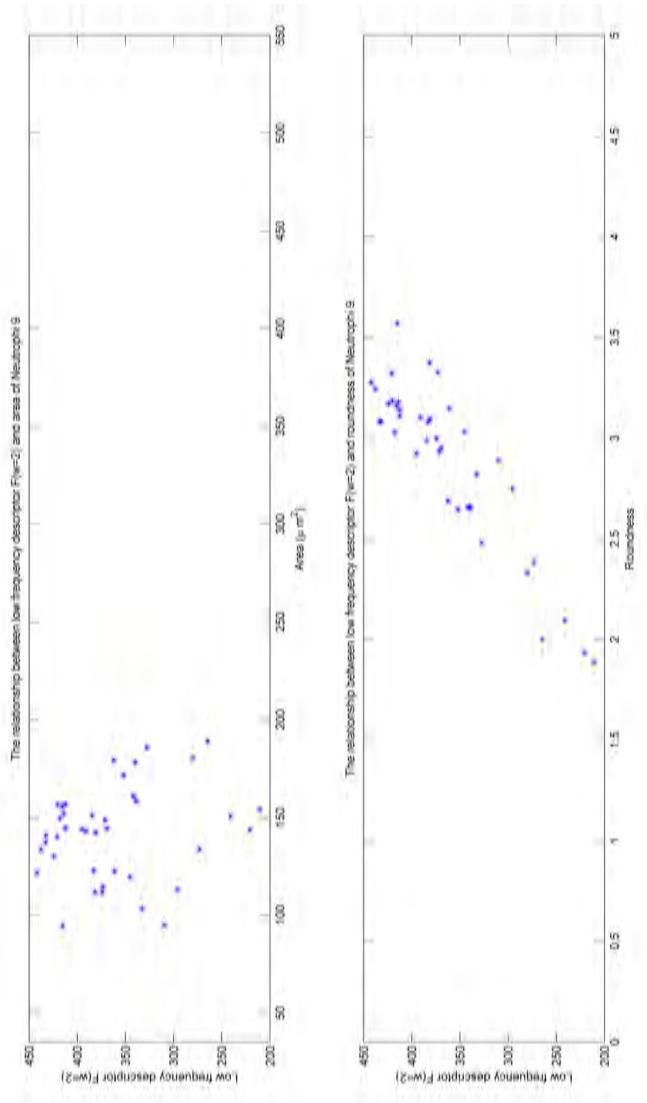

Figure 5.13: The No. 9 Neutrophil LFD $F(w = 2)$ distribution

The top figure shows that there are two parallels with the gradient negative. The bottom one seems to be the combination of the piecewise linear relationship with a positive slope and a piecewise divergence relationship. It also represents one type of neutrophil.

5.13 illustrates the piecewise linear and piecewise diverging relationship. The linear relationship is a direct proportion and at the latter part, the relationship seems diverging. However, in the top figure, there seems to be no relationship. Those points look like two parallels. The segmental linear and segmental diverging relationship represent another module of neutrophils.

In addition to these examples, the distribution of $\boldsymbol{F(w = 2)}$ for all the neutrophils in all the frames is shown in Figure 5.14. It must be noted that $\boldsymbol{F(w = 2)}$ is the unnormalised logarithmic FD. The left hand side and the right hand side of Figure 5.14 are the same, but viewed from different angles. It is clear that for some neutrophils, the colour is light blue in all the frames, while for others the colours are yellow and red in most of the frames. Therefore, it proves that neutrophils have different modules, at least three. One is active, one is "dead" with small movement randomly around the centroid and the third changes between the other two modules in random frames.

Therefore, several conclusions are drawn as follows.

Firstly, roundness has the relationship with the LFD $\boldsymbol{F(w = 2)}$. It is similar to the conclusion from the previous subsection that only roundness has the relationship with neutrophils' feature. Secondly, different relationships between roundness and $\boldsymbol{F(w = 2)}$ represent different neutrophil modules. Some are dead neutrophils, some are active neutrophils, others are neutrophils changing modules between dead and active. The corresponding relation will be analysed in the next chapter. Thirdly, $\boldsymbol{F(w = 2)}$ can be used to detect the module of neutrophils.

5.2.6 Pseudopod Analysis

This subsection is mainly about pseudopod analysis. It is extremely difficult to determine an active positively moving neutrophil, therefore, No. 28 Neutrophil in Frame 12 and Frame 13 was used as a similar example and is represented in Figure 5.15.

Figure 5.15 shows the shape and the logarithmic FD of neutrophil in the continuous two frames when the pseudopod occurs. The top of Figure 5.15 is the estimated retrieved shape boundary in Frame 12 and Frame 13. The blue solid line is the shape in Frame 12 and the red one is in Frame 13. The bottom of Figure 5.15 is the corresponding FD. It is clear that when the pseudopod occurs, both motility and morphology (shape boundary) and shape descriptor changes. Furthermore, the FD height also changes. From the shape changing in the top figure, No. 28 Neutrophil moves toward the top right. The pseudopod perhaps occurs on the top right of No. 28 Neutrophil. The No. 28 Neutrophil first rotated to the top right, and then the pseudopod appeared with considerable velocity in that area.

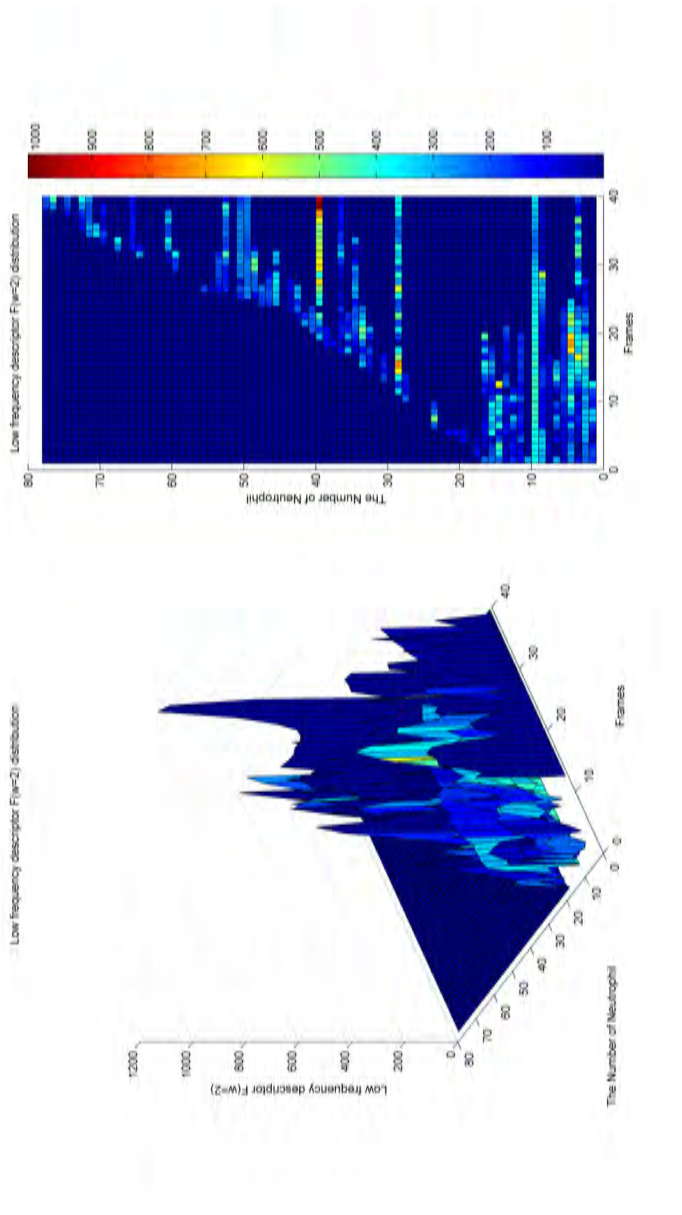

Figure 5.14: The LFD $F(w = 2)$ distribution of roundness.

The LFD distribution of roundness for all the neutrophils in all the frames is plotted as a surface on the left figure. The top view is illustrated on the right. By using the distribution of roundness, several modes of neutrophils are distinguished, such as the active neutrophil.

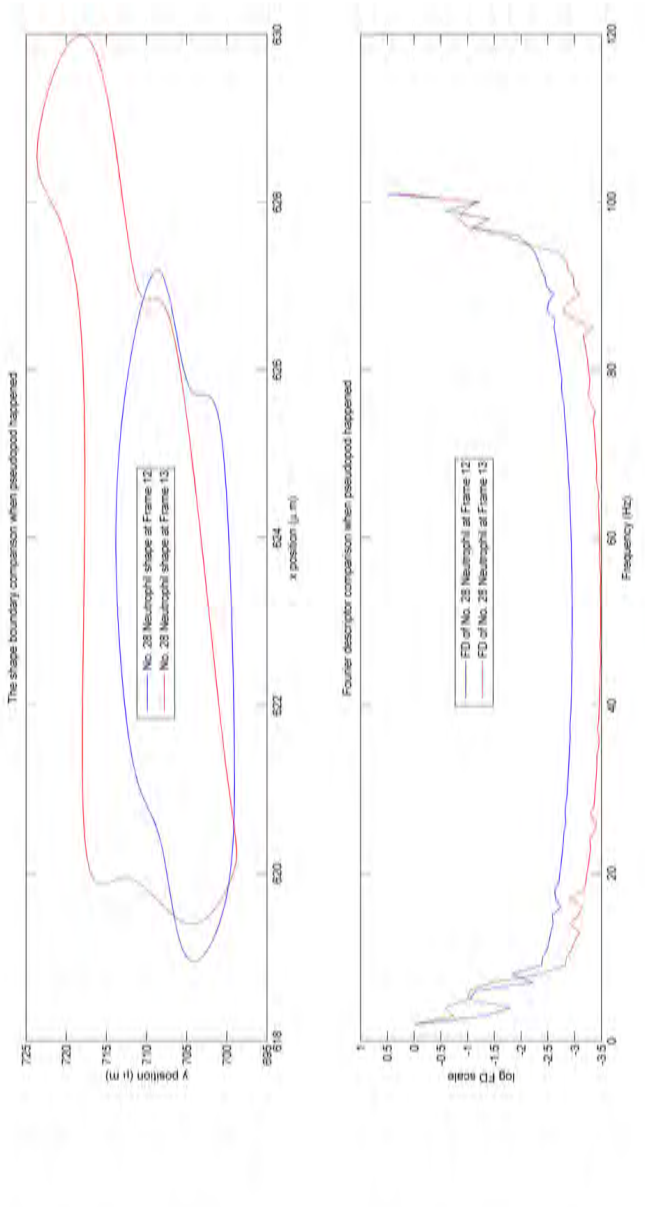

Figure 5.15: **The pseudopod analysis**

The top figure illustrates the process of the pseudopod occurring. It is the shape boundary comparison of the same neutrophil in the continuous two frames. Usually, a huge shape changing (adding or subtracting) in some parts of the neutrophil exists when a pseudopod occurs. For example, in the figure, the red top right part is a pseudopod. The bottom figure represents the logarithmic FDs of the top figure, that is, the logarithmic FD comparison, when the pseudopod occurred.

At the same time, the other part of No. 28 Neutrophil also moved with small velocity on the boundary. Therefore, the velocity on the boundary research is highly significant. Since the length of No. 28 Neutrophil increases and the width changes a little, the height in Frame 13 is less than Frame 12. It thus proves the height conclusion above. Therefore, it is concluded that when the pseudopod occurs, the velocity on the boundary changes with different speeds according to the position of the neutrophil. Additionally, the height of FD may have also change, depending on the direction in which the pseudopod occurs.

5.3 Conclusion

The main idea of this chapter has been shape analysis with effective tools. The first section introduced background knowledge of the shape descriptor. The FD was introduced. The results' part followed the introduction. The FD was implemented. However, the dimensions of FD were different. Therefore, the FD cannot be used as an analysis tool. Next, several tests were implemented to verify the hypotheses. Through mathematics, the FD was transformed into a linear function of control points, which were estimated from the previous chapter. The validation was undertaken and the result indicated that the transform was correct. Therefore, the problem of different sizes was solved by using the control points instead of the real positions and the FD was added in the framework as an analysis tool. Other descriptors were implemented, such as a height descriptor. Finally, pseudopod analysis was undertaken. Innovation and contributions are as follows. Using mathematics, the FD was combined with tracking methodologies to solve the problem of different sizes. The FD with tracking methodologies formed a new framework which can not only track features of the centroid and shape, but can also analyse the features. The LFD, as a new analysis tool, seemed to have the ability to detect the different modes of neutrophils. The distributions of LFD indicated that at least three modes of neutrophils exist. Finally, the mechanism was better understood after the pseudopod analysis.

It should be noted that, the shape's FD is different from the FD of other parameters. Thus far, it is still difficult to determine which parameter or methodology would be the best for shape analysis. Although some results are different from the FD of shape boundary, such as the height and radius descriptor, they are probably another way to analyse the neutrophil information and modules. The combination between centroid tracking, shape tracking and shape descriptor will be considered in the next chapter.

Chapter 6

Modelling Neutrophil Dynamic from Shape and Motion

The previous chapter discussed the descriptors used to analyse neutrophils' characteristics. The FD was the main analysis tool, however, the LFD was proposed since it can detect the neutrophils' mode. Therefore, in this chapter, the combining of shape tracking, centroid tracking, and shape descriptor is implemented to determine a new framework. Neutrophils' motion and morphology, as well as other cells, can be analysed using this framework.

6.1 Mode from Centroid Tracking versus Low Frequency Descriptor

This section mainly focused on the relationship between the mode from centroid tracking and LFD. The mode from centroid tracking was the same methodology, the MM filter, that was used in Chapter Three. There were three models: the constant acceleration model, constant velocity model, and random walk model. LFD $\boldsymbol{F}(\boldsymbol{w} = \mathbf{2})$ was as per the previous chapter. The comparison result is shown in Figure 6.1.

The top of Figure 6.1 illustrates the centroid tracking with different models, where the blue, green and red colours represent the constant acceleration model, constant velocity model and random walk model, respectively. The centroid tracking trajectory has been plotted in black and it is covered with the colour of the dominant models. The purple ball indicates the current position. The middle of Figure 6.1 shows the probability of each mode. The yellow vertical lines are the switching points. For instance, the first switching point is Frame 4, which means in Frame 4, the neutrophils' model changes from the constant acceleration to ran-

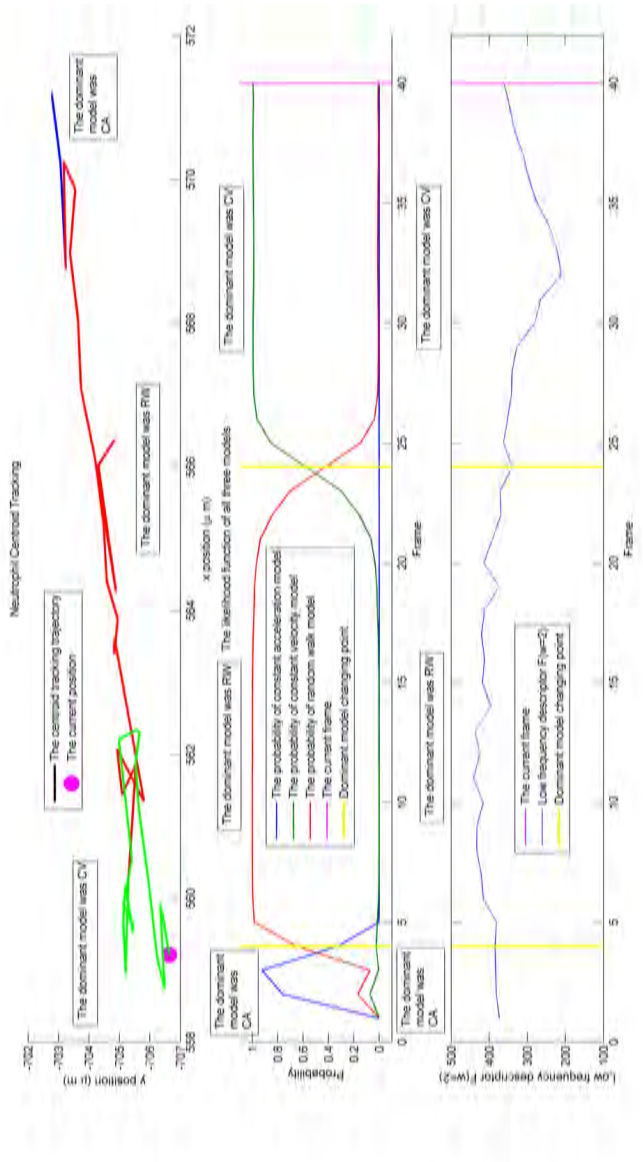

Figure 6.1: **The relationship between the mode based on centroid tracking and the LFD**

The top figure illustrates the centroid tracking trajectory in black. The purple ball is the current position and currently it is at the finishing point. Therefore, the black trajectory is covered with different colours. In the trajectory, the blue part is the constant acceleration model as the dominant model. The red part is the random walk model as the dominant model. The green part is the constant velocity model as the dominant model. The second figure represents the probability distribution. The purple, blue, green and red are the current position, constant acceleration model, constant velocity model and random walk model respectively. The yellow vertical lines are the switching point when the dominant model changes. The bottom figure also plots the switching point to determine whether there is any relationship. The purple line is also the current frame, but the blue line is the LFD $\boldsymbol{F(w=2)}$.

dom walk. The purple line indicates the current frame. The blue, red and green are the same as the top figure. The bottom one represents the LFD $\boldsymbol{F}(\boldsymbol{w} = \mathbf{2})$. The yellow lines and the purple line are the same as the middle figure. The blue curve is the LFD values. When the model chosen by the probability changes, the LFD indicates no obvious response. Therefore, it is concluded that there are no obvious connections between the LFD and the model from centroid tracking. However, the model from centroid tracking does not represent the real neutrophil mode. For example, Figure 5.15 shows that the pseudopod occurs with very little centroid changing and almost retains the same mode from centroid tracking. However, part of the neutrophil is believed to change mode, while the other part remains in the same mode. Therefore, the best method should combine the centroid and the shape mode to judge the relationship and the shape mode detection is likely to be future work.

6.2 Centroid Velocity versus Low Frequency Descriptor

The aim of this section is to determine the link between the centroid velocity and the LFD. The centroid velocity also used the same methodology from Chapter Three. Since the neutrophils were the same as in the previous section, the modes remained the same. The relevant result is shown in Figure 6.2.

In Figure 6.2, the blue, green and red curve are the same as the top figure in Figure 6.1. The starting point is the first frame and the finishing point is the final frame. At the beginning, it seems that the LFD and the centroid velocity are independent. When the dominant model is the constant velocity, it seems that the LFD and the centroid velocity are partially dependent. As part of the green curve in the figure seems like two parallel lines with the gradient positive, the LFD is considered as a direct proportion function of the centroid velocity. However, considering the relationship in total, there is no relationship between the LFD and the centroid velocity. Given of the small centroid velocity, this neutrophil can be considered as "dead" and it makes some small regional random walk movement. Therefore, based on the result from the previous section, it is better to include the shape velocity on the boundary to obtain a more accurate conclusion.

6.3 Shape Velocity Distribution versus Low Frequency Descriptor

As mentioned above, the shape velocity on the boundary was needed. In this section, the LFD and the shape velocity distribution are compared and analysed

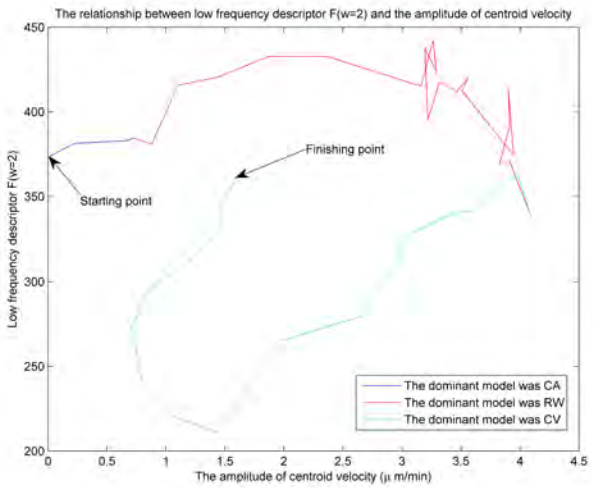

Figure 6.2: **The relationship between the amplitude of the centroid velocity and the LFD**

The starting point and the finishing point are plotted. The blue, red and blue are the constant acceleration dominant model, random walk dominant model and constant velocity dominant model respectively.

to determine the internal connections. The shape velocity distribution means the distribution of velocity on the shape boundary based on different boundary segmentations and different frames. The comparison result is shown in Figure 6.3 and Figure 6.4.

The top figure of Figure 6.3 is the shape velocity distribution of No. 9 Neutrophil. As mentioned above, the shape distribution is based on all the boundary points and the frames. The bottom figure is the same LFD as above. It is extremely difficult to discover any link. Therefore, another figure with different angle views is needed. As such, Figure 6.4 is presented. Figure 6.4 is the same as Figure 6.3 with different angle views. The top figure of Figure 6.4 is the top view distribution. The coloured bar represents the shape velocity amplitude. Compared with other neutrophils, the amplitude of the shape velocity is small, since the No. 9 Neutrophil is changing its mode as noted above. The bottom one is the LFD, which is identical to the one specified above. It is still extremely difficult to reach a conclusion because the shape velocity should be analysed not only with magnitude but also with directions. In addition, just one neutrophil cannot represent all neutrophils' modes. Therefore, it is necessary to analyse all the other modes and check whether there was any connection.

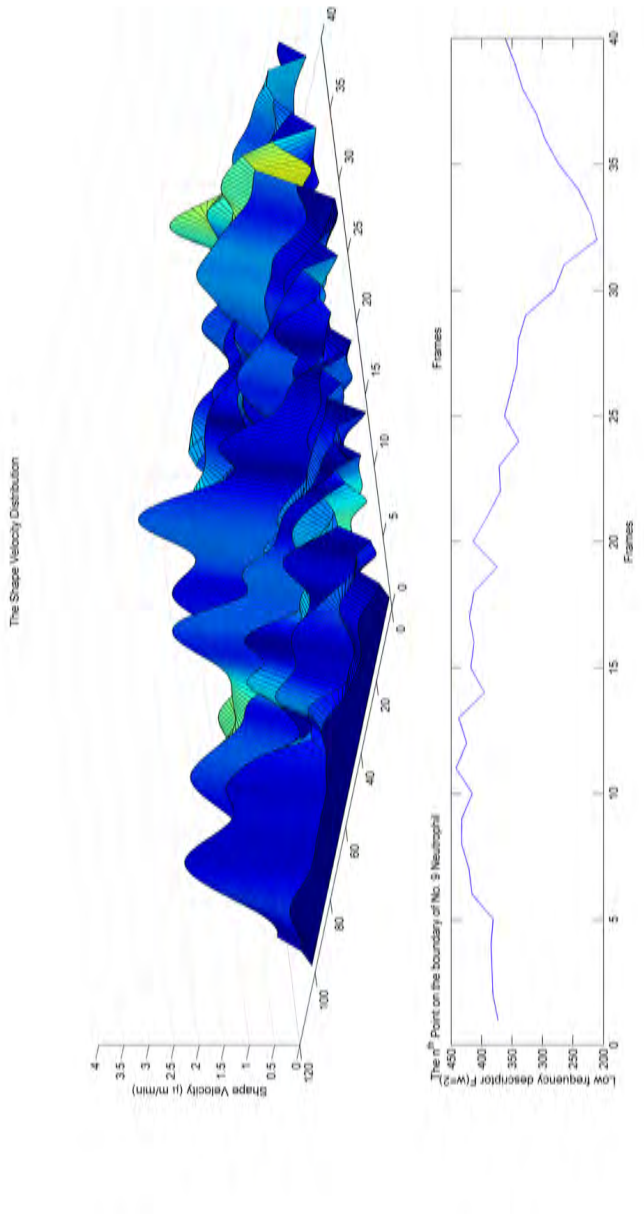

Figure 6.3: The relationship between the amplitude of the shape boundary velocity and the LFD

The top figure represents the distribution of the No. 9 Neutrophil shape boundary velocity. There are 101 boundary points and 40 frames. The bottom figure is the LFD $\boldsymbol{F(w=2)}$ for all the 40 frames.

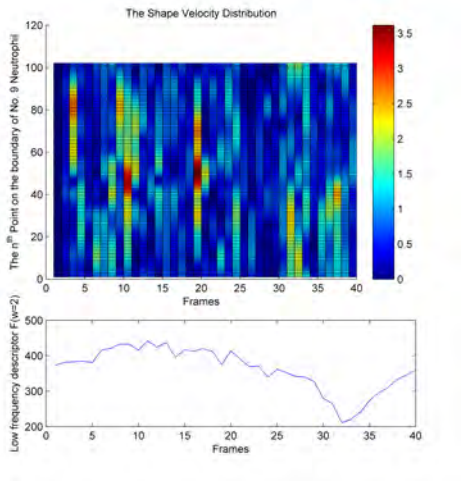

Figure 6.4: **The top view of the shape velocity distribution.**

The top figure is the same as the top one in Figure 6.3 but viewed from the top. The colour bar represents the amplitude of the shape boundary velocity. It is clear that the shape boundary velocities of different points are different for the same neutrophil in different frames. The bottom is the LFD of No. 9 Neutrophil.

6.4 Neutrophil Tracking and Analysis

Given the uncertainty of the above results, it is necessary to combine all the relevant methodologies. Examples of two neutrophils will be shown to combine all the relevant methodologies.

Example 1: No. 49 Neutrophil

No. 49 Neutrophil is a "dead" neutrophil represented in Figure 6.5. The black solid line is the centroid trajectory. The purple square is the current centre. The green arrow is the centroid velocity. The blue solid line is the current shape boundary and the blue dotted line is the next frame shape boundary. The carmine arrows are the shape velocities. From Figure 6.5, it is clear that the centroid trajectory is like a random walk mode and the centroid velocity is minimal for all 40 frames. Furthermore, the velocity on the boundary is small, although the shape of No. 49 Neutrophil is prolate. Combined with Figure 5.11, it is reasonable that the linear relationship, with gradient vertical, between $\boldsymbol{F(w=2)}$ and roundness, can detect the neutrophil mode as "dead". In this current work, the "dead" neutrophil means inactive neutrophils performing small region random walk movements.

Example 2: No. 28 Neutrophil

No. 28 Neutrophil is illustrated as an active neutrophil in Figure 6.6. All the

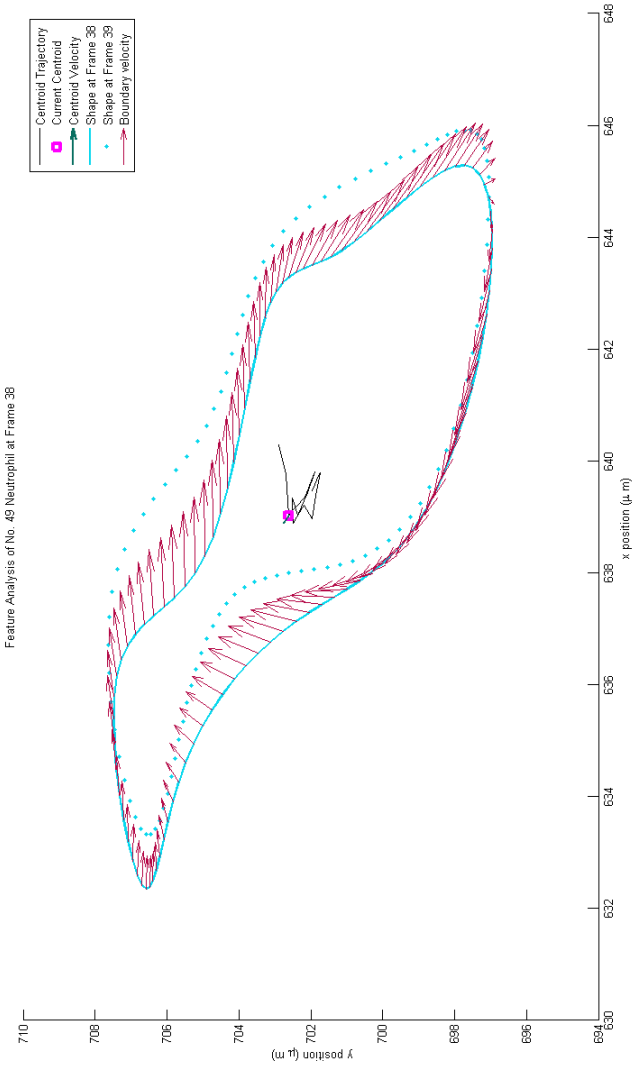

Figure 6.5: **The feature of No. 49 Neutrophil**

The centroid trajectory is plotted in black solid line. The purple square is the current centroid position of No. 49 Neutrophil. The green arrow is the centroid velocity. The length of the green arrow represents the amplitude and the arrow points to the direction, which the centroid moves to in the next frame. The blue solid line is the current shape boundary. The blue dotted line is the shape boundary in the next frame. The carmine arrows are the shape velocity with not only the amplitude but also the direction.

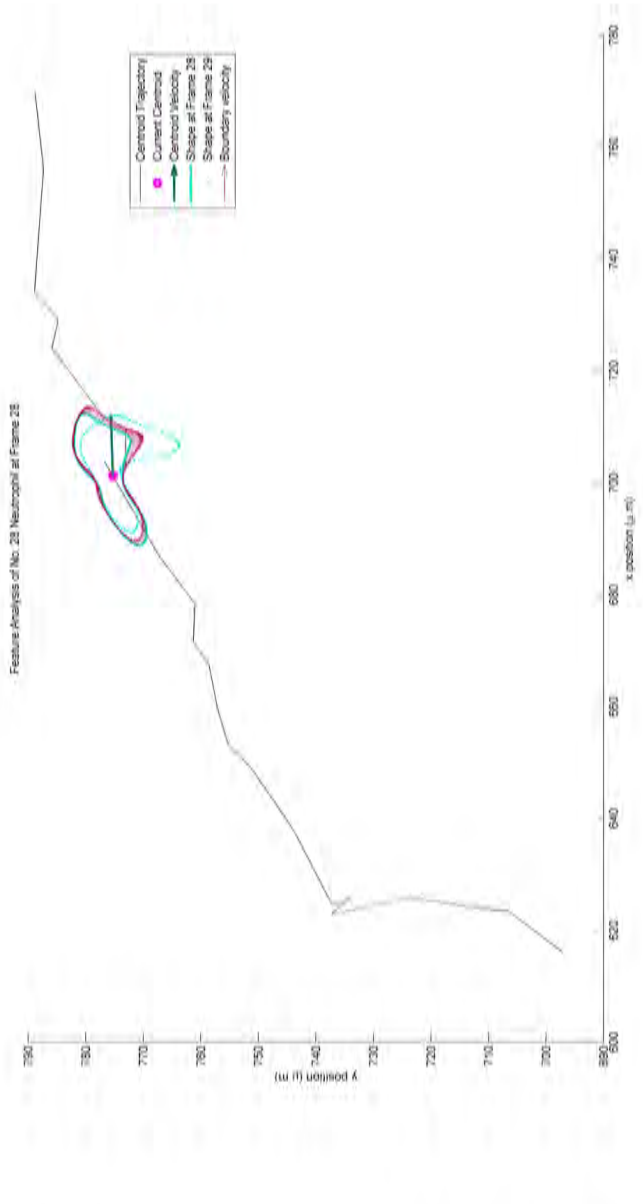

Figure 6.6: The feature of No. 28 Neutrophil

All the colours are the same as mentioned in Figure 6.5. The difference is that the centroid trajectory is long. So are the velocity, centroid and shape.

legends are the same as in Figure 6.5. The differences are as follows. Firstly, the centroid trajectory has a very long distance. Secondly, the centroid velocity magnitude is considerable and the centroid velocity points in one direction. Thirdly, the velocities on the boundary are huge. Combined with Figure 5.12, it is reasonable that the diverging relationship between LFD $\boldsymbol{F(w = 2)}$ and roundness can detect an active neutrophil mode. This kind of active neutrophil usually has enormous movement, including the centroid movement and the shape movement with a specific directional destination.

Combining Example 1 and Example 2, the LFD $\boldsymbol{F(w = 2)}$ can be used to detect the mode of neutrophil.

6.5 Conclusions

This chapter concludes with a summary of the above combination work. This chapter has consisted of five sections and the first three mainly discussed the relationships between the mode from centroid tracking and LFD, between the centroid velocity and LFD, and between the shape velocity distribution and LFD. The fourth section was the summary of all the methods used in previous chapters. Finally the conclusion to this chapter is presented. In this chapter, LFD was the main shape descriptor used and analysed, because of its effective results in the previous chapter. Both modes from centroid tracking and centroid velocity, which were the important results from centroid tracking, were compared with LFD to determine the connections. Shape velocity distribution, as another important result in shape tracking, was also compared with LFD. By using the comparison results represented above, several conclusions can be drawn. Firstly, there were no obvious connections between the mode from centroid tracking and LFD from the results. However, the mode was from centroid not from the shape boundary, which meant that if the shape boundary changed, for example, a pseudopod occurred, while the centroid mode remained the same, the mode from centroid tracking stayed the same but LFD changed, as Figure 5.15 illustrated. LFD, as a feature of neutrophil, presented some information based on the shape boundary. Therefore, it is possible that LFD had some connections with the shape mode, which will be future work. Moreover, it perhaps yielded more interesting results by using the connections between the centroid mode and shape mode. Secondly, there was no obvious link between the centroid velocity and LFD. However, several modes of neutrophils existed, and different modes had different ranges of centroid velocity. In considering the mode of neutrophils, as per the "dead" neutrophil example, there was perhaps an internal link since the LFD did not change in a considerable

range except for Frame 32. A special case, like Frame 32 in the example, would be linked to shape velocity and shape mode. Thirdly, it seemed that shape velocity distribution was independent of LFD. However, the velocity distribution only considered the magnitude and not the angles. Furthermore, different modes of neutrophils had different velocity distributions. Therefore, any mode from centroid tracking, centroid velocity, and shape velocity distribution had no obvious connections with LFD. However, the combination of all these, together with an unknown feature such as the mode from shape boundary, would link to LFD. Finally, not all the neutrophils with prolate morphology were active. In addition, LFD can detect the neutrophil mode combined with other features. Innovation and contributions from this chapter are as follows. LFD was effective in detecting the neutrophil mode with other features of neutrophils and it, as an effective analysis tool, combined with other techniques, together formed the whole framework. Having concluded this chapter, a summary of the complete thesis and future work will be presented in the next chapter.

Chapter 7

Conclusion

The previous chapter discussed the relationship between the LFD and other methodologies. Finally, the new framework was established through combining all the methodologies. This chapter contains two sections, a summary and future plan.

7.1 Summary

This thesis has undertaken research of neutrophils based on the morphology and motility. Neutrophils, as a kind of white blood cell, have a strong ability to move to the site where inflammation happens and to protect the body by phagocytosis or other powerful functions. During the process, neutrophils were thought to change shape caused by the chemotaxis field. "Dead" neutrophils sometimes move randomly or "sleep" in most cases. When the inflammation occurs, the chemotaxis field changes drastically from frame to frame, which makes active neutrophils move, accompanied by shape changing. Therefore, research on neutrophils' morphology and motility is significant. When neutrophils move, a large quantity of information exists, both on the shape boundary and on the centroid. Therefore, this research was conducted on both the centroid tracking and shape tracking, which was covered in chapters three and four. After obtaining significant tracking results, the information from these were analysed using effective tools and has been detailed in Chapter Five. By combining all the information from the above chapters, an effective methodology to solve such a problem, was determined in Chapter Six.

Since the movement of neutrophils is a highly complex process, it requires more than one traditional tracking model. As such, the centroid tracking was implemented by the MM algorithm, which calculates the probability of different models in the same frame and chooses the maximum likelihood model as the dom-

inant model for that frame. The square of centroid estimated velocity magnitude, as an analysis tool, illustrated the distribution of the chemotaxis field. However, all the above methods and analysis only considered centroid information. This was correct, but did not comprise the whole information. There was a significant amount of information on the boundary because the shape changing always happened during the moving process to the inflammatory site, and this information was used in Chapter Four. Innovation and contributions in Chapter Three were as follows. Firstly, the chemotaxis field was estimated using the centroid velocity. It indicated that the chemotaxis field did exist. With information about the chemotaxis field, the mechanism of neutrophils was better understood. The chemotaxis field would be more accurate if it were estimated using the shape velocity. Secondly, the parameters of the KF were fixed, and those parameters were used in the subsequent chapters. Another contribution was that the centroid tracking of neutrophils was applied by the MM filter and this was a new application field. By changing the different models, the motion of neutrophils was estimated more accurately. Finally, the centroid velocity was estimated as an important parameter to analyse the chemotaxis field rather than considering the centroid position tracking only.

Shape tracking is a complex tracking area, combining the fields of computer vision, data association and traditional target tracking. Usually, there are two requirements for shape tracking: the correctness of retrieved shape and the correlation of the same neutrophils in different frames. The former can be guaranteed by the B-Spline method, which belongs to the computer vision area, and the latter can be guaranteed by INNF, which belongs to the data association field. The entire process is completed by a traditional target tracking algorithm. As the focus was on the velocity of all the shape boundary points, it is assumed that all these obey the constant velocity model. Based on the algorithms in Chapter Four, the velocity can be retrieved using the estimator of control points. However, only one module was considered for the neutrophils, and the other two modules, which appeared in Chapter Three, were not applied in Chapter Four. If the MM filter can work on or combine with other algorithms on the shape tracking chapter, the results will be better. Innovation and contributions in Chapter Four are as follows. Firstly, the traditional NNF was improved to solve the problem of neutrophils missing and reappearing. Secondly, the velocity of the boundary was estimated rather than only considering the shape estimation. It is clear that different parts of neutrophils have different velocities. Given this, neutrophils' shapes changed when they moved to the chemotaxis field. Finally, the new framework was established. This would also be suitable for other cells and the velocity on the boundary will

be important in further analysis, which is the future plan.

A shape descriptor is a tool to analyse the shape information as considered in Chapter Five. The focus is the FD of the shape boundary and the descriptors used in this chapter, including LFD. Results were produced based on tests and algorithms detailed in the descriptor section. Innovation and contributions are represented as follows. Firstly, the FD was implemented in a two-dimensional medical image area. Secondly, by comparing the FD with the shape tracking, the framework in the previous chapter was updated. The analysis tool was added in the framework, since the entire tracking process had been simplified. Furthermore, new descriptors were employed in the analysis, such as LFD. Finally, the pseudopod was analysed to understand the mechanism.

Finally, the combination of all the algorithms above presented a highly satisfactory result in detecting the neutrophil mode and also formed a powerful methodology to manage the relevant problems of cells. The methodology was used to detect the mode of neutrophils and through this, the mechanism was better understood.

Based on the limitations mentioned above, the future plan will now be considered.

7.2 Future Plan

Generally, several areas would merit improvement. Firstly, data association methods need to be used and upgraded in both the centroid tracking and the shape tracking to improve the tracking accuracy. The data association mainly focuses on the case of two or more neutrophils coming across, that is, two or more neutrophils covering each other and then separating. It also upgrades the INNF so the neutrophils could disappear in more than one frame. It should be noted that the new method would consider the merged neutrophil as two neutrophils coming across. Therefore the index of both will not break. Furthermore, it will modify the accuracy of both centroid and shape tracking. Secondly, the MM filter will be used or combined with other algorithms in shape tracking. As noted above, the MM filter on the shape boundary can solve a huge number of problems such as classifying different shape modes in one neutrophil. It will also help to determine the connections with LFDs or shape velocity distribution. Thirdly, the Kalman smoother will be utilized to improve the results by reducing the noise and solving the missing data problems. In some results, the Kalman smoother was implemented and it is a powerful tool in reducing noise, although the results were not represented in this thesis. Some neutrophils were missing in a frame, and the

missing data was the estimated initial data without considering the modules and the states. The Kalman smoother rather than the KF, can solve the problem only in the estimated part without the need to add other programs. Finally, new descriptors, such as a radius descriptor, will be explored. Based on the statements in Chapter Five, it would seem reasonable to explore radius descriptors, which can reduce the complex frequency process to time domain process.

The above will be implemented in the future thus concluding this thesis.

Bibliography

[1] H. Akashi and H. Kumamoto. Random sampling approach to state estimation in switching environments. *Automatica*, 13:429–434, 1977.

[2] D. Alspach and H. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions Automatic Control*, 17(4):439–448, 1972.

[3] A. Amini, S. Tehrani, and T. Weymouth. Using dynamic programming for minimizing the energy of active contours in the presence of hard constraints. In *In Proceedings of 2nd International Conference on Computer Vision*, pages 95–99, December 1988.

[4] A. Anderson. A square root formulation of the Kalman covariance equations. *AIAA Journal*, 6:1165–1166, June 1968.

[5] B.D. Anderson and J.B. Moore. *Optimal Filtering*. Prentice-Hall, Incorporation, Englewood Cliffs, New Jersey 07632, 1979.

[6] K. Arbter. Affine-invariant Fourier descriptors. *In J.C. Simon Ed. From Pixels to Features, Elsevier Science Publishers*, pages 153–164, 1989.

[7] H. Asada and M. Brady. The curvature primal sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:2–14, 1986.

[8] A. Atherton and G. Born. Quantitative investigations of the adhesiveness of circulating polymorphonuclear leucocytes to blood vessel walls. *Journal of Physiology and Pharmacology*, 222:447–474, 1972.

[9] N. Ayache and O. Faugeras. Building, registration and fusing noisy visual maps. In *In Proceedings of 1st International Conference on Computer Vision*, volume 7, pages 45–65, 1988.

[10] Y. Bar-Shalom, K. Chang, and H. Blom. Tracking a maneuvering target using input estimation versus the interacting multiple model algorithm. *IEEE Transactions Aerospace and Electronic Systems*, AES-25(2):296–300, March 1989.

[11] Y. Bar-Shalom, X. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. New York: John Wiley & Sons, 2001.

[12] Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. ACADEMIC PRESS, INC., 1988.

[13] R. Battin. *Astronautical Guidance*. McGraw-Hill Book Company, Incorporation, New York., 1964.

[14] G. Beck and G. Habicht. Immunity and the invertebrates. *Scientific American*, 275:60–66, November 1996.

[15] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *Proceedings of Eighth IEEE International Conference on Computer Vision*, pages 454–461, 2001.

[16] S. Berretti, A.D. Bimbo, and P. Pala. Retrieval by shape similarity with perceptual distance and effective indexing. *IEEE Transactions on Multimedia*, 2: 225–239, 2000.

[17] A.D. Bimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19: 121–132, 1997.

[18] A. Blake, R. Curwen, and A. Zisserman. A framework for spatio-temporal control in the tracking of visual contours. *International Journal for Computer Vision*, 11:127–145, 1993.

[19] A. Blake and M. Isard. *Active Contours*. Springer-Verlag London Limited, 1998.

[20] R. Blockley and W. Shyy, editors. *Encyclopedia of Aerospace Engineering, Dynamics and Control*. John Wiley & Sons, Barcelona, 2010.

[21] H. Blom. An efficient filter for abruptly changing systems. In *Proc. 23rd IEEE Conference on Decision and Control, Las Vegas, Nevada*, pages 656–658, December 1984.

[22] H. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Transactions Automatic Control*, 33(8):780–783, August 1988.

[23] H. Blom and Y. Bar-Shalom. Time-reversion of a hybrid state stochastic difference system with a jump-linear smoothing application. *IEEE Transactions Information Theory*, 36(4):836–847, July 1990.

[24] H. Bode and C. Shannon. A simplified derivation of linear least-squares smoothing and prediction theory. *Proceedings IRE*, 38:417–425, 1950.

[25] P. Bogler. Tracking a maneuvering target using input estimation. *IEEE Transactions Aerospace and Electronic Systems*, AES-23:298–310, May 1987.

[26] R. Booton. An optimization theory for time-varying linear systems with nonstationary statistical inputs. *Proceedings IRE*, 40:977–981, 1952.

[27] N. Borregaard. Nneutrophil, from marrow to microbes. *Immunity*, 33:657–670, 2010.

[28] V. Brinkmann, U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann, D. Weiss, Y. Weinrauch, and A. Zychlinsky. Neutrophil extracellular traps kill bacteria. *Science*, 303:1532–1535, 2004.

[29] J. Canny. A computational approach for edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:679–698, 1986.

[30] D. Chetverikov and Y. Khenokh. Matching for shape defect detection. *Lecture Notes in Computer Science*, 1689:367–374, 1999.

[31] R. Cipolla and A. Blake. The dynamic analysis of apparent contours. In *In Proceedings of 3rd International Conference on Computer Vision*, pages 616–625, 1990.

[32] E. Codling, M. Plank, and S. Benhamou. Random walk models in biology. *Journal Royal Society Interface*, 5:813–834, 2008.

[33] J. Cooley and J. Tukey. An algorithm for the machine calculation of comple Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

[34] M. Crowhurst, J. Layton, and G. Lieschke. Developmental biology of zebrafish myeloid cells. *International Journal of Developmental Biology*, 46:483–492, 2002.

[35] R. Curwen, A. Blake, and R. Cipolla. Parallel implementation of Lagrangian dynamics for real-time snakes. In *In Proceedings of British Machine Vision Conference*, pages 29–35, 1991.

[36] M. Daoudi and S. Matusiak. Visual image retrieval by multiscale description of user sketches. *J. Visual Lang. Comput*, 11:287–301, 2000.

[37] E. Davies. *Machine Vision: Theory, Algorithms, Practicalities*. Academic Press, Incorporation, New York, 1997.

[38] L. Dotta, L. Tassone, and R. Badolato. Clinical and genetic features of warts, hypogammaglobulinemia, infections and myelokathexis (whim) syndrome. *Current Molecular Medicine*, 11:317–325, 2011.

[39] D.S.Broomhead and D. Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. *Royal Signals and Radar Establishment Memorandum*, pages 1–34, 1988.

[40] S. Dubois and F. Glanz. An autoregressive model approach to two-dimensional shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:55–66, 1986.

[41] G. Dudek and J. Tsotsos. Shape representation and recognition from multi-scale curvature. *Computer Vision and Image Understanding*, 68:170–189, 1997.

[42] O.D. E. Meijering and I. Smal. Methods for cell and particle tracking. *Imaging and Spectroscopic Analysis of Living Cells*, 504:183–200, 2012.

[43] D. Eberly. *Geometric methods for analysis of ridges in N-dimensional images*. PhD thesis, University of North Carolina at Chapel Hill, 1994.

[44] J. EI-Benna, P. Dang, M. Gougerot-Pocidalo, J. Marie, and F. Braut-Boucher. p47phox, the phagocyte nadph oxidase/nox2 organizer: structure, phosphorylation and implication in diseases. *Exp. Mol. Med.*, 41:217–225, 2009.

[45] G. Eichmann, C. Lu, and M. Jankowski. Shape representation by gabor expansion. *Hybrid Image and Signal Processing II*, 1297:86–94, 1990.

[46] A. Farina. Application of knowledge-based techniques to tracking function. *In Knowledge-Based Radar Signal and Data Processing*, 6:1–34, 2006.

[47] R.A. Fisher. On an absolute criterion for fitting frequency curves. *Messenger of Mathmatics*, 41:155–160, 1912.

[48] B. Fournier and C. Parkos. The role of neutrophils during intestinal inflammation. *Mucosal Immunology*, 61:119–161, 2012.

[49] H. Freeman. On the encoding of arbitrary geometric configurations. *IRE Transactions on Electronic Computers*, EC-10:260–268, 1961.

[50] K. Fu. *Syntactic Methods in Pattern Recognition*. Academic Press, Incorporation, New York, 1974.

[51] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, Incorporation, New York, 1972.

[52] C.F. Gauss. *Theoria motus corporum coelestiorum in sectionibus conics solem ambientum*. Cambridge University Press, 1809.

[53] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Prentice Hall, 2010.

[54] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings -F*, 140:107–113, 1993.

[55] A. Goshtasby. Description and discrimination of planar shape using shape matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7:738–743, 1985.

[56] W. Grimson. *Object recognition by computer*. MIT Press, 1990.

[57] W.I. Grosky and R. Mehrotra. Index-based object recognition in pictorial data management. *Computer Vision, Graphics, and Image Processing*, 52:416–436, 1990.

[58] A. Guimaraes-Costa, M. Nascimento, G. Froment, R. Soares, F. Morgado, F. Conceicao-Silva, and E. Saraiva. Leishmania amazonensis promastigotes induce and are killed by neutrophil extracellular traps. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 106, pages 6748–6753, 2009.

[59] F. Gustafsson. *Adaptive Filtering and Change Detection*. New York: John Wiley & Sons, 2000.

[60] F. Gustafsson and G. Hendeby. Some relations between extended and unscented Kalman filters. *IEEE Transactions on Signal Processing*, 60(2):545–555, 2012.

[61] J. Hammersley and K. Morton. Poor man's Monte Carlo. *Journal of the Royal Statistical Society B*, 16:23–38, 1954.

[62] J. Handschin and D. Mayne. Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *International Journal of Control*, 9:547–559, 1969.

[63] C. Haslett. Resolution of inflammation and the role of apoptosis in the tissue fate of granulocytes. *Clinical Science*, 83:639–648, 1992.

[64] C. Haslett. Granulocyte apoptosis and its role in the resolution and control of lung inflammation. *American Journal of Respiratory and Critial Care Medicine*, 160:S5–S11, 1999.

[65] R. Helmick, W. Blair, and S. Hoffman. Interacting multiple-model approach to fixed-interval smoothing. In *Proc. 32nd IEEE Conference on Decision and Control, San Antonlo, Taxas*, volume 4, pages 3052–3057, December 1993.

[66] P. Henson. Dampening inflammation. *Nature Immunology*, 6:1179–1181, 2005.

[67] S. Holland. Chronic granulomatous disease. *Clinical Reviews in Allergy and Immunology*, 38:3–10, 2010.

[68] C. Hong, Y. Kidani, N. A-Gonzalez, T. Phung, A. Ito, X. Rong, K. Ericson, H. Mikkola, S. Beaven, L. Miller, W. Shao, P. Cohen, A. Castrillo, P. Tontonoz, and S. Bensinger. Coordinate regulation of neutrophil homeostasis by liver x receptors in mice. *Journal of Clinical Investigation*, 122:337–347, 2012.

[69] M.K. Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transaction*, 8:179–187, 1962.

[70] C.L. Huang and D.H. Huang. A content-based image retrieval system. *Image and Vision Computing*, 16:149–163, 1998.

[71] A. Huertas and G. Medione. Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(5):651–664, 1986.

[72] J. Iivarinen and A. Visa. Shape recognition of irregular objects. *Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling*, SPIE 2904:25–32, 1996.

[73] J. Jablonska, S. Leschner, K. Westphal, S. Lienenklaus, and S. Weiss. Neutrophils responsive to endogenous ifn-beta regulate tumor angiogenesis and growth in a mouse tumor model. *Journal of Clinical Investigation*, 120:1151–1164, 2010.

[74] B. Jahne and H. HauBcker, editors. *Computer Vision and Applications*. ACADEMIC PRESS, 2000.

[75] A. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, Incorporation, New York, 1970.

[76] A.M. J.B. Beltman and R. de Boer. Analysing immune cell migration. *Nature Reviews Immunology*, 9:789–798, 2009.

[77] S. Jeannin. Mpeg-7 visual part of experimentation model version 5.0. *ISO/IEC JTC1/SC29/WG11/N3321*, 2000.

[78] J.D. Jong and L. Zon. Use of zebrafish system to study primitive and definitive hematopoiesis. *Annual Review of Genetics*, 39:481–501, 2005.

[79] S.J. Julier and J.K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. *InternatiInternational Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, 3068:182–193, 1997.

[80] S. Julier and J. Uhlmann. The scaled unscented transform. In *in Proceedings of American Control Conference*, pages 4555–4559, 2002.

[81] K. Kaawaase, F. Chi, J. Shuhong, and Q.B. Ji. A review on selected target tracking algorithms. *Information Technology Journal*, 10:691–702, 2011.

[82] 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*, 7:0035182, 2012.

[83] R. Kalman. A new approach to linear filtering and prediction problems. *Transactions ASME, Journal of Basic Engineering*, 82:35–45, 1960.

[84] K.Arbter, W. Snyder, H. Burkhardt, and G. Hirzinger. Application of affine-invariant Fourier descriptors to recognition of 3-D objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:640–647, 1990.

[85] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.

[86] R. Kennedy, Y. Lee, B. Roy, C. Reed, and R. Lippmann. *Solving Data Mining Problems Through Pattern Recognition*. Prentice-Hall, PTR, Upper Saddle River, New Jersey, 1997.

[87] A. Khare and U. Tiwary. Daubechies complex wavelet transform based moving object tracking. In *Computational Intelligence in Image and Signal Processing*, volume 7, pages 663–687, 2007.

[88] M. Kliot and E. Rivlin. Invariant-based shape retrieval in pictorial databases. *Computer Vision and Image Understanding*, 71:182–197, 1998.

[89] E. Kolaczkowska and P. Kubes. Neutrophil recruitment and function in health and inflammation. *Nature Reviews Immunology*, 13:159–175, 2013.

[90] A. Kolmogorov. *Foundations of the theory of probability*. Springer, New York, 1933.

[91] K. Li, E. Miller, M. Chen, T. Kanade, L. Weiss, and P. Campbell. Cell population tracking and lineage construction with spatiotemporal context. *Medical Image Analysis*, 12:546–566, 2008.

[92] S. Li. Shape matching based on invariants. In *In Omidvar(ed), Shape Analysis, Progress in Neural Networks*, volume 6, pages 203–228, 1999.

[93] X. Li and Y. Bar-Shalom. Tracking in clutter with nearest neighbor filters: Analysis and performance. *IEEE Transactions on Aerospace and Electronic Systems*, 32:995–1010, 1996.

[94] S. Liao and M. Pawlak. On image analysis by moments. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 18:254–266, 1996.

[95] G. Lieschke. Zebrafish– an emerging genetic model for the study of cytokines and hematopoiesis in the era of functional genomics. *International Journal of Hematology*, 73:23–31, 2001.

[96] G. Lieschke. Fluorescent neutrophils throw the spotlight on inflammation. *blood*, 108(13):3961–3962, December 2006.

[97] C. Lin and R. Chellappa. Classification of partial 2D shapes using Fourier descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9: 686–690, 1987.

[98] G. Litman, J. Cannon, and L. Dishaw. Reconstruction immune phylogeny: new perspectives. *Nature Reviews Immunology*, 5:866–879, 2005.

[99] G. Lu and A. Sajjanhar. Region-based shape representation and similarity measure suitable for content-based image retrieval. *Multimedia Systems*, 7: 165–174, 1999.

[100] R.G. Lyons. *Understanding Digital Signal Processing*. Prentice Hall, 2011.

[101] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 207(1167):187–217, 1980.

[102] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press, 1982.

[103] R. Megens, K. Kemmerich, J. Pyta, C. Weber, and O. Soehnlein. Intravital imaging of phagocyte recruitment. *Thrombosis and Haemostasis*, 105:802–810, 2011.

[104] R. Mehrotra and J. Gary. Similar-shape retrieval in shape data management. *IEEE Computer*, 28:57–62, 1995.

[105] R. Menegazzi, E. Decleva, and P. Dri. Killing by neutrophil extracellular traps: fact or folklore? *Blood*, 119:1214–1216, 2012.

[106] J. Mestas and C. Hughes. Of mice and not men: differences between mouse and human immunology. *Journal of Immunology*, 172:2731–2738, 2004.

[107] O. Mitchell and T. Grogan. Global and partial shape discrimination for computer vision. *Optical Engineering*, 23:484–491, 1984.

[108] F. Mokhtarian and A. Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:34–43, 1986.

[109] F. Mokhtarian, S. Abbasi, and J. Kittler. Efficient and robust retrieval by shape content through curvature scale space. *International Workshop on Image DataBases and Multimedia Search*, 17:539–544, 1996.

[110] M. Morf and T. Kailath. Square-root algorithms for least-squares estimation. *IEEE Transactions Automatic Control*, AC-20(4):487–497, August 1975.

[111] J. Moura. What is signal processing? *IEEE Signal Processing Magazine*, 26(6): 6–6, 2009.

[112] W. Nauseef. How human neutrophils kill and degrade microbes: An integrated view. *Immunol. Rev.*, 219:88–102, 2007.

[113] A.V. Oppenheim and R.W. Schafer. *Digital Signal Processing*. Prentice Hall, 1975.

[114] A.V. Oppenheim, R.W. Schafer, and J.R. Buck. *Discrete-Time Signal Processing*. Prentice Hall, 1999.

[115] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems Man and Cybernetics*, 9(1):62–66, 1979.

[116] A.K.H.K. P.A. Negulescu, T.B. Krasieva and M. Cahalan. Polarity of t cell shape, motility, and sensitivity to antigen. *Immunity*, 4:421–430, 1996.

[117] T. Pavlidis. *Algorithms for Graphics and Image Processing*. Computer Science Press, 1982.

[118] K. Pearson. The problem of the random walk. *Nature*, 72:294, July 1905.

[119] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12: 629–639, 1990.

[120] A. Peters. Just how big is the pulmonary granulocyte pool? *Clinical Science*, 94:7–19, 1998.

[121] M. Peura and J. Iivarinen. Efficiency of simple shape descriptors. In *In Proceedings of the Third International Workshop on Visual Form, Capri, Italy*, pages 443–451, May 1997.

[122] J. Pillay, I. den Braber, N. Vrisekoop, L. Kwast, R. de Boer, J. Borghans, K. Tesselaar, and L. Koenderman. In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. *Blood*, 116:625–627, 2010.

[123] J. Prewitt. Object enhancement and extraction. In *Picture Processing and Psychopictorics*, pages 75–149, 1970.

[124] N.C. R. Delgado-Gonzalo and M. Unser. A new hybrid Bayesian-variational particle filter with application to mitotic cell tracking. volume 48, pages 1917–1920, 2011.

[125] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi. Tracking deforming objects using particle filtering for geometric active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1470–1475, August 2007.

[126] T. Rauber. Two-dimensional shape description. Technical report, University Nova de Lisboa, 1994.

[127] N. Ray, S. Acton, and K. Ley. Tracking leukocytes in vivo with shape and size constrained active contours. *IEEE Transactions on Medical Imaging*, 21: 1222–1235, 2002.

[128] S. Renshaw, C. Loynes, D. Trushell, S. Elworthy, P. Ingham, and M. Whyte. A transgenic zebrafish model of neutrophilic inflammation. *Blood*, 108(13): 3976–3978, December 2006.

[129] C. Reyes-Aldasoro, Y. Zhao, D. Coca, S. Billings, V. Kadirkamanathan, G. Tozer, and S. Renshaw. Analysis of immune cell function using in vivo cell shape analysis and tracking. *ResearchGate*, pages 1–10, 2009.

[130] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter*. ARTECH HOUSE, Incorporation, 2004.

[131] L.G. Roberts. Machine perception of three-dimensional solids. Technical report, MIT Lincoln Laboratory Technical Report No 315 Boston, 1965.

[132] W.J. Rucklidge. Efficient locating objects using Hausdorff distance. *Int. J. Comput. Vision*, 24:251–270, 1997.

[133] Y. Rui, A. She, and T. Huang. A modified Fourier descriptor for shape matching in MARS. *Image Databases and Multimedia Search*, pages 165–177, 1997.

[134] A. Sage and J. Melsa. *Estimation Theory with Applications to Communications and Control*. McGraw-Hill Book Company, Incorporation, New York, 1971.

[135] T. Saitoh, J. Komano, Y. Saitoh, T. Misawa, M. Takahama, T. Kozaki, T. Uehata, H. Iwasaki, H. Omori, S. Yamaoka, N. Yamamoto, and S. Akira. Neutrophil extracellular traps mediate a host defense response to human immunodeficiency virus-1. *Cell host and microbe*, 12:109–116, 2012.

[136] B. Scassellati, S. Slexopoulos, and M. Flickner. Retrieving images by 2d shape: a comparison of computation methods with human perceptual judgments. In *SPIE Proceedings on Storage and Retrieval for Image and Video Databases II*, volume 2185, pages 2–14, 1994.

[137] G. Schmidt. Designing nonlinear filters based on Daum's theory. *Journal of Guidance, Control and Dynamics*, 16:371–376, 1993.

[138] G. Scott. The alternative snake – and other animals. In *In proceedings of 3rd Alvey Vision Conference*, pages 341–347, 1987.

[139] C. Serhan. Novel lipid mediators and resolution mechanisms in acute inflammation: to resolve or not? *American Journal of Pathology*, 177:1576–1591, 2010.

[140] L. Shapiro and A. Rosenfeld, editors. *Computer Vision and Image Processing*. ACADEMIC PRESS, INC., 1992.

[141] J. Shi, G. Gilbert, Y. Kokubo, and T. Ohashi. Role of the liver in regulating numbers of circulating neutrophils. *Blood*, 98:1226–1230, 2001.

[142] I. Sobel. *Camera Models and Machine Perception*. PhD thesis, Stanford University, 1970.

[143] O. Soehnlein and L. Lindbom. Phagocyte partnership during the onset and resolution of inflammation. *Nature Reviews Immunology*, 10:427–439, 2010.

[144] M. Sonka, V. Hlavac, and R. Boyle. Image processing, analysis and machine vision. *Journal of Electronic Imaging*, 5:423, 1996.

[145] D.M. Squire and T.M. Caelli. Invariance signatures: Characterizing contours by their depatures from invariance. *Computer Vision and Image Understanding*, 77:284–316, 2000.

[146] C. Summers, S. Rankin, A. Condliffe, N. Singh, A. Peters, and E. Chilvers. Neutrophil kinetics in health and disease. *Trends in Immunology*, 31:318–324, 2010.

[147] G. Taubin and D.B. Cooper. *Object recognition based on moment (or Algebraic) Invariants*. MIT Press, Cambridge, MA, 1992.

[148] A. Taza and C. Suen. Discrimination of planar shapes using shape matrices. *IEEE Transactions on Systems Man and Cybernetics*, 19:1281–1289, 1989.

[149] M. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70:920–930, 1980.

[150] C.H. Teh and R.T. Chin. On image analysis by the methods of moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:496–513, 1988.

[151] D. Terzopoulos. On matching deformable models to images. *Topical Meeting on Machine Vision, Technical Digest Series*, 12:160–167, 1987.

[152] D. Terzopoulos and R. Szeliski. *Tracking with Kalman snakes*. MIT Press, 1992.

[153] C. Thisse and L. Zon. Organogenesis–heart and blood formation from the zebrafish point of view. *Science*, 295:457–462, 2002.

[154] Q. Tieng and W. Boles. Recognition of 2D object contours using the wavelet transform zero-crossing representation. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 19:910–916, 1997.

[155] U. von Andrian, J. Chambers, L. McEvoy, R. Bargatze, K. Arfors, and E. Butcher. Two-step model of leukocyte-endothelial cell interaction in inflammation: distinct roles for LECAM-1 and the leukocyte beta 2 integrins in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 88:7538–7542, 1991.

[156] N. Wiener. *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*. John Wiley & Sons, Incorporation, New York, 1949.

[157] N. Wiener. *Non linear problems in random theory*. Cambridge, MA, Technology Press and New York, Wiley, 1958.

[158] Y. Xiong, C.H. Huang, P.A. Lglesias, and P.N. Devreotes. Cell navigate with a local-excitation, global-inhibition-biased excitable network. *Proceeings of the National Academy of Sciences of the United States of America*, 107:17079–17086, 2010.

[159] H. Yang, S. Lee, and K. Lee. Recognition of 2D object contours using starting-point-independent wavelet coefficient matching. *J. Visual Commun. Image Represent.*, 9:171–181, 1998.

[160] I. Yong, J. Walker, and J. Bowie. An analysis technique for biological shape. i. *Graphics Image Process*, 25:357–370, 1974.

[161] L. Zadeh and J. Ragazzini. An extension of Wiener's theory of prediction. *Journal of Applied Physics*, 21:645–655, 1950.

[162] C. Zeidler, M. Germeshausen, C. Klein, and K. Welte. Clinical implications of ELA2-, HAX1-, and G-CSF-receptor (CSF3R) mutations in severe congenital neutropenia. *British Journal of Haematology*, 144:459–467, 2009.

[163] D. Zhang and G. Lu. Generic Fourier descriptor for shape-based image retrieval. In *IEEE International Conference on Multimedia and Expo (ICME2002)*, volume 1, pages 425–428, 2002.

[164] D. Zhang and G. Lu. Review of shape representation and descriptioin techniques. *Elsevier*, pages 1–19, 2003.

[165] D. Zhang and G. Lu. A comparison of shape retrieval using Fourier descriptors and short-time Fourier descriptors. In *Proceedings of the Second IEEE Pacific-Rim Conference on Multimedia*, volume 2195, pages 855–860, 2001.