

2017-01-01

Image Data Mining For Quality Control Of Nanofiber Reinforced Piezoelectric Nanocomposites Based On Sem Images

Zhonghua Hu

University of Texas at El Paso, cloud7@live.cn

Follow this and additional works at: https://digitalcommons.utep.edu/open_etd



Part of the [Electrical and Electronics Commons](#), and the [Industrial Engineering Commons](#)

Recommended Citation

Hu, Zhonghua, "Image Data Mining For Quality Control Of Nanofiber Reinforced Piezoelectric Nanocomposites Based On Sem Images" (2017). *Open Access Theses & Dissertations*. 668.
https://digitalcommons.utep.edu/open_etd/668

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

IMAGE DATA MINING FOR QUALITY CONTROL OF NANOFIBER
REINFORCED PIEZOELECTRIC
NANOCOMPOSITES BASED ON SEM IMAGES

ZHONGHUA HU

Doctoral Program in Electrical and Computer Engineering

APPROVED

Tzu-Liang Tseng, Ph.D., Chair

Jianguo Wu, Ph.D., Co-Chair

Sergio Cabrera, Ph.D.

Paras Mandal, Ph.D.

Virgilio Gonzalez, Ph.D.

Eric Smith, Ph.D.

Charles H Ambler, Ph.D.
Dean of the Graduate School

Copyright ©

by

Zhonghua Hu

2017

IMAGE DATA MINING FOR QUALITY CONTROL OF NANOFIBER
REINFORCED PIEZOELECTRIC
NANOCOMPOSITES BASED ON SEM IMAGES

by

ZHONGHUA HU, Master of Science

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering

(Industrial and Systems Engineering Track)

THE UNIVERSITY OF TEXAS AT EL PASO

May 2017

Acknowledgements

The research presented in this dissertation benefited from the valuable insights and support of many people. It is my pleasure to express my sincere gratitude to all of them.

First and foremost, I would like to express the deepest appreciation to my committee chair Prof. Bill Tseng who gave me continuous full support during all my UTEP life which cross Master and PhD study time. I benefit from not only his expertise in academic research, but also his great personality in all other aspects. Without his guidance and encouragement this dissertation would not have been possible.

I would also like to thank my doctoral committee co-chair Prof. Jianguo Wu. I am so lucky that supervised by him who has the attitude and the substance of a genius: he continually and convincingly conveyed a spirit of adventure in regard to research. When I was struggling in the scientific research, he always give me some innovated suggestion and ways to help me out.

In addition, I would be grateful to my doctoral committee members Prof. Sergio Cabrera, Prof. Paras Mandal, Prof. Virgilio Gonzalez and Prof. Eric Smith for serving on my committee and providing helpful suggestions to improve my dissertation.

My thanks also go to my friends in my research group: Dr. Aditya, Akundi, Hoejin Kim, Yuxin Wen, Honglun Xu, Carlos A Garcia Rosales, Thank you for being supportive and accompanying me in my PhD studies. You really make my life in UTEP enjoyable and memorable.

Lastly and most importantly, I'm blessed and I thank God for every day for everything that happens to me. Also I want to thank my wife Hua Xu and my parents for their unconditional love, support and encouragement. Without my wife's tender loving care, I cannot hang on there until this mission accomplished. I also want to give my special thanks to my parents. Whether I have good time or tough time, they provide me endless love and support to realize my own dream with no hesitation. I am very happy born in this family and so proud of it.

Abstract

Recently, portable and wireless nano-scaled devices have been widely used in environmental monitoring, medical implants, defense technology, industrial safety, and personal electronics, such as the nanowire-based gas and chemical sensors and programmable nanowire circuit for nano-processors. There has been an increasing demand for high energy density capacitors that can be engineered for various applications in advanced devices. In it, piezoelectric nanofiber materials play critical role in producing new products. A piezo-ceramic polymer composite contains a polymer of high dielectric strength and high permittivity ceramic inclusions, making it well suited as a high energy density capacitor. Furthermore, the research shows that the well aligned piezoelectric nanofibers generate better dielectric permittivity performance than random distributed ones. Hence, how to evaluate the nanofiber alignment for quality control becomes very critical. However, the extremely small size of nanofiber makes the quality control evaluation process very difficult. Currently, the standard quality inspection technique is the morphology analysis of nanofibers embedded in the base material based on microscopic images, e.g., scanning electron microscope (SEM) images. Visual checking of these SEM images is often adopted to evaluate the nanofiber alignment, which is often subjective, inaccurate and time-consuming. Therefore, how to automatically extract the nanofibers (number, sizes, locations and orientations) is highly desirable. The objective of this research is to fill such need by developing various image data mining methods for automatical nanofiber alignment evaluations through the SEM images. These image data mining methods are proposed to extract the information of the nanofibers through SEM images. The first method use a series of image processing algorithms (thinning, windowing, transforming and convolution process) to get the probability distribution to find the orientation of the nanofibers in SEM images. Second method use Hough transform based algorithms to segment the nanofibers from SEM images. The third method use a statistics method (cost function based multiple changepoints method) base on the boundary data of the nanofibers to extract the parameters (size, location, orientation) of nanofibers in SEM images. Finally, through

the comparison of these image data mining methods, the advantages and weak points of each method can be summarized to help people improve them in future quality control results.

Table of Contents

Acknowledgements	iv
Abstract	v
Table of Contents	vii
List of Tables	ix
List of Figures	x
Chapter 1 Introduction	1
1.1 Introduction.....	1
1.2 Background and literature review	2
1.3 Motivation and research objects	6
Chapter 2 Pattern Mining of Aligned Nanofibers	8
2.1 Image Generation through simulation	9
2.2 Image Processing Procedure	10
2.3 Validation with simulation images.	30
2.4 Experiments in real SEM images.....	53
Chapter 3 Image Segmentation Based on Hough Transform	60
3.1 Introduction of Hough Transform.....	60
3.2 Partitioning Hough Transform.....	63
3.3 Simulation Case Study.....	64
3.4 Edge based partitioning hough transform.....	70
Chapter 4 Nanofiber segmentation based on Boundary detection and multiple Changepoints models	79
4.1 Boundary Method	79
4.2 Change Point Method	85
4.3 Mathematics Geometry Method	98
Chapter 5 Discussion and Conclusion	101
5.1 Research Work to Date	101
5.2 Future Work.....	101

References	103
Vita.....	107

List of Tables

Table 2.1 6 different orientation probability distribution cases.	31
Table 2.2 Procedure of finding orientation distribution in simulation nanofibers' images	32
Table 2.3 RSS of simulation validation	41
Table 2.4 Two new cases for Robustness Validation	42
Table 2.5 RSS of methods in case 7	46
Table 2.6 RSS in Case 8	49
Table 2.7 Comparison between different interrogate windows size with 200 nanofibers simulation.....	51
Table 2.8 Comparison between different interrogate windows size with 2000 nanofibers simulation.....	52
Table 3.1. Partitioning Hough transform algorithm for nanofiber segmentation.	64
Table 3.2 Simulation Case Study Results Data	67
Table 3.3 Some advanced edge detector algorithms.....	75
Table 3.4 Table 3.4. Edge based partitioning Hough transform algorithm for nanofiber segmentation.	76
Table 3.5 Simulation case study results	78
Table 4.1 Major procedures in the statistical method in this chapter	79
Table 4.2 Moore boundary tracking algorithm working process.....	81
Table 4.3 Some Online change point detection method list	87
Table 4.4 Pseudocode of Optimal Partitioning algorithm	90
Table 4.5 Pseudocode of PELT algorithm.....	91
Table 4.6 OP and PELT method computing time with/out Parallel Computing in seconds	98
Table 4.7 Accuracy of simulation case study using Boundary based multiple changepoints method.....	100

List of Figures

Figure 1.1 Global Nano-material market	1
Figure 1.2 Illustration of the alignment of PZT nanofibers in the PVDF matrix	4
Figure 1.3 One Scanning Electron Microscope system	5
Figure 1.4 Flowchart of this dissertation	7
Figure 2.1 The procedure of image processing	8
Figure 2.2 (a) (b) (c) (d) (e) (f) Simulation nanofiber images with different orientation distribution	10
Figure 2.3 (a) (b) (c) (d) different shapes of structuring elements	12
Figure 2.4 (a) The original image (b) Use largest disks to touch two or more boundaries	14
(c) continue scan the rest untouched boundaries by decreased the size of disks	14
(d) the dash line represent the “Skeleton” of the original image	14
Figure 2.5 (a) Part of Gaussian smoothed image (b) Part of image after skeleton operation	15
Figure 2.6 Prewitt masks	16
Figure 2.7 Sobel masks	17
Figure 2.8 (a) Part of the Gaussian smoothed image (b) Part of the Ridges image	18
Figure 2.9 (a) Original image (b) After 2-D DFT (c) Centered spectrum (d) After Log Transformation	20
Figure 2.10 (a) original image divided by 6 interrogate windows	22
(b) the Fourier spectrum of each interrogate window	22
Figure 2.11 Schematic diagram of the Fan shape mask	23
Figure 2.12 Fan shape masks when $\Delta\theta = 9^\circ, 5^\circ$ and 3°	26
Figure 2.13 Line shape masks when $\Delta\theta = 9^\circ, 5^\circ$ and 3°	29
Figure 2.14 (a) Probability distribution of Fan shape method of simulation case 1	34
(b) Probability distribution of Line shape method of simulation case 1	34
Figure 2.15 (a) Probability distribution of Fan shape method of simulation case 2	35
(b) Probability distribution of Line shape method of simulation case 2	35
Figure 2.16 (a) Probability distribution of Fan shape method of simulation case 3	36
(b) Probability distribution of Line shape method of simulation case 3	36
Figure 2.17 (a) Probability distribution of Fan shape method of simulation case 4	38
(b) Probability distribution of Line shape method of simulation case 4	38
Figure 2.18 (a) Probability distribution of Fan shape method of simulation case 5	39
(b) Probability distribution of Line shape method of simulation case 5	39
Figure 2.19 (a) Probability distribution of Fan shape method of simulation case 6	41
(b) Probability distribution of Line shape method of simulation case 6	41
Figure 2.20(a)(b)Example Simulation image with 2000 nanofibers	43
Figure 2.21 (a) Probability distribution of Fan shape method of simulation case 7	45
(b) Probability distribution of Line shape method of simulation case 7	45
Figure 2.22 2000 nanofibers’ simulation image after “Skeleton” operation	45
Figure 2.23 (a) Probability distribution of Fan shape method of simulation case 8 using $\Delta\theta = 5^\circ$	47
(b) Probability distribution of Line shape method of simulation case 8 using $\Delta\theta = 5^\circ$	47
Figure 2.24 (a) Probability distribution of Fan shape method of simulation case 8 using $\Delta\theta = 2^\circ$	49
(b) Probability distribution of Line shape method of simulation case 8 using $\Delta\theta = 2^\circ$	49

Figure 2.25 (a) Probability distribution of Fan shape method of simulation case 7 using $\omega = 800$ and $\omega = 100$ with 200 nanofibers	50
(b)) Probability distribution of Line shape method of simulation case 7 using $\omega = 800$ and $\omega = 100$ with 200 nanofibers	51
Figure 2.26 (a) Probability distribution of Fan shape method of simulation case 7 using $\omega = 800$ and $\omega = 100$ with 2000 nanofibers	52
(b)) Probability distribution of Line shape method of simulation case 7 using $\omega = 800$ and $\omega = 100$	52
With 2000 nanofibers.....	52
Figure 2.27 SEM image 1	54
Figure 2.28 (a) Probability distribution of Fan shape method of SEM image 1	55
(b) Probability distribution of Line shape method of SEM image 1	55
Figure 2.29 SEM image 2	56
Figure 2.30 (a) Probability distribution of Fan shape method of SEM image 2.....	57
(b) Probability distribution of Line shape method of SEM image 2.....	57
Figure 2.31 (a) (b) SEM image 3	58
Figure 2.32 (a) Probability distribution of Fan shape method of SEM image 3.....	59
(b) Probability distribution of Line shape method of SEM image 3.....	59
Figure 3.1 Illustration of Hough transform.....	61
Figure 3.2 Hough Transform flowchart	62
Figure 3.3 (a) Low density SEM simulation.....	65
(b) Median density SEM simulation	65
Figure 3.3 (c) High Density SEM simulation	65
(d) Extremely High Density SEM simulation.....	65
Figure 3.4 (a) 30 nanofibers simulation using simple Hough Transform method.....	66
(b) 100 nanofibers simulation using simple Hough Transform method	66
(c) 200 nanofibers simulation using simple Hough Transform method	66
(d) 400 nanofibers simulation using simple Hough Transform method	66
Figure 3.5 (a) 30 nanofibers simulation using Partitioning Hough Transform method.....	66
(b) 100 nanofibers simulation using Partitioning Hough Transform method.....	67
Figure 3.5 (c) 200 nanofibers simulation using Partitioning Hough Transform method.....	67
(d) 400 nanofibers simulation using Partitioning Hough Transform method.....	67
Figure 3.6 Accuracy Analysis of the Segmentation Method	68
Figure 3.7 Wrong Detection of simple Hough Transform.....	68
Figure 3.8 Problem exist after “Skeleton” operation	69
Figure 3.9 corner nanofibers miss detected	70
Figure 3.10 2D mask.....	72
Figure 3.11 Roberts masks.....	73
Figure 3.12 Prewitt masks.....	73
Figure 3.13 Sobel masks	73
Figure 3.14 Laplacian masks	74
Figure 3.15 (a) 30 nanofibers simulation using Edge based Partitioning Hough Transform method.....	77
(b) 100 nanofibers simulation using Edge based Partitioning Hough Transform method	77
Figure 3.15 (c) 200 nanofibers simulation using Edge based Partitioning Hough Transform method.....	77

(d) 400 nanofibers simulation using Edge based Partitioning Hough Transform method	77
Figure 3.16 Accuracy Analysis of simulation case study	78
Figure 4.1 (a) 8-neighbors Moore Neighborhood	80
(b) 4-neighbors Moore Neighborhood	80
Figure 4.2 (a) given binary region R	82
(b) Initiate the start point and find the first b_1	82
Figure 4.2 (c) Set $b = b_1, c = c_1$	82
(d) repeat the step 3 and 4	82
Figure 4.2 (e) meet the stopping criterion in step 5	83
(f) Full boundary points of region R	83
Figure 4.3(a) a simulation SEM image with 10 nanofibers	84
(b) Result after using the bwboundaries function in matlab	84
Figure 4.4 Boundary of 3 connected nanofibers	85
Figure 4.5 Example of line representation on the computer display system	93
Figure 4.6 Change points detection when penalty (a) $\beta = 0.01$ (b) $\beta = 0.1$ (c) $\beta = 1$ (d) $\beta = 2$	96
Figure 4.7 The time cost of OP method and PELT method	96
Figure 4.8 The time cost/the length of boundary data	97
Figure 4.9 Performance Analysis	97
Figure 4.10 (a) 30 nanofibers simulation using Boundary based multiple changepoints method	99
(b) 100 nanofibers simulation using Boundary based multiple changepoints method	99
Figure 4.10 (c) 200 nanofibers simulation using Boundary based multiple changepoints method	99
(d) 400 nanofibers simulation using Boundary based multiple changepoints method	99
Figure 4.11 An example of one redundant wrong detected line	100

Chapter 1 Introduction

1.1 INTRODUCTION

Nowadays, nanotechnology has been greatly developed in the recent several decades. The nanotechnology now has a more generalized description by “National Nanotechnology Initiative”, i.e., the manipulation of matter with at least one dimension sized from 1 to 100 nanometers. Scientists found that when the diameters of polymer fiber materials are shrunk from micrometers to submicrons or nanometers, some amazing characteristics would appear. These characteristics or properties include large ratio (1000 times) of surface area to volume, flexibility in surface functionalities and superior mechanical performance like stiffness and tensile strength which would be much better than any other conventional materials we knew [1]. Figure 1.1 illustrates the growth of the nano-material market through 2010 to 2014.

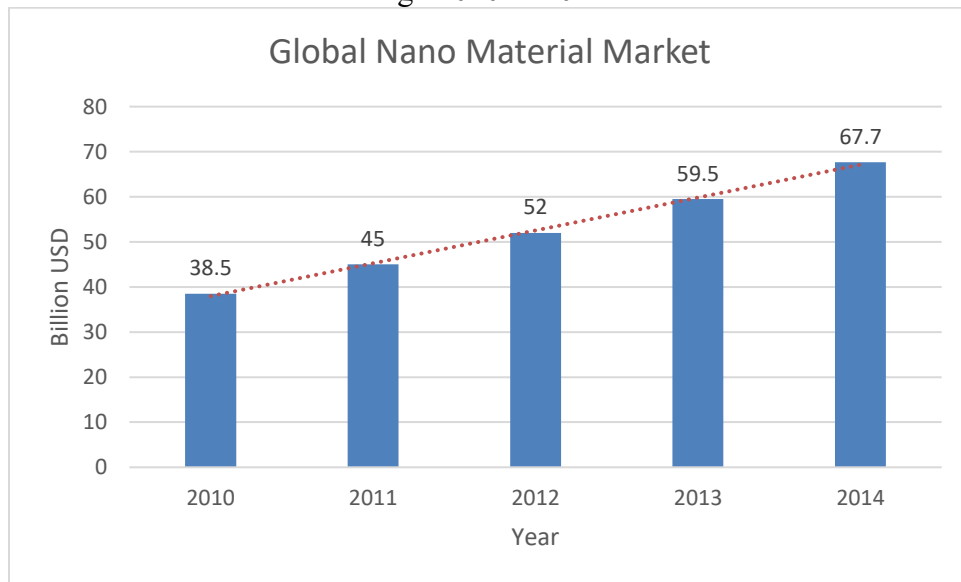


Figure 1.1 Global Nano-material market

Recently, there's a large demand of developing in electronics focuses on portable and flexible devices for applications. These applications could involve integration with the human body like health monitors, artificial muscles, sensors networks and engineered tissue constructs. On the other hand, in this “Industrial 4.0” era, the concept “Internet of Things” describe the

applications of portable and wireless devices which need new power sources beyond rechargeable batteries. Moreover, the power source should be scalable. Hence in these areas, flexible, rugged and lightweight constructions are critical requirements. Nanocomposites with high energy densities have received broad interest for use in advanced electronic devices and electric power systems since they can be designed to offer a combination of both high dielectric permittivity and dielectric strength[2, 3]. With the help of nanotechnology, new nanofiber material like piezoelectric material provide the ability to bend and stretch also they are attractive for pressure sensors and mechanical energy harvests [4].

1.2 BACKGROUND AND LITERATURE REVIEW

Piezoelectric effect is a unique property of certain crystal which generate an electric field or current if subjected to physical stress. Piezoelectric polymers are especially promising for devices with this type of functionality, because they can be exploit deformations induced by small forces, through pressure, mechanical vibration, elongation/compression, bending or twisting. These materials combine structural flexibility, ease of processing and good chemical resistance, with large sensitive areas, simplicity in device design and associated potential for low cost implementation [5].

Among piezoelectric nanofibers, piezoelectric zinc oxide (ZnO) is one of the earliest used to build the Nano-generators. In the later years, PZT piezoelectric materials has been developed. PZT is a ceramic material which show good piezoelectric performance using in nanofibers based energy harvesters. These several years, PVDF (Polyvinylidene fluoride) [5] nanofibers have been developed which have the unique good combination of material properties in flexibility, lightweight, biocompatibility and availability in ultra-long lengths various thicknesses and shapes. These excellent properties make PVDF nanofibers be a good candidate for organic Nano-generator and energy harvesting applications in wearable and implantable devices [6].

Piezoelectric nanofibers have been widely used as reinforcement to form various nanocomposites with significantly enhanced mechanical and other functional properties. The

morphology (e.g., spatial distribution, fiber orientation) of nanofibers in the base material plays a critical role in determining the enhancement of material properties. In structural applications, uniform distribution of nanofibers in terms of both spatial location and orientation is desirable to achieve the best mechanical properties. However, in some other applications, alignment of nanofibers with the same orientation is more preferable. Recently research shows that composites with aligned filler can influence the dielectric properties of a material, specifically the dielectric permittivity and breakdown strength.[7, 8] Bowen et al. created anisotropic filler for composites by using a dielectrophoretical assembly process. It was found that the sample with aligned filler had a dielectric constant approximately three times higher than that with random filler; however, the improved permittivity came at the expense of the dielectric strength. [9] This drop in dielectric strength has also been reported in the case of composites with randomly dispersed ceramic filler, otherwise referred to as 0-3 composites. Tomer et al. used the dielectrophoretical assembly method to obtain aligned barium titanate in a silicone elastomer thermoset polymer and found the dielectric permittivity in the applied electric field direction was much larger than a random one; however, the breakdown strength decreased when compared to a nanocomposite with random orientation. In the dielectric nanocomposites manufacturing used in high performance capacitors, well-aligned nanofibers could increase both permittivity and breakdown strength, and therefore increase the energy density of capacitors[10]. These results indicate that the orientation of the filler of nano composites must be carefully controlled to obtain high dielectric permittivity and breakdown strength required to achieve materials with high energy density[11]. The alignment of the nanowires can lead to energy densities up to 51.6% greater than nanocomposites with random alignment at 20% volume fraction. The maximum calculated energy density of the nanocomposites is as high as 1.28 J cm at volume fraction of 40% PZT nanowires have in PVDF with axis of alignment in the direction of the electric field. These results indicate that aligned PZT nanowires have promising potential to produce nanocomposites with high energy density[10]. Nanocomposites with aligned PZT nanofibers were prepared through uniaxial strain assembly as shown in Figure 1.2.

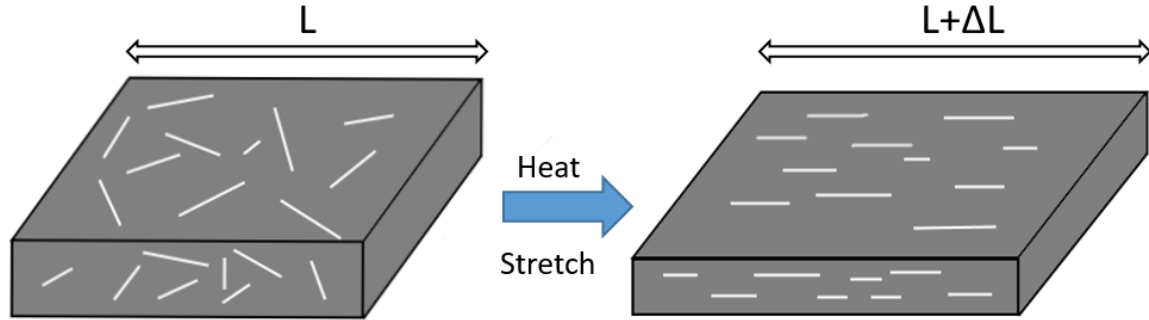


Figure 1.2 Illustration of the alignment of PZT nanofibers in the PVDF matrix

Currently, the standard quality inspection technique is morphology analysis of nanofibers imbedded in the base material based on microscopic images, e.g., scanning electron microscope (SEM) images. The scanning electron microscope (SEM) uses a focused beam of high-energy electrons to generate a variety of signals at the surface of solid specimens [68]. The signals that derive from electron-sample interactions reveal information about the sample including external morphology (texture), chemical composition, and crystalline structure and orientation of materials making up the sample. In most applications, data are collected over a selected area of the surface of the sample, and a 2-dimensional image is generated that displays spatial variations in these properties. Areas ranging from approximately 1 cm to 5 microns in width can be imaged in a scanning mode using conventional SEM techniques (magnification ranging from 20X to approximately 30,000X, spatial resolution of 50 to 100 nm). The SEM is also capable of performing analyses of selected point locations on the sample; this approach is especially useful in qualitatively or semi-quantitatively determining chemical compositions (using EDS), crystalline structure, and crystal orientations (using EBSD).

Accelerated electrons in an SEM carry significant amounts of kinetic energy, and this energy is dissipated as a variety of signals produced by electron-sample interactions when the incident electrons are decelerated in the solid sample. These signals include secondary electrons (that produce SEM images), backscattered electrons (BSE), diffracted backscattered electrons

(EBSD that are used to determine crystal structures and orientations of minerals), photons (characteristic X-rays that are used for elemental analysis and continuum X-rays), visible light (cathodoluminescence–CL), and heat. Secondary electrons and backscattered electrons are commonly used for imaging samples: secondary electrons are most valuable for showing morphology and topography on samples and backscattered electrons are most valuable for illustrating contrasts in composition in multiphase samples (i.e. for rapid phase discrimination). X-ray generation is produced by inelastic collisions of the incident electrons with electrons in discrete orbitals (shells) of atoms in the sample. As the excited electrons return to lower energy states, they yield X-rays that are of a fixed wavelength (that is related to the difference in energy levels of electrons in different shells for a given element). Thus, characteristic X-rays are produced for each element in a mineral that is "excited" by the electron beam. SEM analysis is considered to be "non-destructive"; that is, x-rays generated by electron interactions do not lead to volume loss of the sample, so it is possible to analyze the same materials repeatedly. Figure 1.3 shows the SEM devices.



Figure 1.3 One Scanning Electron Microscope system

1.3 MOTIVATION AND RESEARCH OBJECTS

The morphology analysis is often based on visual inspection of SEM images, which is subjective and lack of quantitative measurement of distribution quality of nanofibers. Therefore, the object of this research is to develop several data mining techniques to extract the features of the nanofibers through the SEM image automatically. After that, we may evaluate the accuracy of these methods, and then find the best methods to extract the nanofibers through SEM image, Eventually, we may utilize these data (length, width, location, orientation) to conduct the quality analysis.

The rest part of the dissertation is organized as follows. In Chapter 2, a new image processing method is implement to find the probability distribution of the orientation of the nanofibers in the SEM image. This could help users to get directionality of the nanofibers which is related to the physical performance. Chapter 3 presents a segmentation method which is based on Hough Transform technique commonly used in the digital image processing. In this paper, we made some modification of the Hough Transform method to improve the detection results. Chapter 4 proposes an innovative method which uses a statistical method to find critical pattern information in the image, it provide a totally different prospective compared with the traditional image processing methods. Chapter 5 summarizes the work to date and introduce the future work. Figure 1.4 illustrates the flowchart of this dissertation.

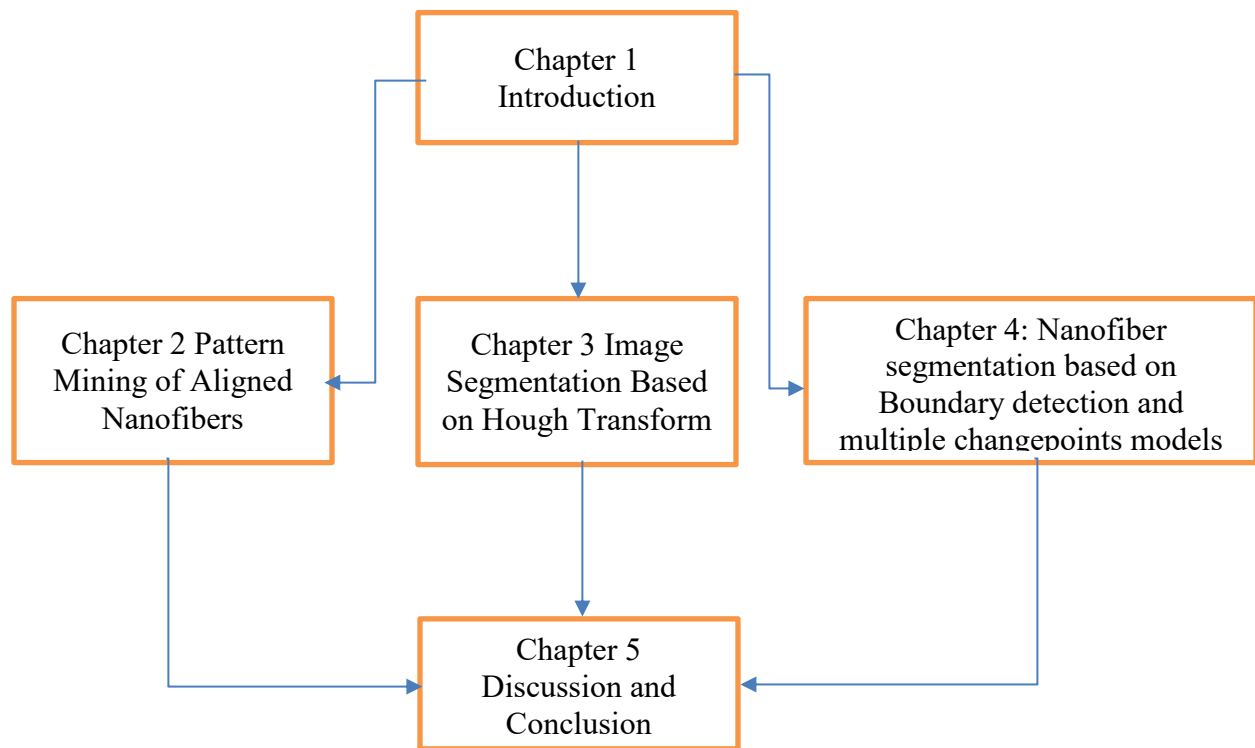


Figure 1.4 Flowchart of this dissertation

Chapter 2 Pattern Mining of Aligned Nanofibers

In Chapter 1, it is mentioned that the orientation of the nanofiber alignment is a critical quality feature of the piezoelectric nanoconposites. To optimize the physical performance of the piezoelectric nanoconposites, an image processing method based on SEM images for quality control of orientation alignment is developed in this chapter. This method includes several steps and a quantitative analysis approach to check the orientation of the nanofibers is implemented to estimate the kernel feature of nanofiber structures. Compared with isotropic orientation, well-aligned cases provide great electronic capacity of the piezoelectric nanofibers. Experimental images were collected using SEM device. The whole image processing procedure of this present study is shown in Figure 2.1[12].

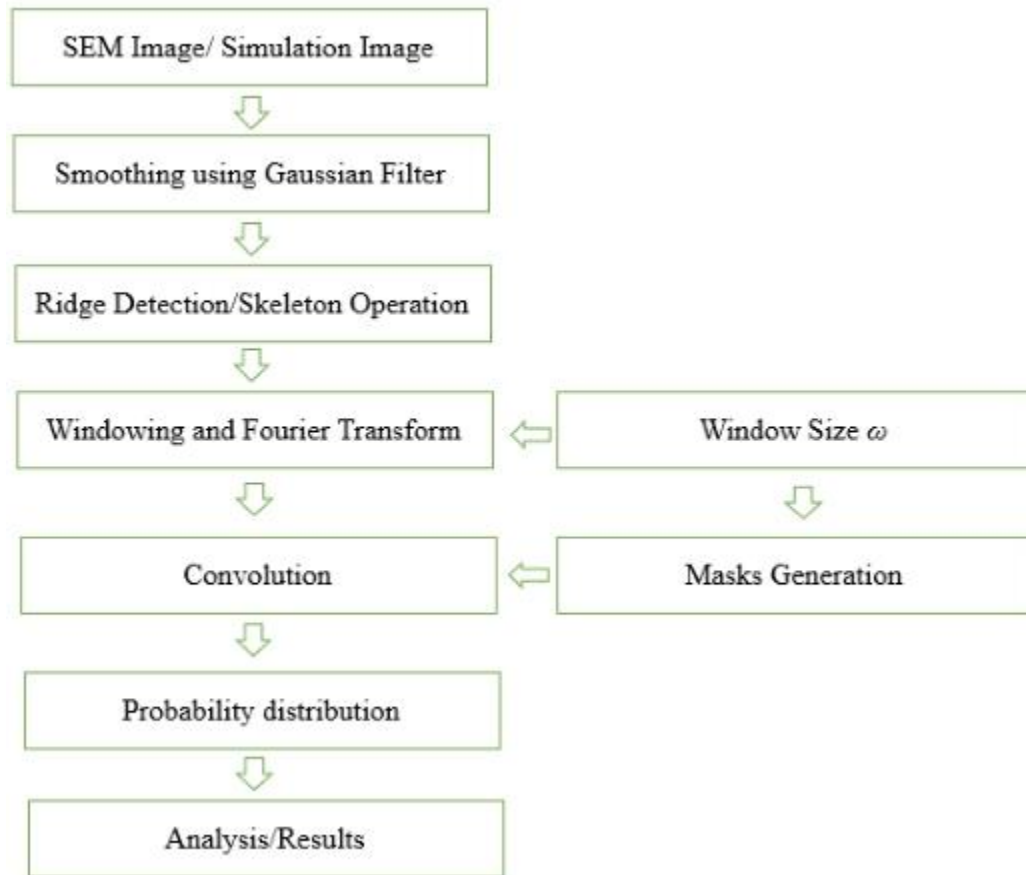


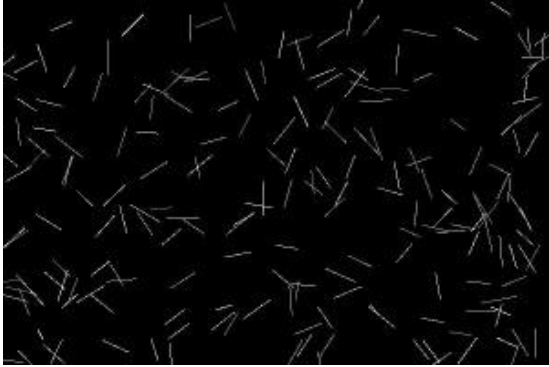
Figure 2.1 The procedure of image processing

2.1 IMAGE GENERATION THROUGH SIMULATION

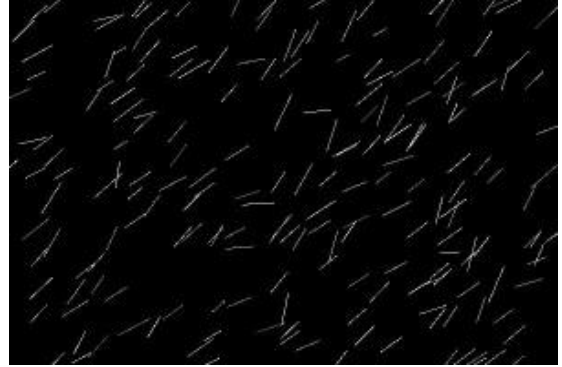
In the present research, we use the simulated images first. The simulated images filled with nanofibers were generated randomly to quantitatively evaluate the method we developed. The images were generated with 8-bit gray level. The artificial nanofibers were randomly distributed with a uniform density $\rho = 0.0005$ nanofibers per *pixel* and the length of the nanofibers are $l = 120 \pm 20$ pixels, the width of the nanofibers is a fixed value $2\sigma = 4$ pixels. For each pixel, the intensity (in Equation 2.1) was obtained by integrating one-dimensional Gaussian function related to the distance between the integration point in a pixel and the nanofiber r_i over the pixel,

$$I_{simulation}(x, y) = \sum_{i=1}^N \int_{y-0.5}^{y+0.5} \int_{x-0.5}^{x+0.5} I_i e^{\left(-\frac{r_i^2}{2\sigma^2}\right)} dx dy \quad (2.1)$$

Where N is the total number of nanofibers and I_i is the maximum intensity of each nanofiber with a normal distribution of 192 ± 32 . Simulation fibers are generated 5% random noise was introduced to simulate realistic imaging noise. Four example simulation images with different angular distribution were generated as shown in Figure 2.2.



(a)



(b)



(c)



(d)



(e)



(f)

Figure 2.2 (a) (b) (c) (d) (e) (f) Simulation nanofiber images with different orientation distribution

2.2 IMAGE PROCESSING PROCEDURE

2.2.1 Smoothing using Gaussian Filter

Before applying other image processing methods, the images were smoothed by using Gaussian low pass filter to reduce image noise. For a given image $I(x, y)$, x and y represent the pixel coordinate in the spatial domain, the smoothed image is $L(x, y, t)$ was obtained by convoluting with the Gaussian filter $g(x, y, t)$ as shown in Equation (2.2).

$$L(x, y, t) = g(x, y, t) * I(x, y) \quad (2.2)$$

$$\text{where } g(x, y, t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}$$

where t is the scale parameter. In this study, the scale parameter was appropriately selected according to the width of the nanofibers on the pixel scale.

2.2.3 Skeleton Operation

Skeleton operation is one of the commonly used morphological image processing method of interest which can be applied to binary images. Morphological techniques probe an image with a small shape or template called a structuring element. The structuring element (SE) is positioned at all possible locations in the image and it is compared with the corresponding neighborhood of pixels. Some operations test whether the element "fits" within the neighborhood, while others test whether it "hits" or intersects the neighborhood[13]. Figure 2.3, show some structuring elements with different shapes. The X denote the centers of the SE which can be used to influence some certain tasks. Notice that in Figure 2.3, SEs (c) and (d) have different representations which can produce different results in some cases.

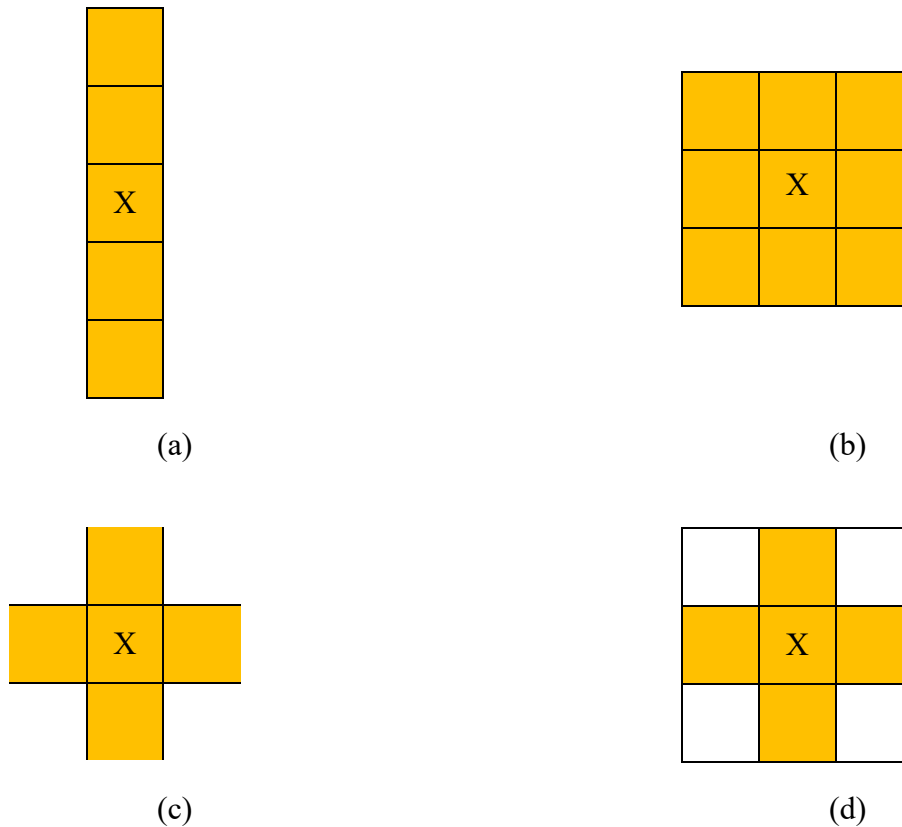


Figure 2.3 (a) (b) (c) (d) different shapes of structuring elements

A typical morphology method involves two operations: dilation and erosion. Most of the morphological processing methods are based on these operations[13].

Erosion operation of A by structuring element E is denoted as $A \ominus E$. The mathematical formulation is shown in Equation 2.3

$$A \ominus E = \{z | (E)_z \subseteq A\} \quad (2.3)$$

where z are the points left after the erosion operation. Since E has to be contained in A is equivalent to there is no common elements in E which is the background, the above formula of erosion operation can be expressed like in another statement shown in Equation 2.4 as below,

$$A \ominus E = \{z | (E)_z \cap A^c = \emptyset\} \quad (2.4)$$

where A^c is the complement of A which is the background and \emptyset is the empty set which stand for no common elements shared by structuring element E and the background.

On the other hand, the operation “Dilation” of A by E, denoted as $A \oplus E$, the formula of “Dilation” operation is shown in Equation 2.5.

$$A \oplus E = \{z | (\hat{E})_z \cap A \neq \emptyset\} \quad (2.5)$$

The equation is based on reflecting E about its origin, and shifting this reflection by z. The dilation of A by E then is the set of all displacements, z, such that \hat{E} and A overlap by at least one element. Based on this interpretation, the equation can be written equivalently in Equation 2.6 as shown below

$$A \oplus E = \{z | [(\hat{E})_z \cap A] \subseteq A\} \quad (2.6)$$

In conclusion, erosion operation shrink the components of an image and on the opposite, dilation operation expands them.

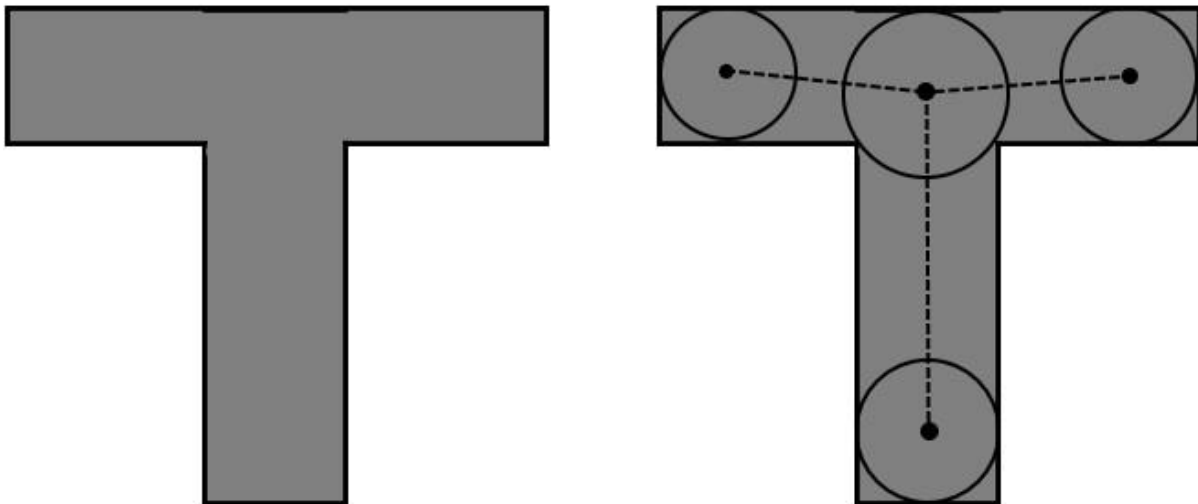
Next method “Opening” operation uses both erosion and dilation operation. Opening operation generally smooths the contour of an object, breaks narrow isthmuses, and eliminates thin protrusions. The algorithm of the opening operation is to do the erosion operation first and then do the dilation operation with the same structuring element.

The opening of set A by structuring element E , denoted as $A \circ B$, the equation 2.7 represent the “Opening” operation by using “Erosion” and “Dilation” operation.

$$A \circ B = (A \ominus B) \oplus B \quad (2.7)$$

Figure 2.4 shows, the notion of a skeleton, $S(A)$, of a set A is intuitively simple. It is deduced by

- (a) If z is a point of $S(A)$ and $(D)_z$ is the largest disk centered at z and contained in A , one cannot find a larger disk (not necessarily centered at z) containing $(D)_z$ and included in A . The disk $(D)_z$ is called a *maximum disk*.
- (b) The disk $(D)_z$ touches the boundary of A at two or more different places.



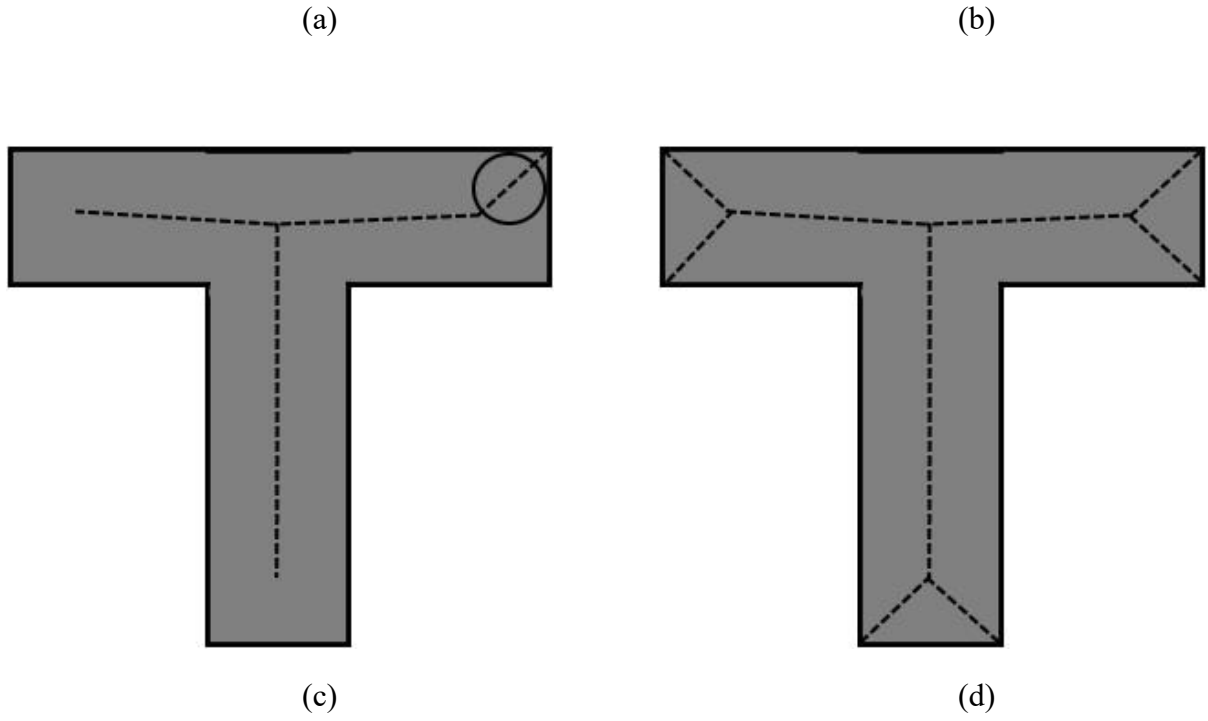


Figure 2.4 (a) The original image (b) Use largest disks to touch two or more boundaries
 (c) continue scan the rest untouched boundaries by decreased the size of disks
 (d) the dash line represent the "Skeleton" of the original image

The skeleton of A can be expressed in terms of erosion operations and opening operations. Serra [14] represent it as Equation (2.8)

$$S(A) = \bigcup_{k=0}^K S_k(A) \quad (2.8)$$

With

$$S_k(A) = (A \ominus kB) - (A \ominus kB) \circ B \quad (2.9)$$

where B is a structuring element, and $(A \ominus kB)$ indicates k successive erosions of A :

$$(A \ominus kB) = ((\dots ((A \ominus b) \ominus b) \ominus \dots) \ominus B) \quad (2.10)$$

k times, and K is the last iterative step before A erodes to an empty set. In other words,

$$K = \max\{k | (A \ominus kB) \neq \emptyset\} \quad (2.11)$$

The formulation give in the above equations and states that $S(A)$ can be obtained as the union of the *skeleton subsets* $S_k(A)$.

Figure 2.5 shows the comparison of the Gaussian smoothed image and the image after skeleton operation was applied to a binary image derived from the smoothed image.



Figure 2.5 (a) Part of Gaussian smoothed image (b) Part of image after skeleton operation

2.2.3 Ridge Detection

In mathematics, the ridges of a smoother function of two variables are a set of curves whose points are, in one or more ways to be made precise below, local maxima of the function in at least one dimension. For image processing, ridge detection is a mechanism which can extract the local maxima of the images, these ridges are made up with numerous thin frames one or two pixel thick.

The basic theory of the ridge detection method is derived by Tony Lindeberg[15]. Here, we denote the spatial derivatives of $L(x, y, t)$ as Equation (2.12)

$$L_{x^a y^b}(x, y, t) = \partial_{x^a y^b}(x, y, t) \quad (2.12)$$

where a and b are orders of differentiation. For each pixel (x, y) in the smoothed image $L(x, y, t)$, a local coordinate system (p, q) aligned along the principal curvature direction is introduced. Here, the p direction is the slope that maximizes the intensity change and the direction q is orthogonal to the direction p .

To do the ridge detection in this research, denote L_{pp} and L_{qq} as the eigenvalues of the following Hessian Matrix

$$H = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \quad (2.13)$$

of the smoothed image L with a coordinate transformation applied to local directional derivative operators as follows below Equation (2.14) and (2.15)

$$\partial_p = \sin\beta \partial_x - \cos\beta \partial_y \quad (2.14)$$

$$\partial_q = \cos\beta \partial_x + \sin\beta \partial_y \quad (2.15)$$

where p and q are coordinates of the rotated coordinate system.

In equation 2.4, L_{xx} , L_{yy} and L_{xy} can be computed by applying derivative operations in the x and y direction. Both Prewitt [16] operator and Sobel operator [17] can be used to compute discrete derivative operations, the masks are shown in Figure 2.6 and Figure 2.7.

-1	-1	-1	-1	0	1
0	0	0	-1	0	1
1	1	1	-1	0	1

Figure 2.6 Prewitt masks

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Figure 2.7 Sobel masks

It can be shown that the mixed derivative L_{pq} in the transformed coordinate system is zero if we choose an inclination angle of a ridge, β , in the q axis with respect to the x -axis given by Equation (2.16) and (2.17)

$$\cos\beta = \sqrt{\frac{1}{2} \left(1 + \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}} \right)} \quad (2.16)$$

$$\sin\beta = \text{sgn}(L_{xy}) \sqrt{\frac{1}{2} \left(1 - \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}} \right)} \quad (2.17)$$

Then, a formal differential geometric definition of the ridges of $L(x, y, t)$ can be expressed as the set of points that satisfy Equation (2.18).

$$L_p = 0, L_{pp} \leq 0, |L_{pp}| \geq |L_{qq}| \quad (2.18)$$

In image processing method, the ridges can be represent like as follow Equation (2.19)

$$\begin{aligned} \text{ridges} &= \text{Last} \left[\text{Eigenvalues} \left[\begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix} \right] \right. \\ &\quad \left. = \frac{1}{2} \left(L_{xx} + L_{yy} + \sqrt{L_{xx}^2 + 4L_{xy}^2 - 2L_{xx}L_{yy} + L_{yy}^2} \right) \right] \end{aligned} \quad (2.19)$$

In this present research, the ridges images were subsequently converted to binary (black and white) images. Figure 2.8 shown the comparison between the Gaussian smoothed image and the resulting ridges obtained from the image.



Figure 2.8 (a) Part of the Gaussian smoothed image (b) Part of the Ridges image

2.2.4 Fourier spectral approach

Fourier transform decomposes a function or a signal into the frequency components that make it up. In digital image processing, the image could be looked as a 2-D discrete signal, then after the 2-D discrete Fourier transform, the image data is changed from the spatial domain $f(x, y)$ to the frequency domain $F(u, v)$. The equation of 2-D discrete Fourier transform (DFT) is shown in Equation (2.20)

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp \left[-j2\pi \left(\frac{ux}{M} + \frac{vy}{N} \right) \right] \quad (2.20)$$

where $M \times N$ is the image size data.

Since the 2-D DFT is complex in general, it can be expressed in polar form as below Equation (2.21)

$$F(u, v) = |F(u, v)| \exp[j\phi(u, v)] \quad (2.21)$$

where $\phi(u, v)$ is the phase angle.

$$\phi(u, v) = \tan^{-1}\left(\frac{I(u, v)}{R(u, v)}\right) \quad (2.22)$$

where $I(u, v)$ is the imaginary part and real part is $R(u, v)$, then the magnitude is

$$|F(u, v)| = \sqrt{R^2(u, v) + I^2(u, v)} \quad (2.23)$$

Is called the Fourier spectrum. When $u = 0$ and $v = 0$ then

$$F(0,0) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp\left[-j2\pi\left(\frac{0 \times x}{M} + \frac{0 \times y}{N}\right)\right] = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \quad (2.24)$$

which indicates that the zero frequency term is proportional to the average of $f(x, y)$. That is,

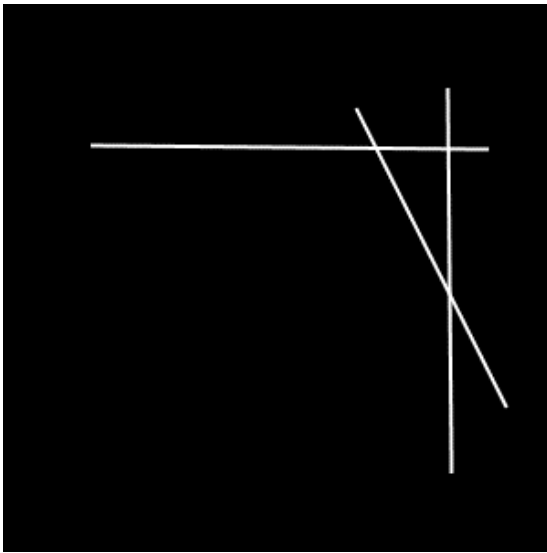
$$F(0,0) = MN \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) = MN \bar{f}(x, y) \quad (2.25)$$

where \bar{f} denotes the average value of f . Then

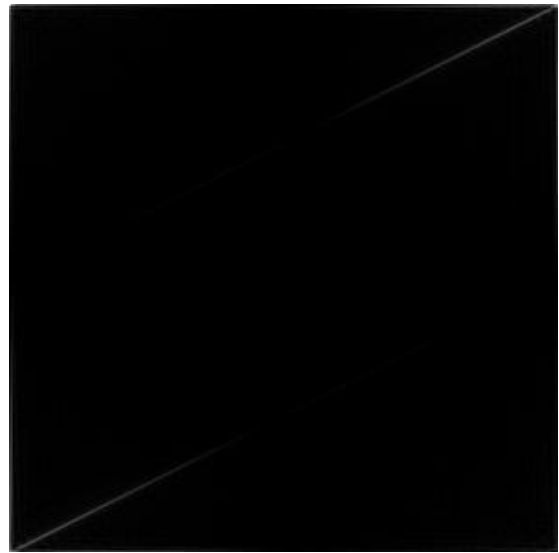
$$F(0,0) = MN |\bar{f}(x, y)| \quad (2.26)$$

Since the result of multiplying by M by N usually is very large, the center of the frequency domain $F(0,0)$ is the largest component of the spectrum by a factor that can be several orders of magnitude larger than other spectrum components.

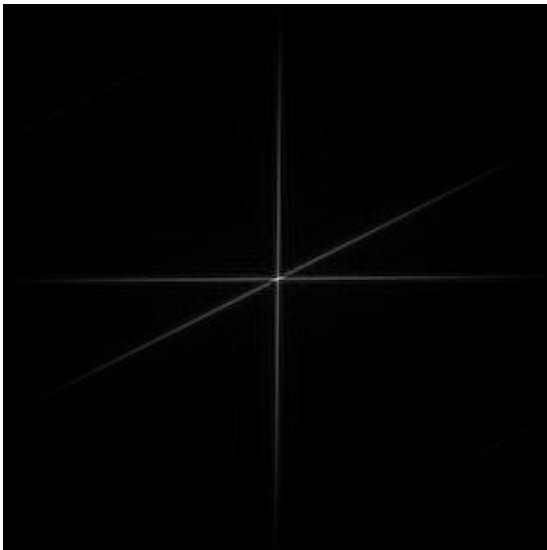
Figure 2.9 shows one example of an image to be shown in the frequency domain, Figure 2.9 (a) is the original image, Figure 2.9 (b) is the Fourier spectrum of the original image, Figure 2.9 (c) is the centered Fourier spectrum and Figure 2.9 (d) shows increased visibility of the detail after a log transformation.



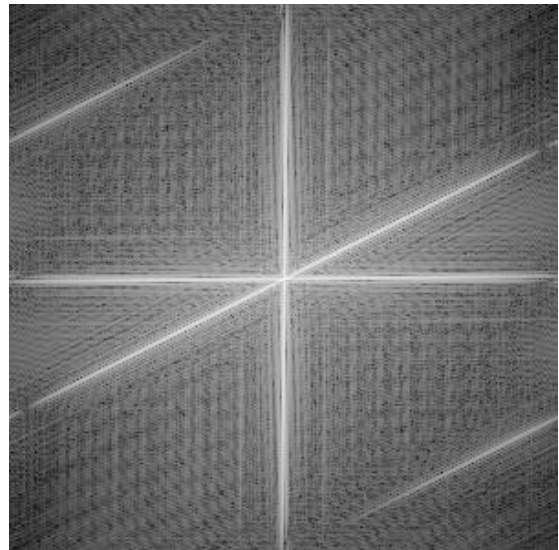
(a)



(b)



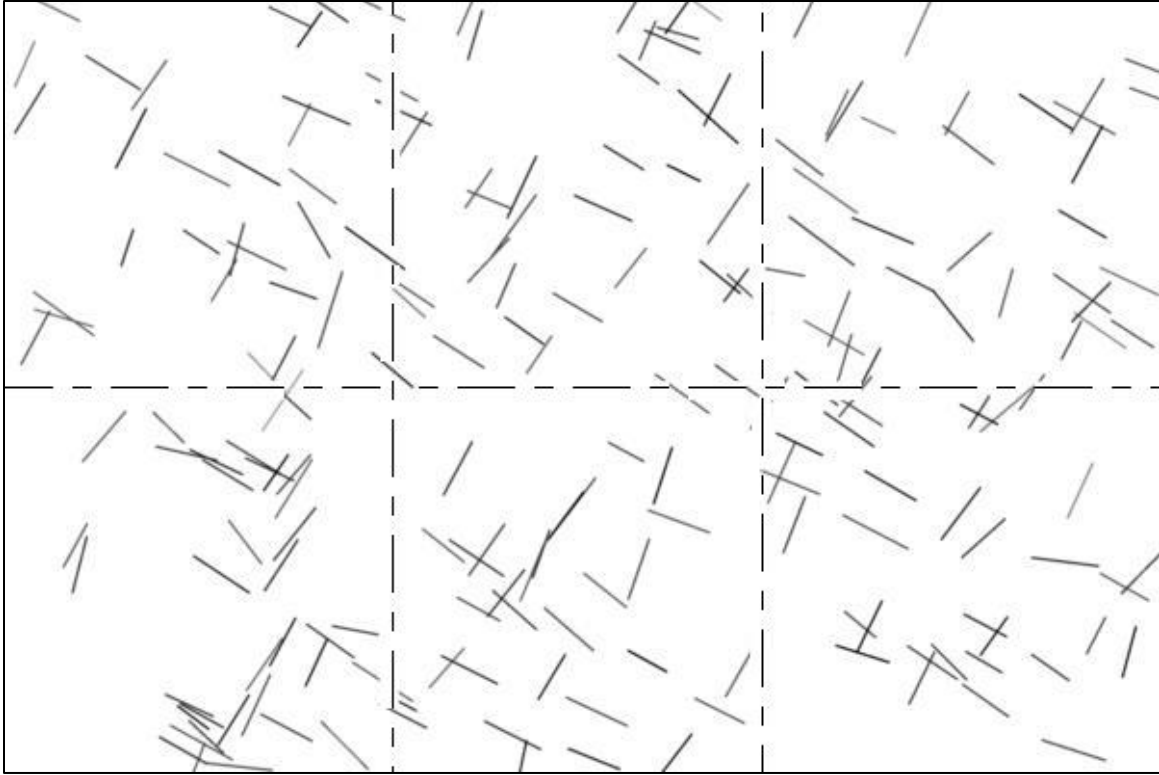
(b)



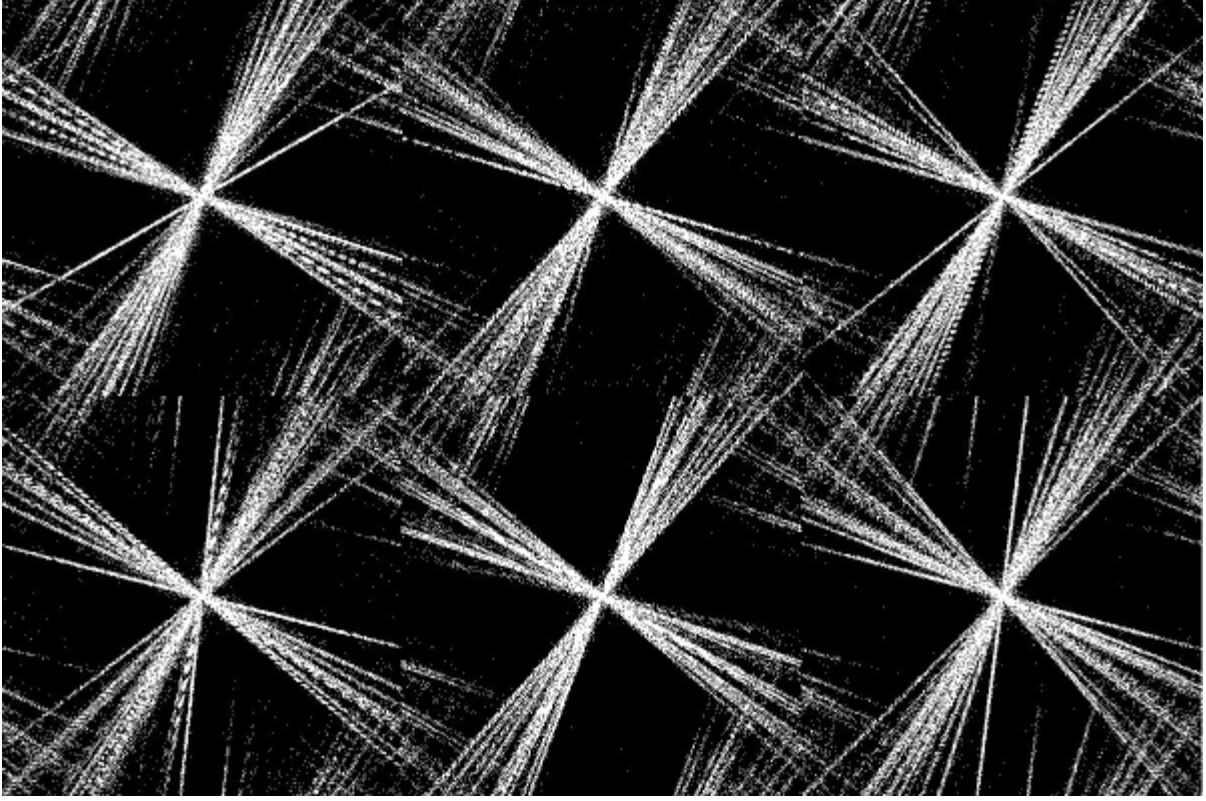
(d)

Figure 2.9 (a) Original image (b) After 2-D DFT (c) Centered spectrum (d) After Log Transformation

We may find that in the Fourier spectrum image, the prominent lines indicate the principal direction in the original image. We may use this feature to get the orientation of the nanofibers. Figure 2.10 (a)(b) show the original image divided by 6 interrogate window and Fourier spectrum of each window. The bright lines' orientation in each spectrum corresponding the orientation of the nanofibers in the original interrogate windows.



(a)



(b)

Figure 2.10 (a) original image divided by 6 interrogate windows

(b) the Fourier spectrum of each interrogate window

2.2.5 Detection Orientation with Fan and Line shape Masks

In section 2.2.4, we know that the Fourier spectrum is ideally suited for describing the directionality of 2-D patterns in an image. Thus, the task is to use some method to detect the orientation of these 2-D patterns. In the Fourier spectrum of an image, there are some brighter lines. To obtain the orientation of these lines, a contour-texture based on user-defined mask was adapted. A mask is a set of patterns that represent a set of response vector images, therefore, masks in this research should resemble lines that are widened along a single direction.

In this research, two shape masks would be created and analyzed to investigate the performances. The first one is fan shape masks method. The fan shape masks form a series of square images with ω pixels, and each image has two symmetric sectors with an internal angle θ .

The inner diameters of the sectors are selected to be ω and $\frac{1}{8}\omega$ pixels, respectively. The inner diameter is applied to eliminate correlations associated with same displacements. Figure 2.11 shows schematic diagram illustrating the fan shape mask generation procedure.

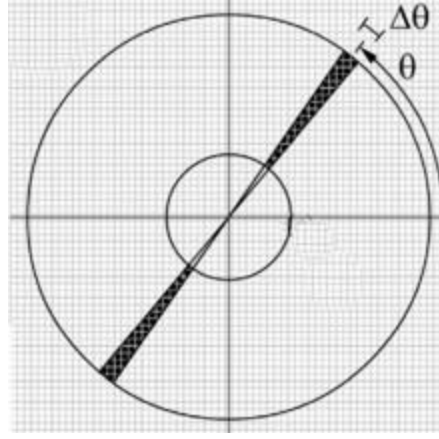


Figure 2.11 Schematic diagram of the Fan shape mask

Thus, the fan mask are defined as in equation

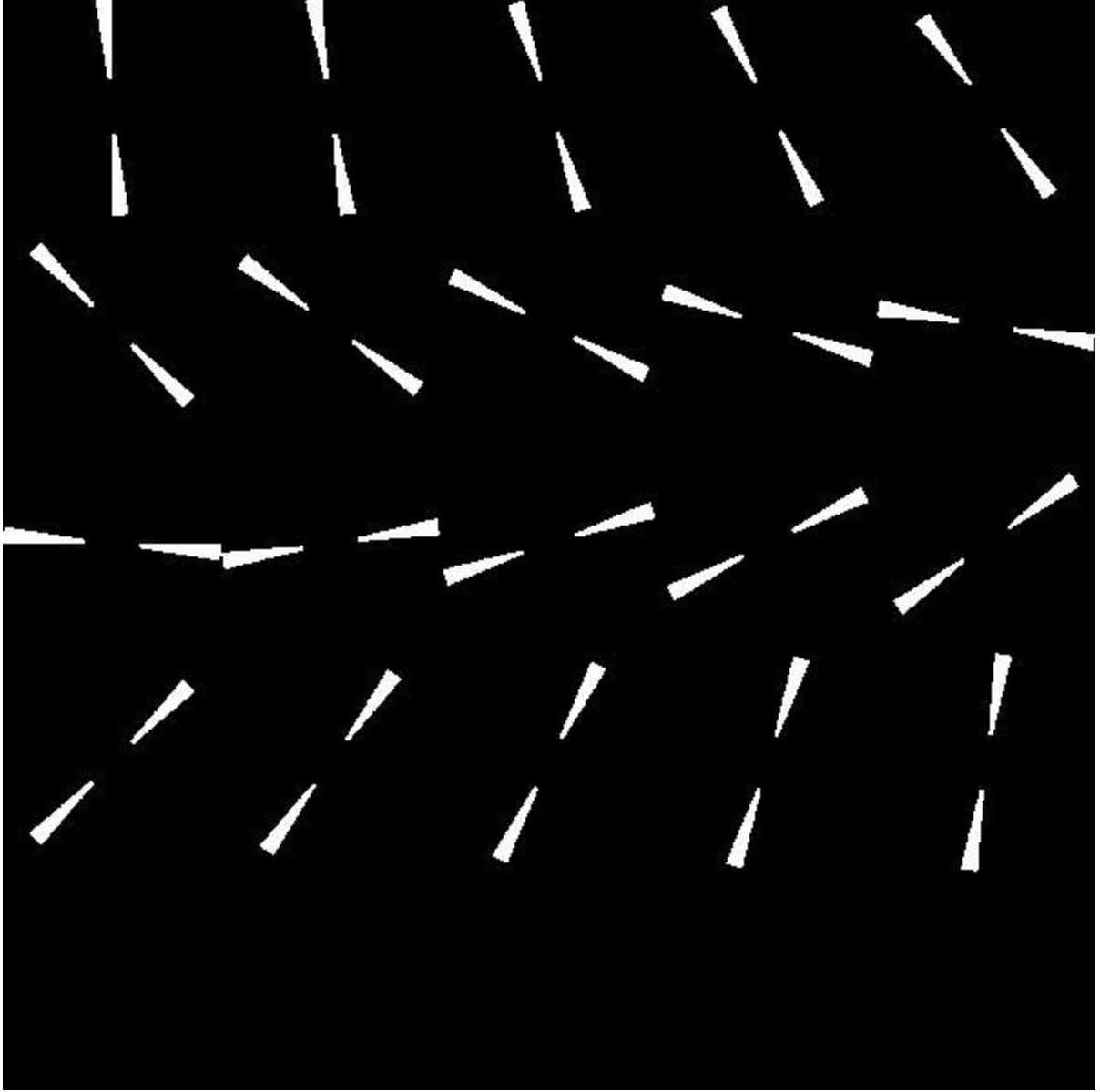
$$M(x, y, \theta) = \int_{y-0.5}^{y+0.5} \int_{x-0.5}^{x+0.5} B(x, y, \theta) dx dy \quad (2.27)$$

where $B(x, y, \theta)$

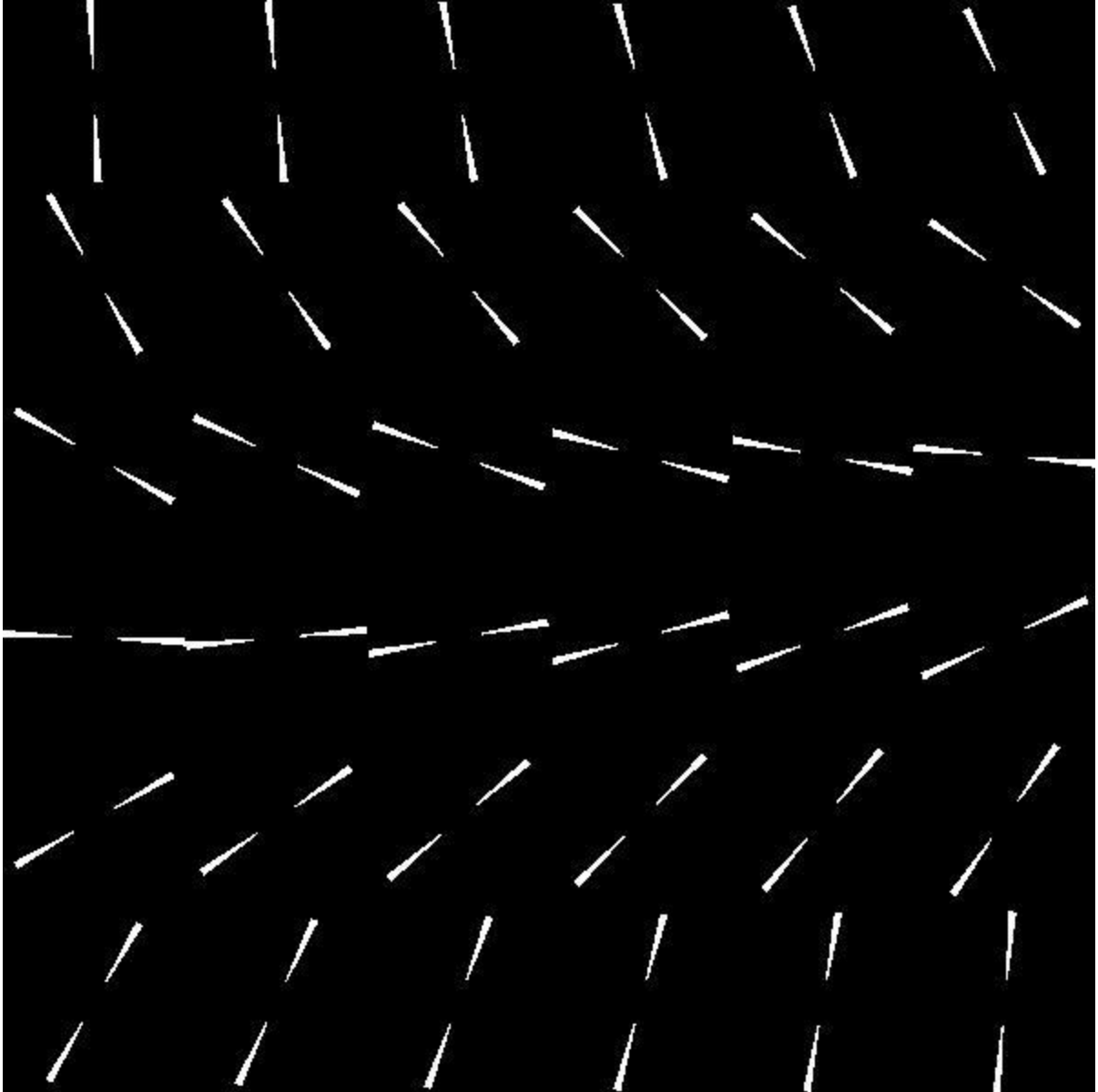
$$= \begin{cases} 1, & \text{if } \theta - \frac{\Delta\theta}{2} \leq \tan^{-1}\left(\frac{y}{x}\right) < \theta + \frac{\Delta\theta}{2} \\ & \text{and } \frac{\omega}{8} \leq \sqrt{x^2 + y^2} \leq \frac{\omega}{2} \\ 0, & \text{otherwise} \end{cases}$$

Here, $M(x, y, \theta)$ is a single fan mask, x and y are the coordinates of the displacements in the interrogated square region, and θ is the orientation angle of the mask. The value (0 or 1) at each pixel of a mask is obtained by integrating the sector function $B(x, y, \theta)$ over a pixel, as shown in Figure 2.8. To cover all direction, the total number of masks should be $i = \pi/\Delta\theta$. M

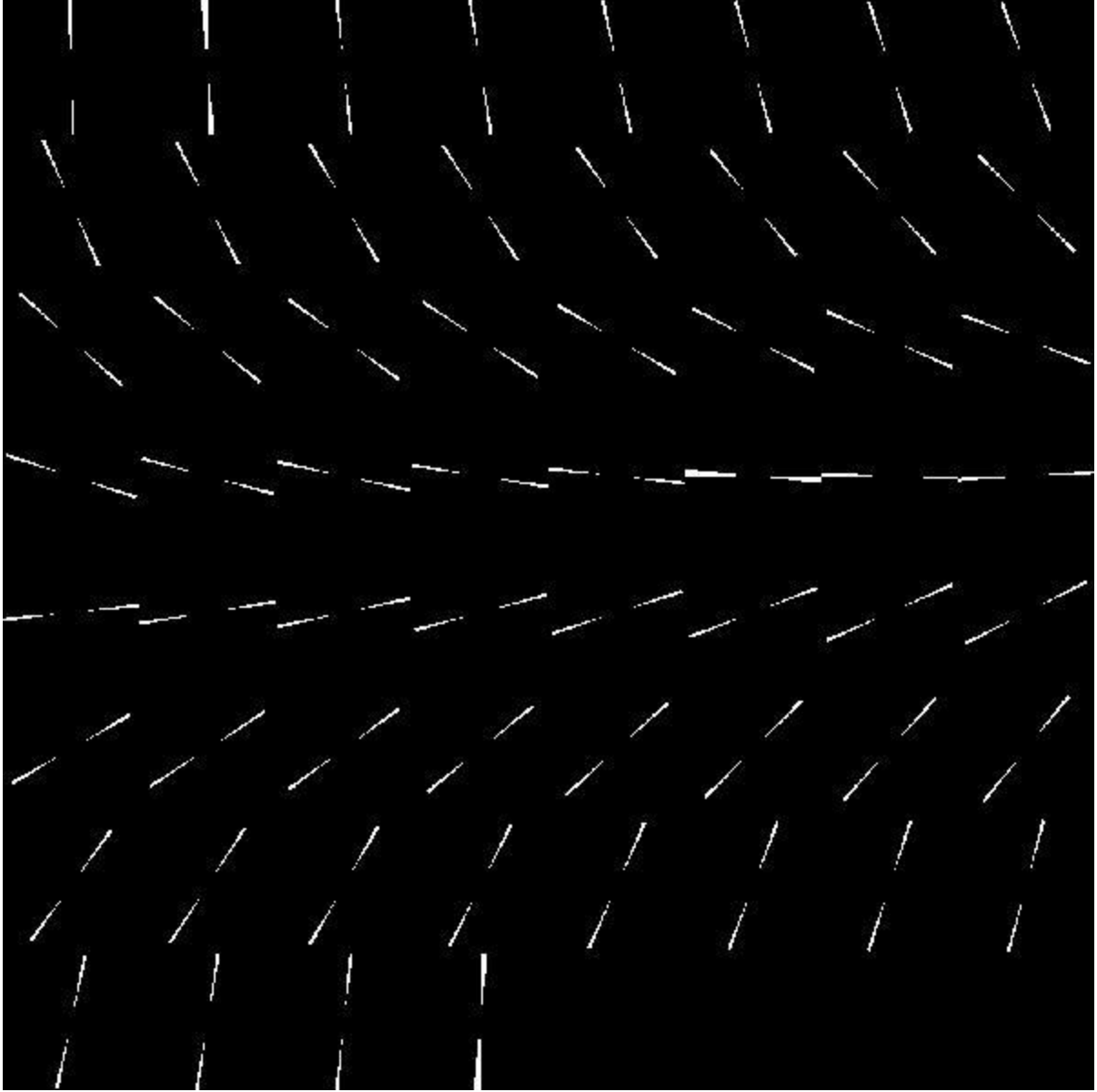
should be chosen as $M < \frac{\pi\omega}{2\sqrt{2}}$ for guaranteeing one complete pixel in the diagonal masks. Figure 2.12 (a) (b) (c) shows the fan shape masks for different parameters $\Delta\theta$ when $\Delta\theta = 9^\circ, 5^\circ$ and 3° , $i = 20, 36$ and 60 respectively.



(a)



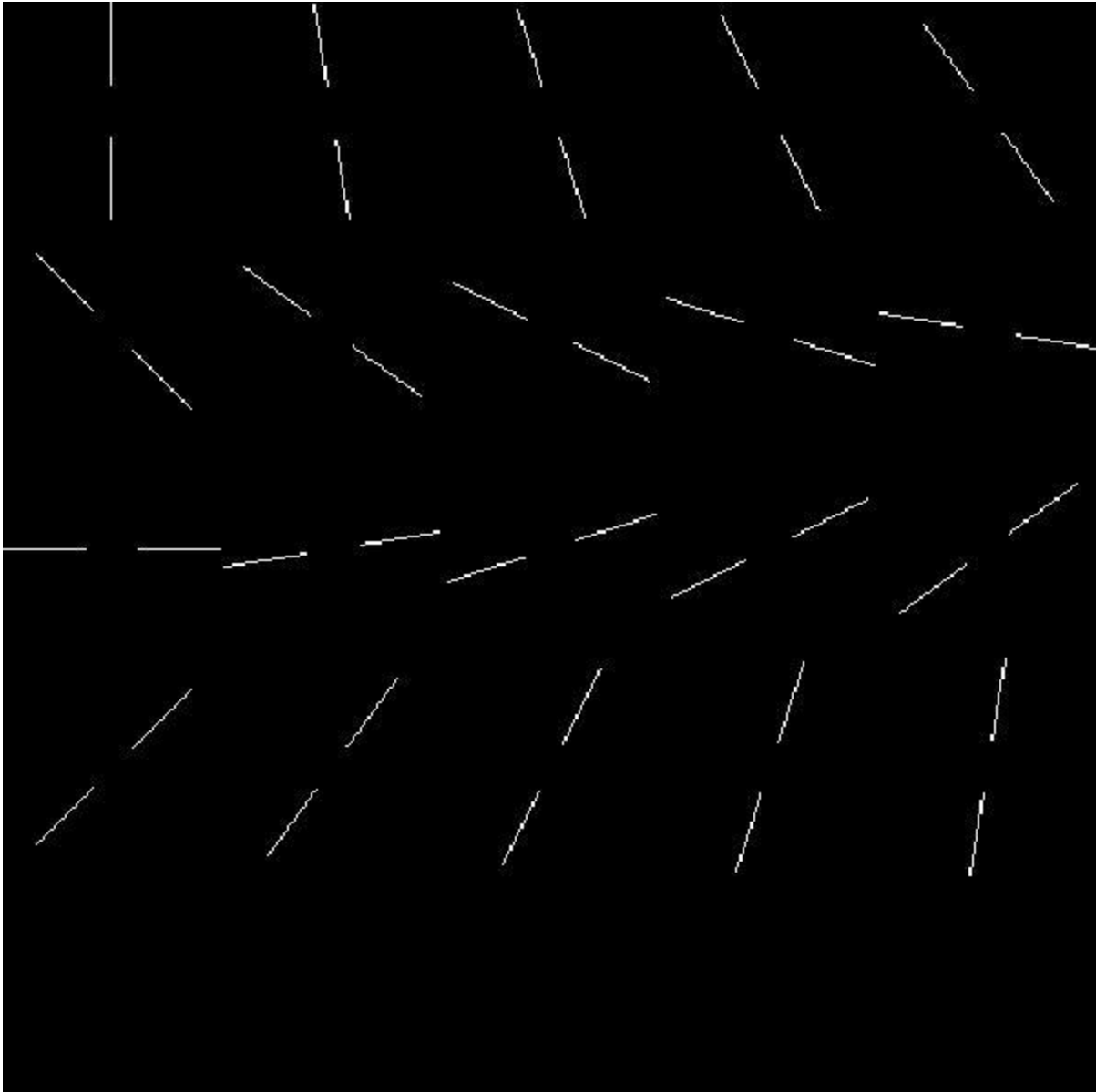
(b)



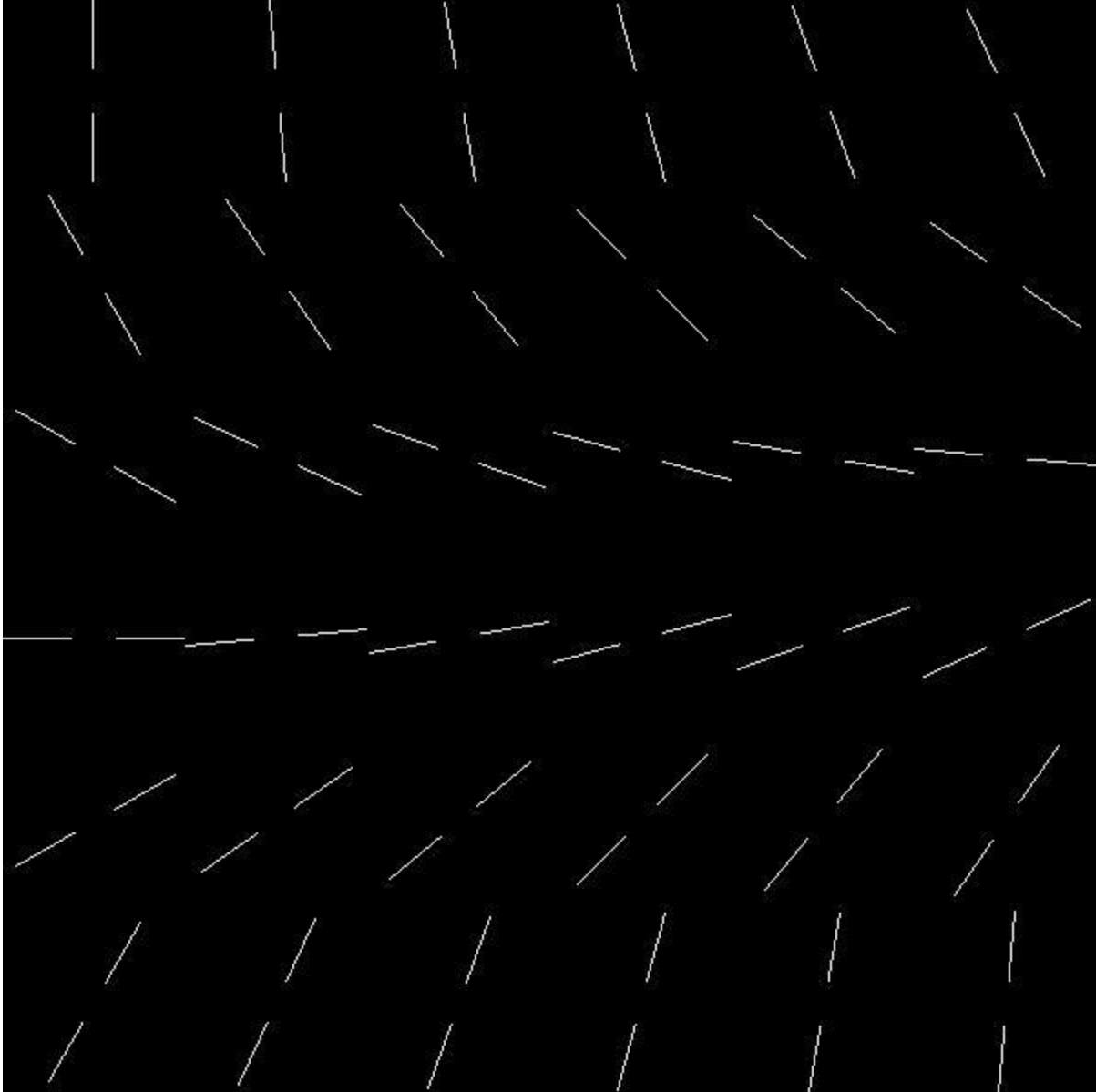
(c)

Figure 2.12 Fan shape masks when $\Delta\theta = 9^\circ, 5^\circ$ and 3°

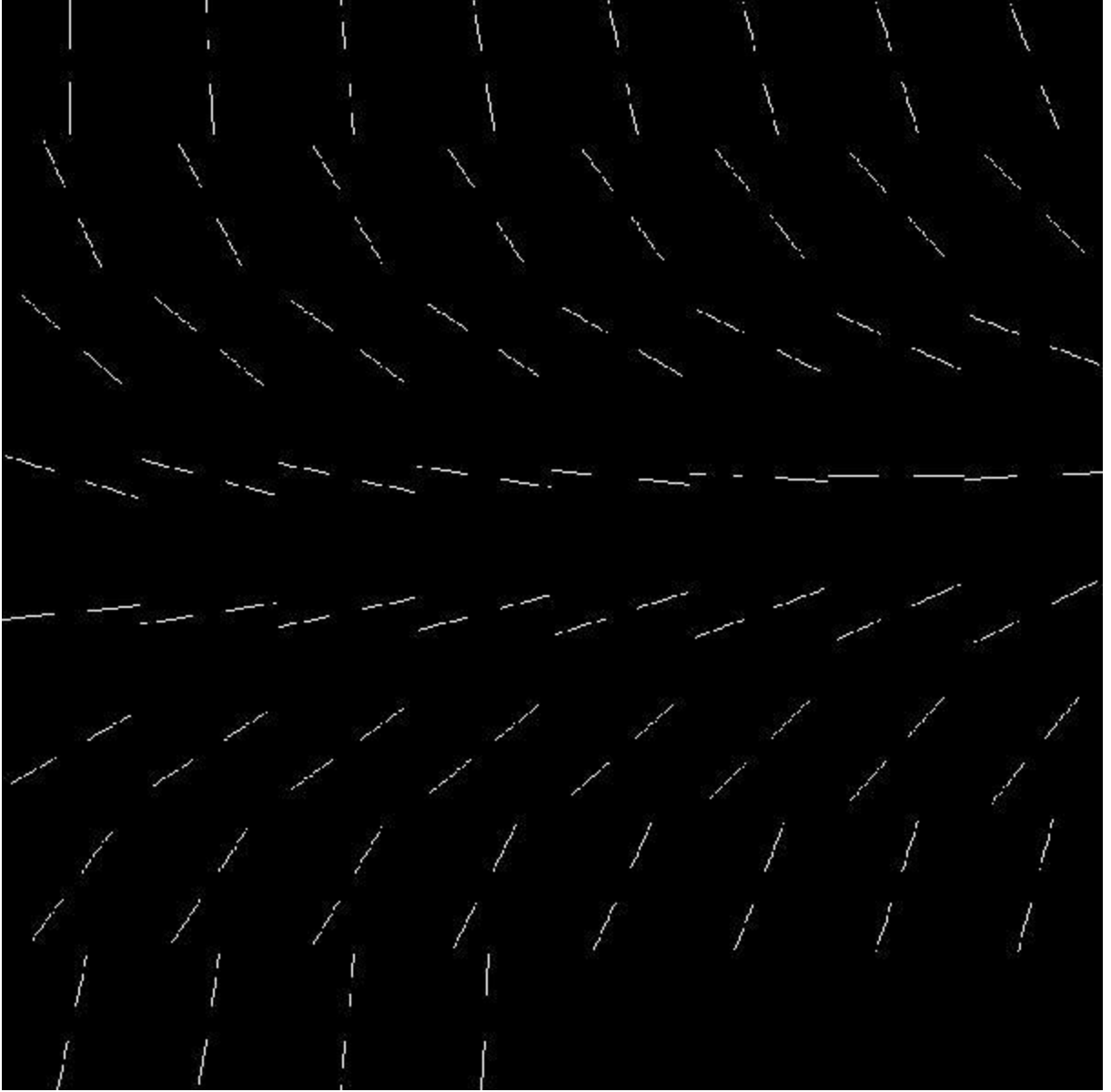
For the second method we use it the line shape masks, similar with the previous method, the line shape masks method also remove the inner diameter with $\omega/8$ area to prevent the uncorrected filtering. As the name, the line shape mask consist of a one-pixel thick single line. Figure 2.13 (a) (b) (c) show some masks group with different parameter $\Delta\theta = 9^\circ, 5^\circ, 3^\circ$.



(a)



(b)



(c)

Figure 2.13 Line shape masks when $\Delta\theta = 9^\circ, 5^\circ$ and 3°

Before we implement the method, we may set a partially interrogated square region with ω pixels length and width, it should be the same size with the created masks. Then, the image has been separated by several connected section by the square region. After that, each region will be transform from spatial domain to frequency domain by the Fourier transform then scanned by different masks to detect the orientation of the lines in this region. After one region done all the

interrogation by the masks, the window move to the next, repeat the iteration of scanning until the window move to the last section of the image. At last, calculate the detected orientation of all the sections, get the overall orientation distribution. The probability distribution of the orientation can be estimated as Equation 2.29

$$P(\theta) = \sum_{x=1}^{\omega} \sum_{y=1}^{\omega} S(x, y) \cdot M(x, y, \theta) \quad (2.28)$$

The peaks of the $P(\theta)$ could represent the orientation in this interrogation. Accumulate the data from all the interrogation windows can get the probability density function (PDF) distribution.

2.3 VALIDATION WITH SIMULATION IMAGES.

2.3.1 Simulation condition

The present method was method was evaluated using Monte Carlo simulations. The angular distribution function $p(\theta)$ is listed in Table 2.1 and a total of 100 images with 2400×1600 $pixel^2$ with 200 simulated nanofibers in each were generated for each case. The scale parameter $t = \text{width of nanofiber} = 4$ is will be also used.

Table 2.1 6 different orientation probability distribution cases.

Case Number	Case	Probability density function(PDF)
1	Random angle(uniform distribution)	$P_1(\theta) = 1/180$
2	Single Normal distribution with $\mu = 50, \sigma = 15$	$P_2(\theta) = \frac{1}{\sqrt{2\pi} \cdot 15^2} \exp \left[-\frac{(\theta - 50)^2}{2 \cdot 15^2} \right]$
3	Single Normal distribution with $\mu = -30, \sigma = 5$	$P_3(\theta) = \frac{1}{\sqrt{2\pi} \cdot 5^2} \exp \left[-\frac{(\theta - (-30))^2}{2 \cdot 5^2} \right]$
4	Two normal distribution with $\mu_1 = 30, \sigma_1 = 10$ & $\mu_2 = -60, \sigma_2 = 10$	$P_4(\theta) = \frac{1}{\sqrt{2\pi} \cdot 10^2} \exp \left[-\frac{(\theta - 30)^2}{2 \cdot 10^2} \right] + \frac{1}{\sqrt{2\pi} \cdot 10^2} \exp \left[-\frac{(\theta - (-60))^2}{2 \cdot 10^2} \right]$
5	3 peaks	$P_5(\theta) = \begin{cases} \frac{1}{3}, \theta = -70 \\ \frac{1}{3}, \theta = -30 \\ \frac{1}{3}, \theta = 30 \end{cases}$
6	One angle, $\theta = 45$	$P_6(\theta) = 1$

2.3.2 Simple Monte Carlo test with 100 iteration

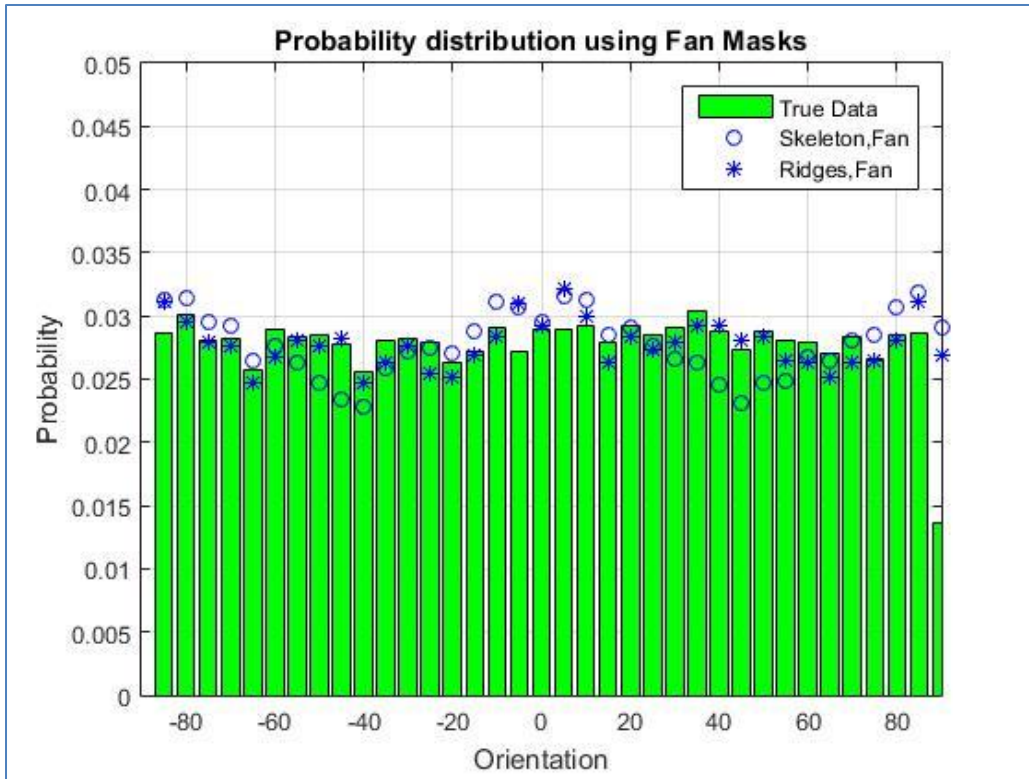
To validate this research method's accuracy, we set the interrogate window size $\omega = 800$, the angle scan resolution $\Delta\theta = 5^\circ$, would be used to test the performance with different masks. Later on, we will also do the comparison of the choice of ω and $\Delta\theta$. Table 2.2 shows the pseudo code of the method process in validation of simulation image.

Table 2.2 Procedure of finding orientation distribution in simulation nanofibers' images

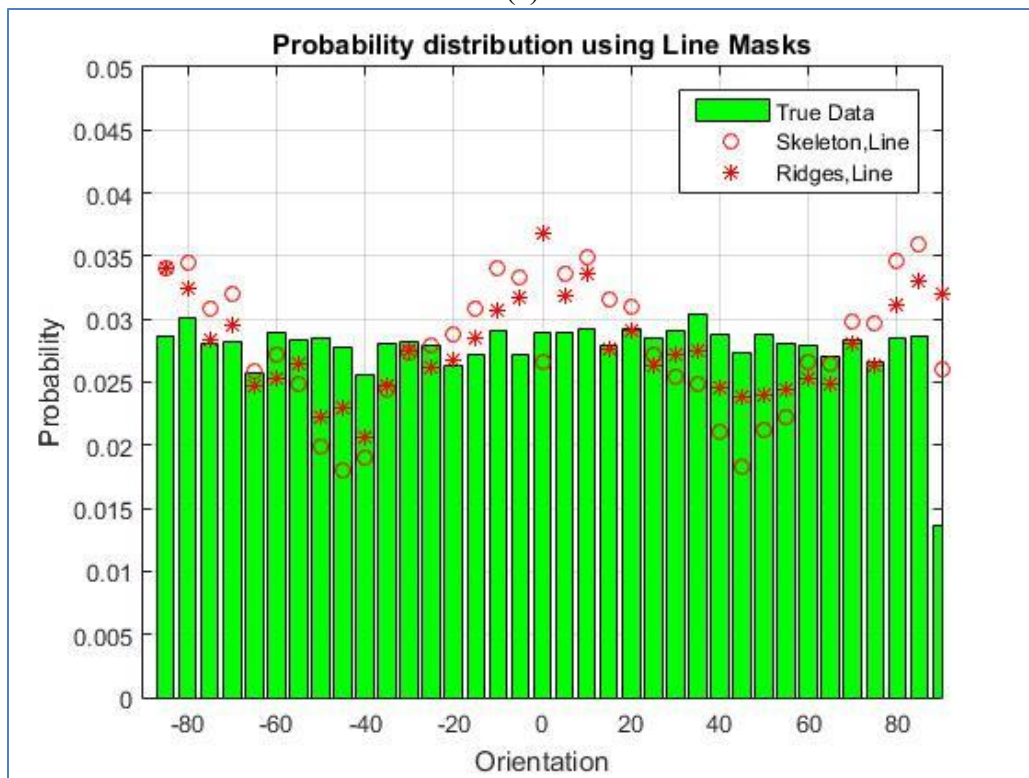
-
- For Monte Carlo simulation $i = 1:100$
 - Generate simulation image using orientation PDF $P(\theta)$,
 - Use Gaussian filter to smooth the image
 - Generate “Skeleton” and “Ridges” images
 - Divide above generated images into t square interrogate window's size ω
 - Build the Fan shape and Line shape masks based on ω and $\Delta\theta$
 - For $j = 1:t$
 - Use Fourier Transform to transfer the interrogate window image from spatial domain to frequency domain
 - Convolute the Fourier spectrum image of the interrogate window with the masks for all the different angle
 - Sum the value of each interrogate window of each angle and get the orientation distribution
-

After the Monte Carlo test, we may conclude some information by observing the distribution plots.

1) Nanofibers with $P_1(\theta)$ orientation is random angle by uniform distribution which means the nanofibers are isotropic. Figure 2.14(a) shows the $P_1(\theta)$ distribution result based on Fan shape masks. It is shown when using Fan masks, both ridges images and skeleton images are closed to the true data, in “Ridges” images, the results are a little better than in “Skeleton” images. But for the Line shape masks in Figure 2.14(b), the performance is not good in “Skeleton” images. In “Ridges” images, the results are the same, it cannot detect the right shape of the probability distribution of the true data.



(a)

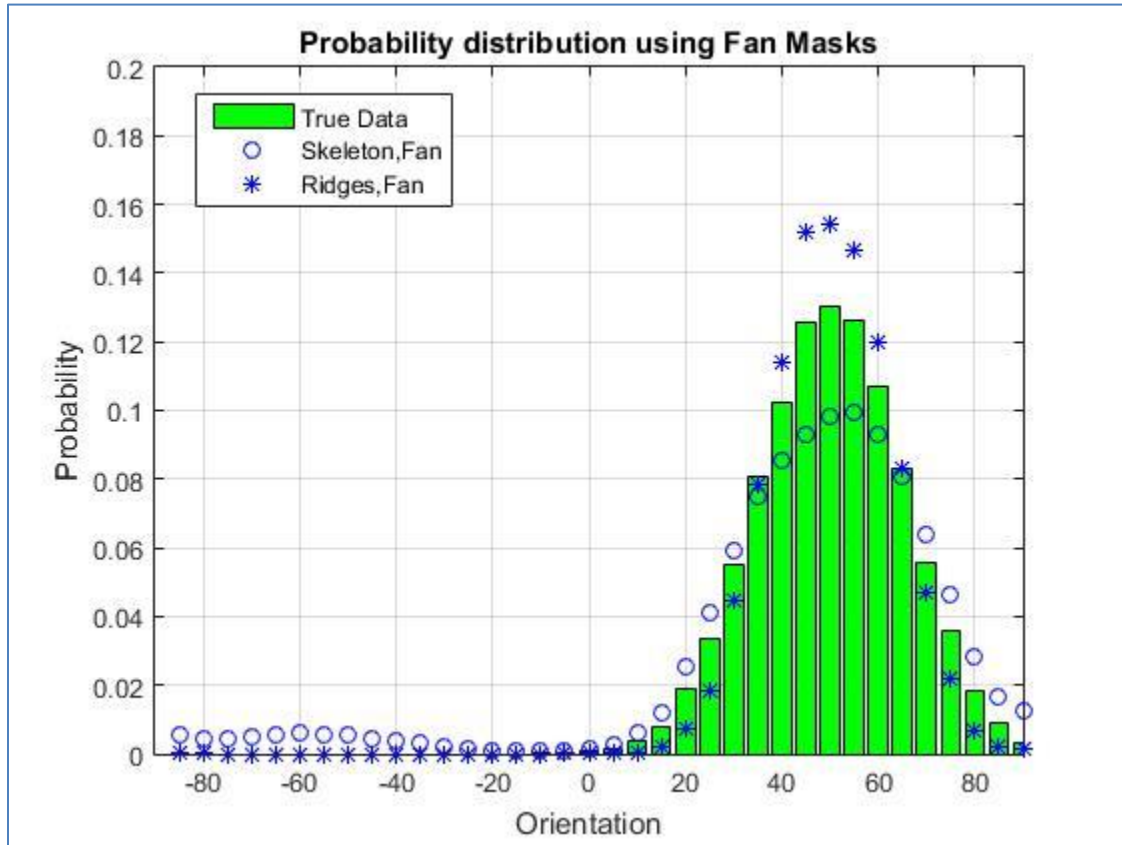


(b)

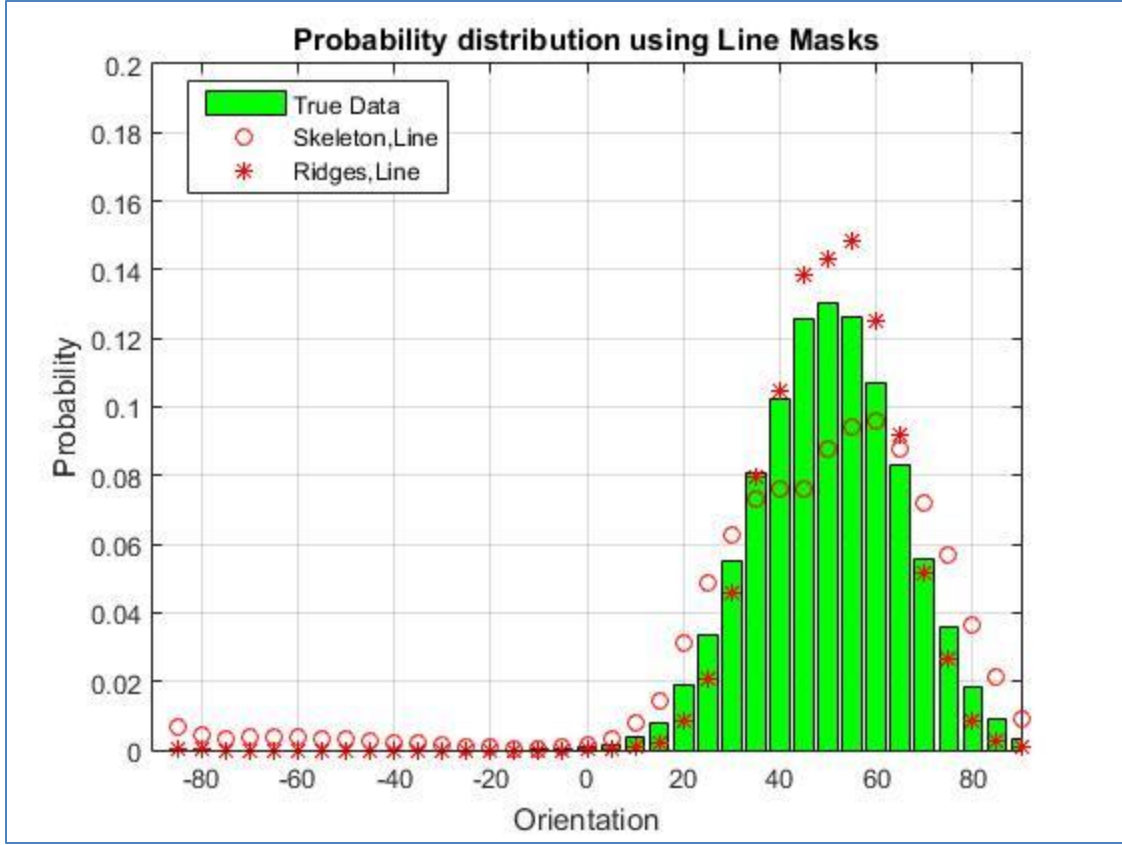
Figure 2.14 (a) Probability distribution of Fan shape method of simulation case 1

(b) Probability distribution of Line shape method of simulation case 1

2) $P_2(\theta)$ is a normal distribution with a bigger stand deviation 15. In Figure 2.15(a), we may find that Fan shape masks on the “Ridges” image will reflect more intensively than “Skeleton” image on the area close to the $\mu = 50$. Also, the dilation operations did some improvement. For the Line shape masks in case 2 in Figure 2.15(b), it has the similar result in “Ridges” images with Fan shape masks.



(a)

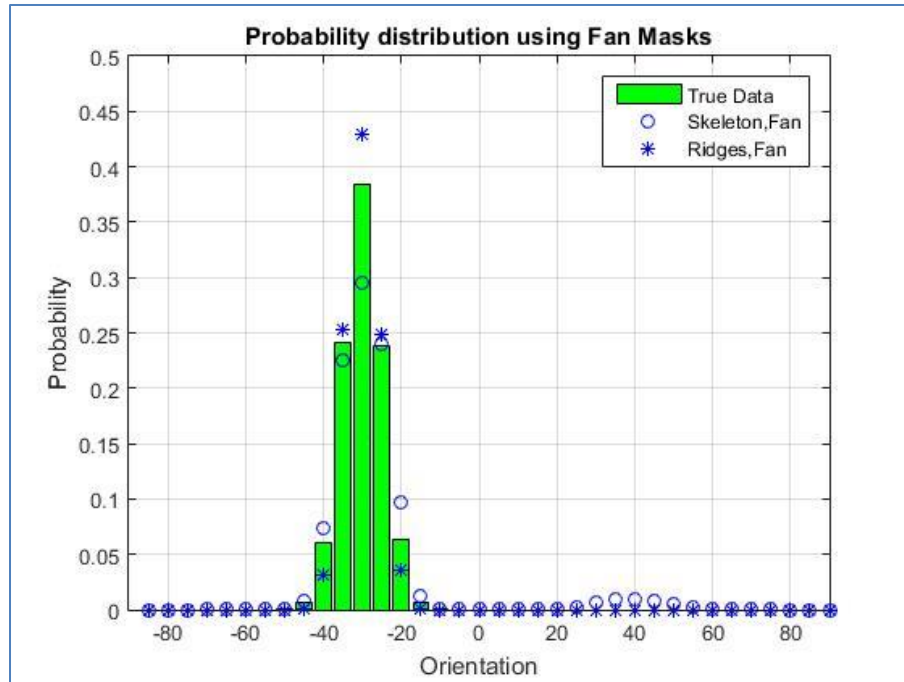


(b)

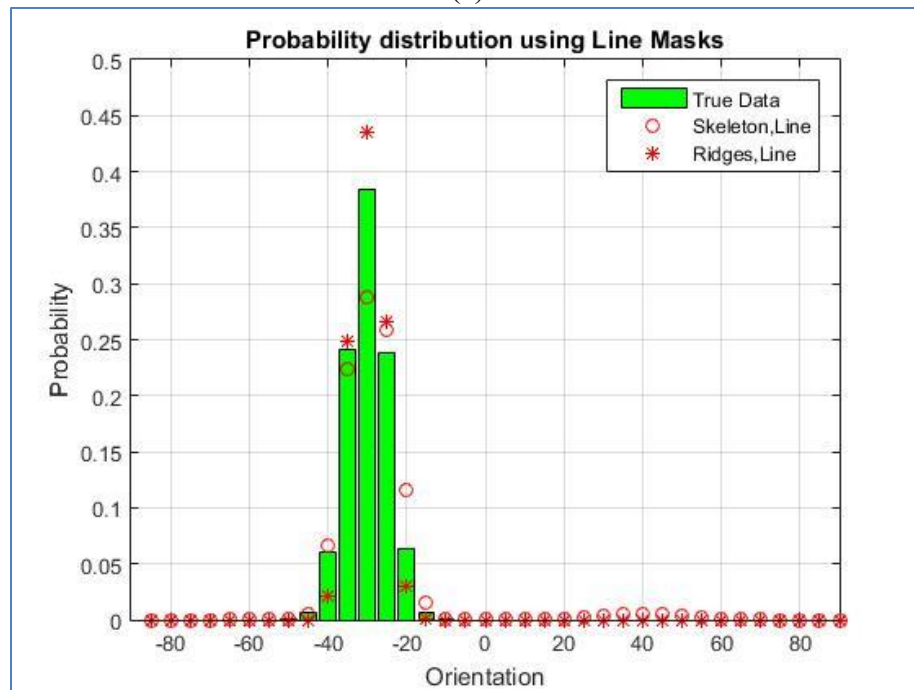
Figure 2.15 (a) Probability distribution of Fan shape method of simulation case 2

(b) Probability distribution of Line shape method of simulation case 2

3) $P_3(\theta)$ is also a normal distribution with a smaller standard deviation $\sigma = 5$, thus the simulation data is more convergent around $\mu = -30$. In Figure 2.15(a), Fan shape masks in both "Skeleton" and "Ridges" images can detect the probability distribution correctly. For the line shape masks test in Figure 2.16(b), it has the similar results,



(a)

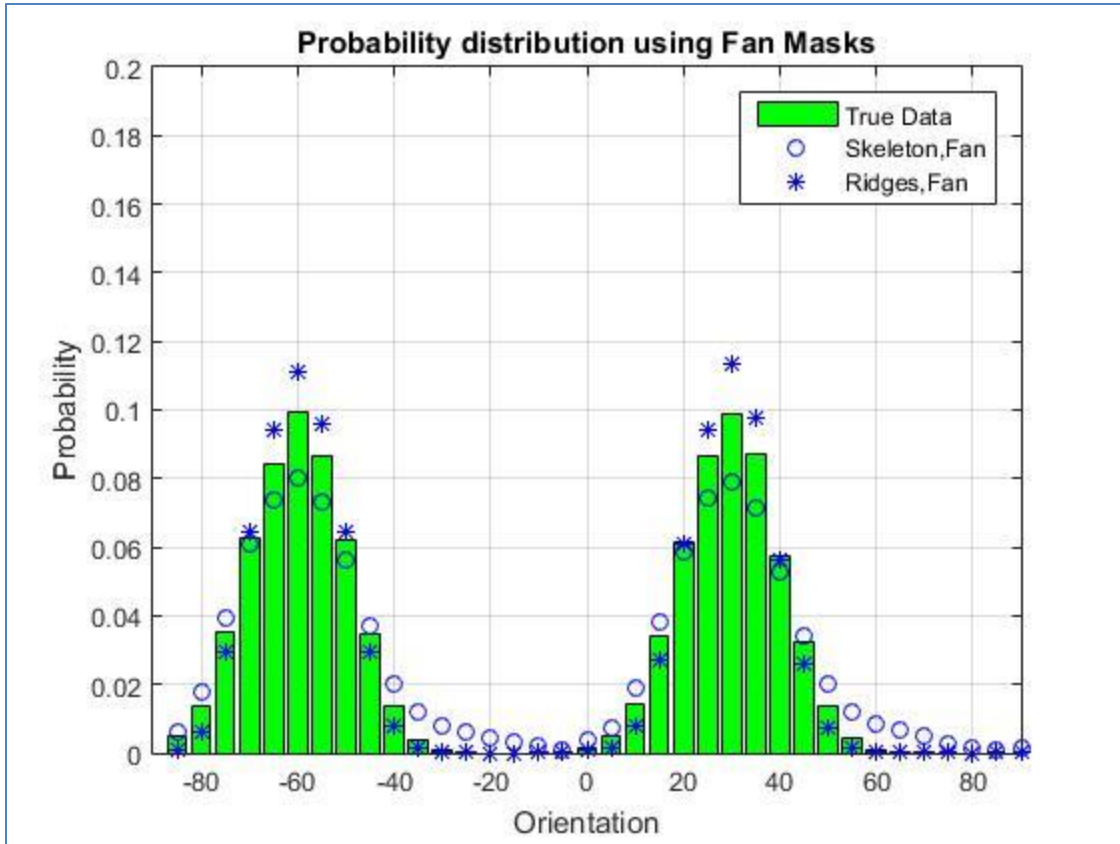


(b)

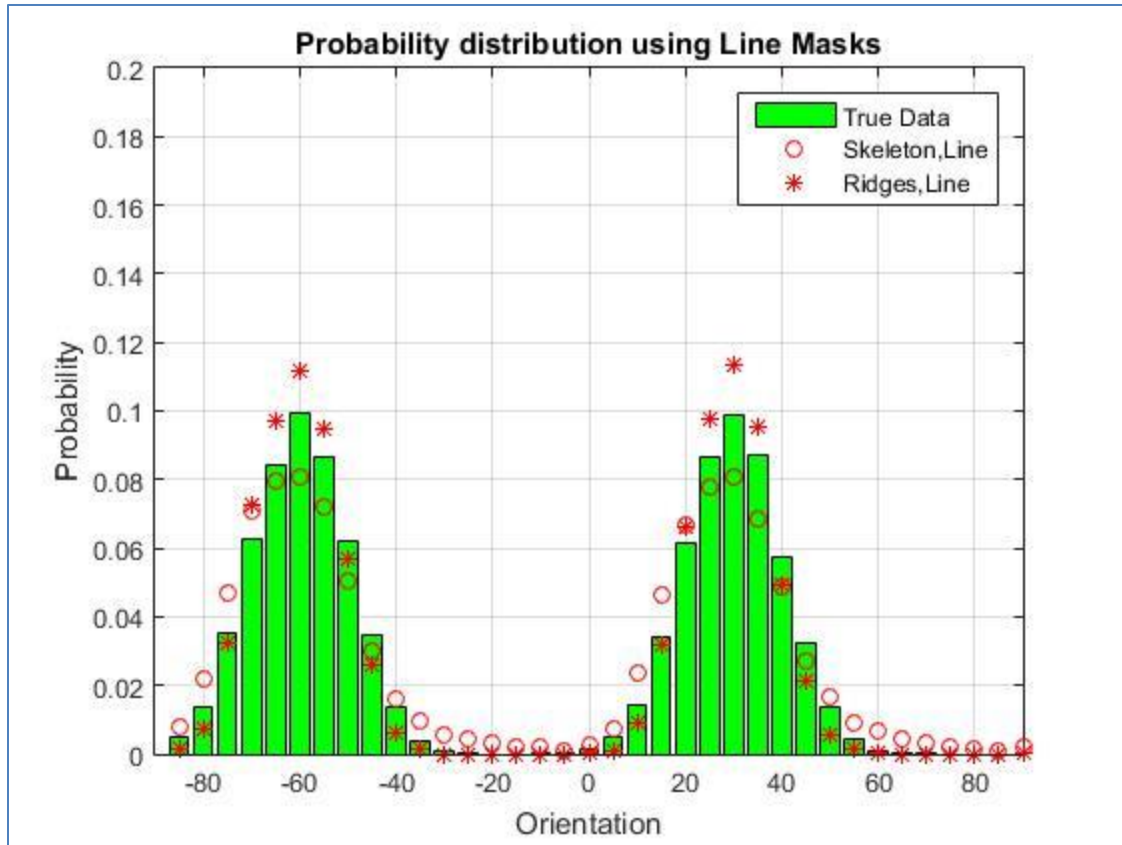
Figure 2.16 (a) Probability distribution of Fan shape method of simulation case 3

(b) Probability distribution of Line shape method of simulation case 3

4) $P_4(\theta)$ is a two normal distribution function, it has two bell shape distributions. From Figure 2.17 (a) (b) both fan shape and line shape masks provide good performance. And we may find that the Ridges method is more sensitive than the “Skeleton” method.



(a)

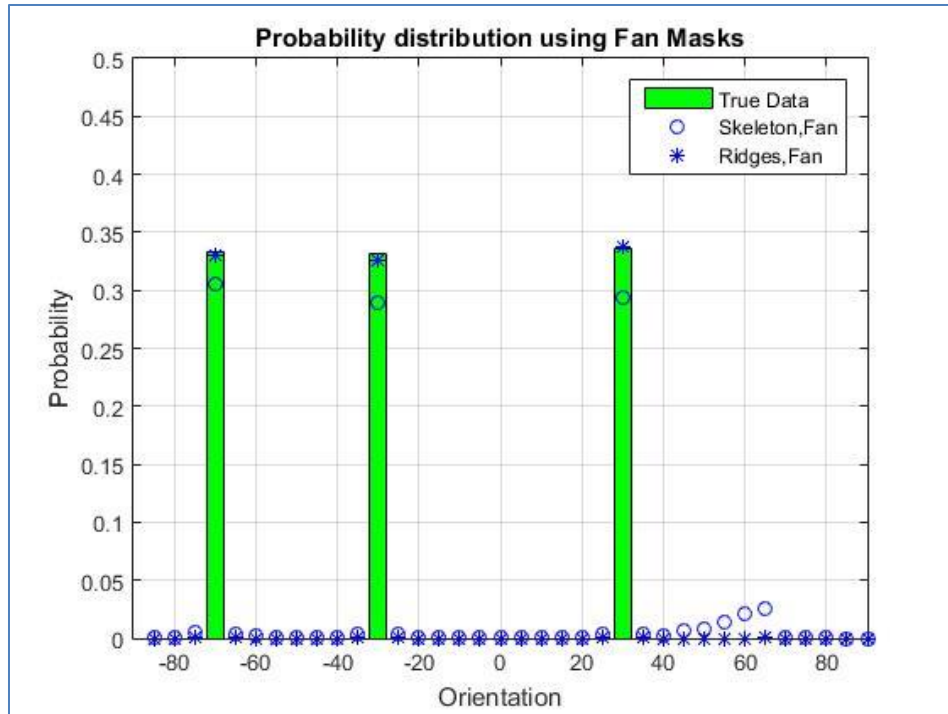


(b)

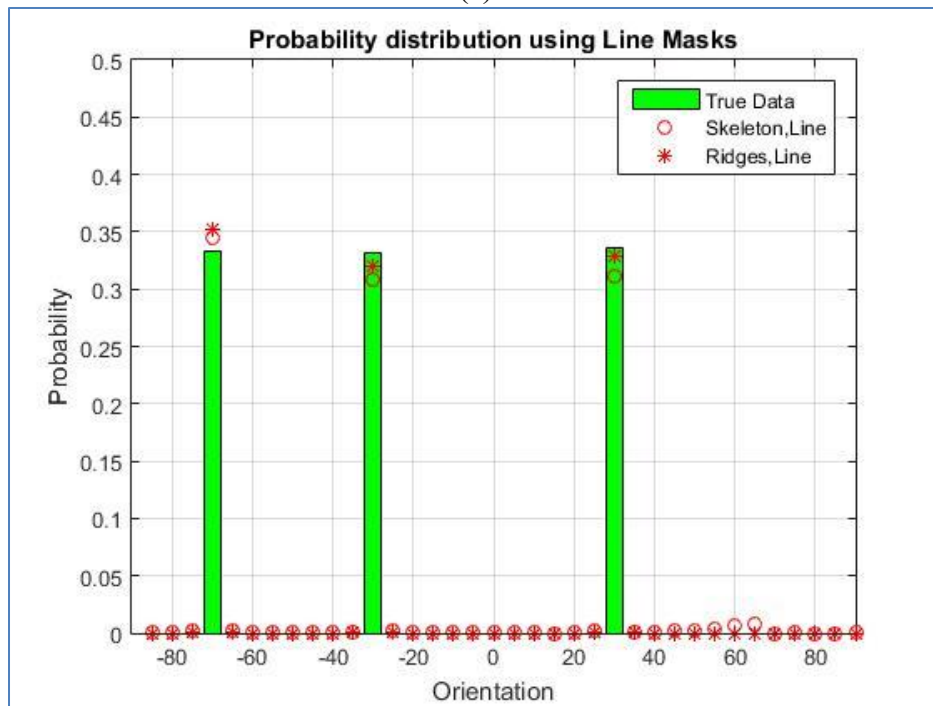
Figure 2.17 (a) Probability distribution of Fan shape method of simulation case 4

(b) Probability distribution of Line shape method of simulation case 4

5) $P_5(\theta)$ is a distribution which has three peaks with same probability $1/3$. All the method in “Ridges” images get perfect results which match the true data.



(a)

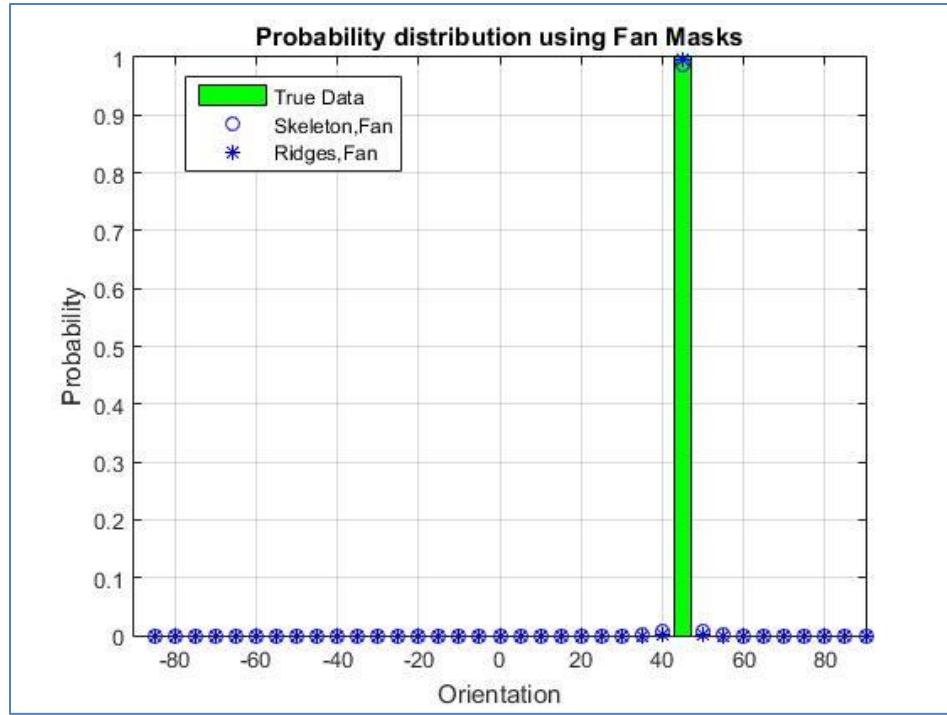


(b)

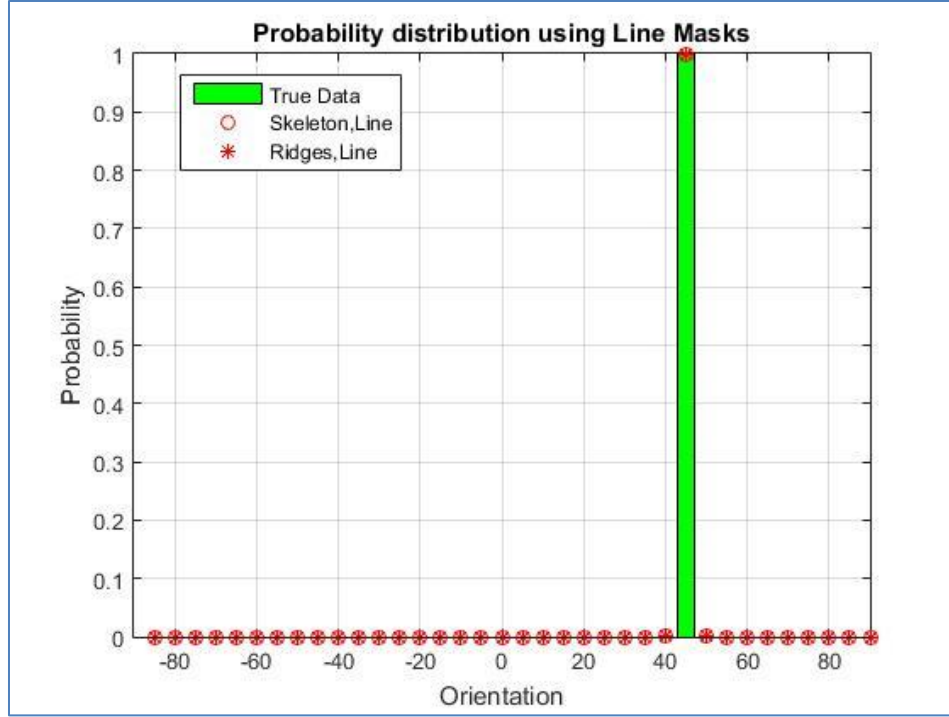
Figure 2.18 (a) Probability distribution of Fan shape method of simulation case 5

(b) Probability distribution of Line shape method of simulation case 5

6) $P_6(\theta)$ is one single jump distribution which means all the nanofibers aligned perfect well with one direction, all the methods in this research can get good results.



(a)



(b)

Figure 2.19 (a) Probability distribution of Fan shape method of simulation case 6

(b) Probability distribution of Line shape method of simulation case 6

Table 2.3 shows the Residual Sum of Squares of each method in these 6 cases. The least RSS is shown bold in each case. We may conclude that the line shape masks may have advantage when the orientation of nanofibers converge very closely while fan shape masks could provide high accuracy in all the situation.

Table 2.3 RSS of simulation validation

	case 1	case 2	case 3	case 4	case 5	case 6
RSS SK Fan	0.00044372	0.0042	0.01	0.002	0.0057	0.00025488
RSS Ridges Fan	0.00025154	0.003	0.0039	0.0011	0.000046774	0.000043977
RSS SK Line	0.001	0.0081	0.0131	0.0023	0.0014	8.8654E-06
RSS Ridges Line	0.00072903	0.0019	0.0061	0.0015	0.00059149	4.1624E-08

2.3.3 Testing for Robustness

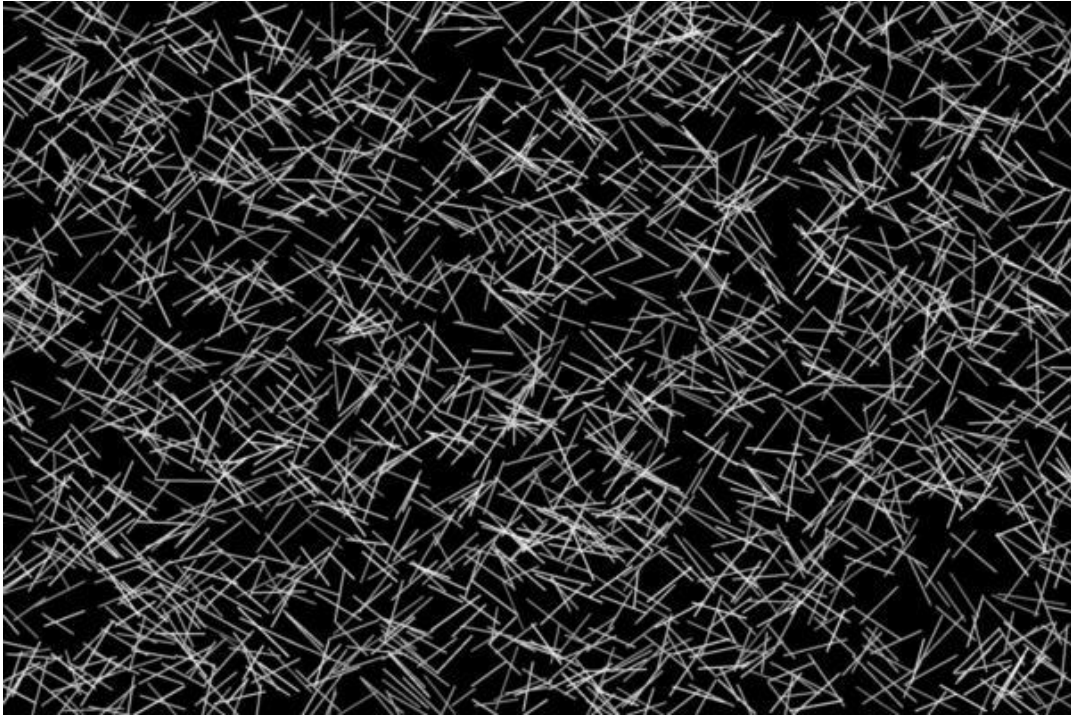
Previous section test the proposed method's performance in normal situation. Here, we may increase the number of nanofibers in simulation image to test whether these methods are robustic.

Set the number of nanofibers to 2000, which is 10 times as the previous condition. Also we use two new $P_7(\theta)$ and $P_8(\theta)$ orientation distribution function to generate the simulation image which would be validate by the above methods. Table 2.4 illustrate the new functions.

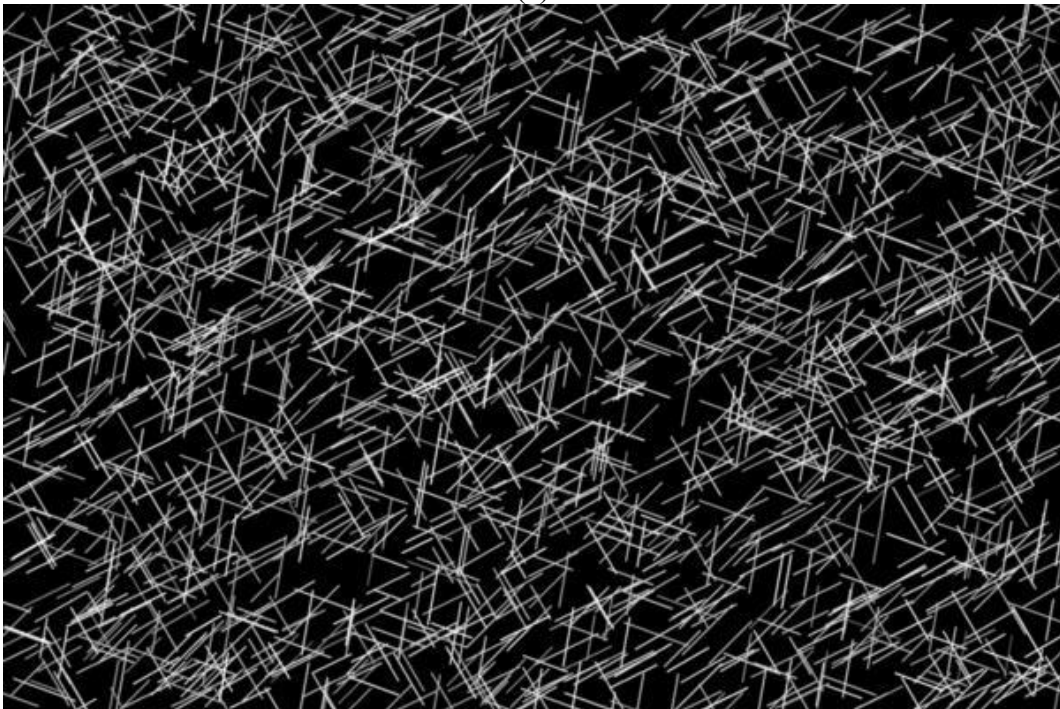
Table 2.4 Two new cases for Robustness Validation

Case Number	Case	Probability Density Function
7	2 normal distribution with $\mu_1 = 45$ and $\sigma_1 = 20$, $\mu_2 = -55$ and $\sigma_2 = 15$	$P_7(\theta) = \frac{1}{\sqrt{2\pi} \cdot 20^2} \exp \left[-\frac{(\theta - 40)^2}{2 \cdot 20^2} \right] + \frac{1}{\sqrt{2\pi} \cdot 15^2} \exp \left[-\frac{(\theta - (-55))^2}{2 \cdot 15^2} \right]$
8	6 peaks	$P_8(\theta) = \begin{cases} \frac{1}{6}, -70 \\ \frac{1}{6}, -30 \\ \frac{1}{6}, 10 \\ \frac{1}{6}, 45 \\ \frac{1}{6}, 60 \\ \frac{1}{6}, 64 \end{cases}$

Figure 2.20 (a) (b) shown the examples of the complex nanofibers simulation image with $P_7(\theta)$ and $P_8(\theta)$. Since $P_7(\theta)$ has two normal distribution with larger stand deviation and $P_8(\theta)$ has more peaks, it is much more complex and difficult to get the orientation of the nanofibers by human being's vision.



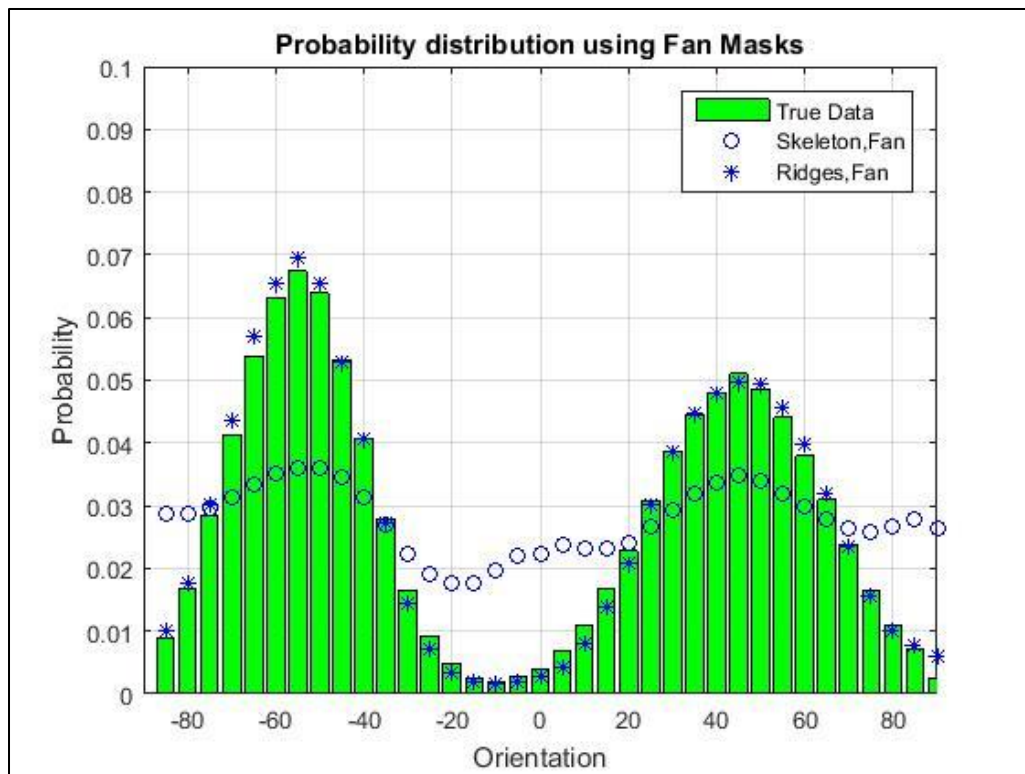
(a)



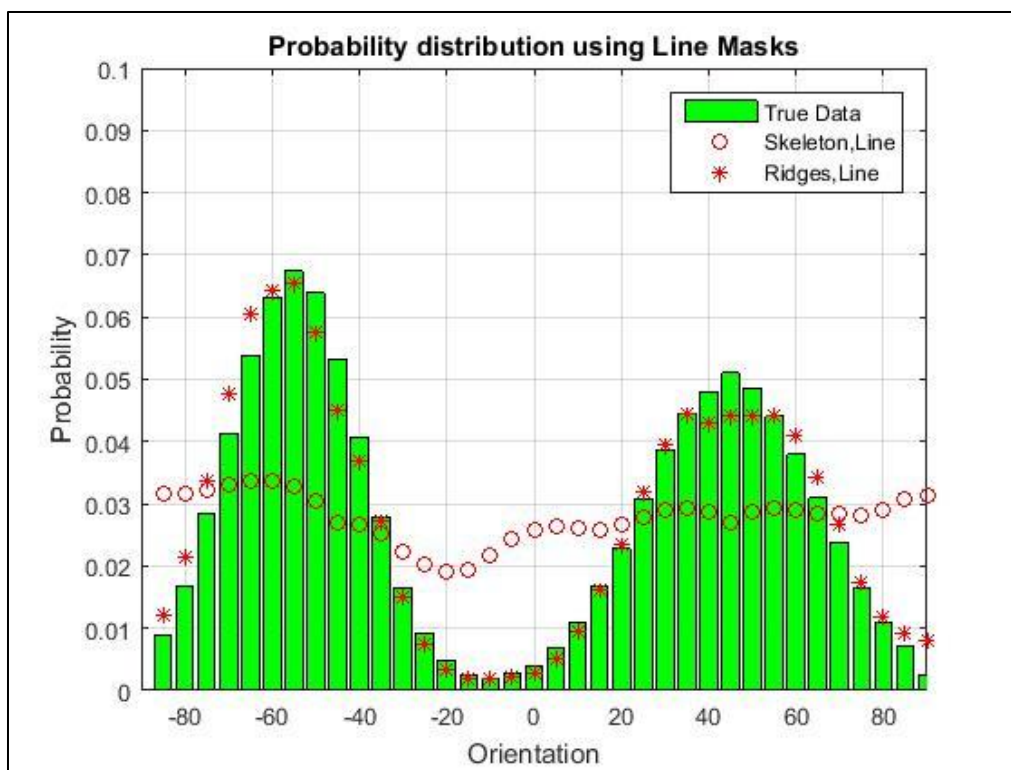
(b)

Figure 2.20(a)(b)Example Simulation image with 2000 nanofibers

7) From the distribution results in Figure 2.21(a) (b), we may find for case 7 which contain two normal distribution with bigger stand deviation, the fan shape masks method in “Ridges” image represent the best fit with the true generated nanofibers’ distribution. The orientation’s accuracy of line shape masks method in “Ridges” image is a bit lower. On the other hand, for “Skeleton” images, since it cannot represent the original images due to the simulated nanofibers are too dense to discriminate, thus the issue in Figure 2.22 would made the result not good. The RSS of the methods in case 7 is shown in Table 2.5



(a)



(b)

Figure 2.21 (a) Probability distribution of Fan shape method of simulation case 7

(b) Probability distribution of Line shape method of simulation case 7

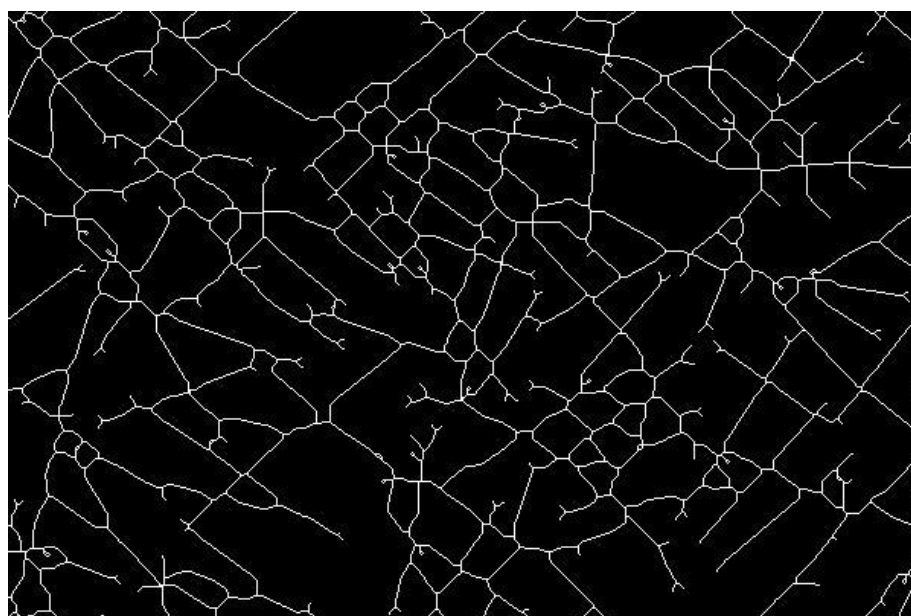
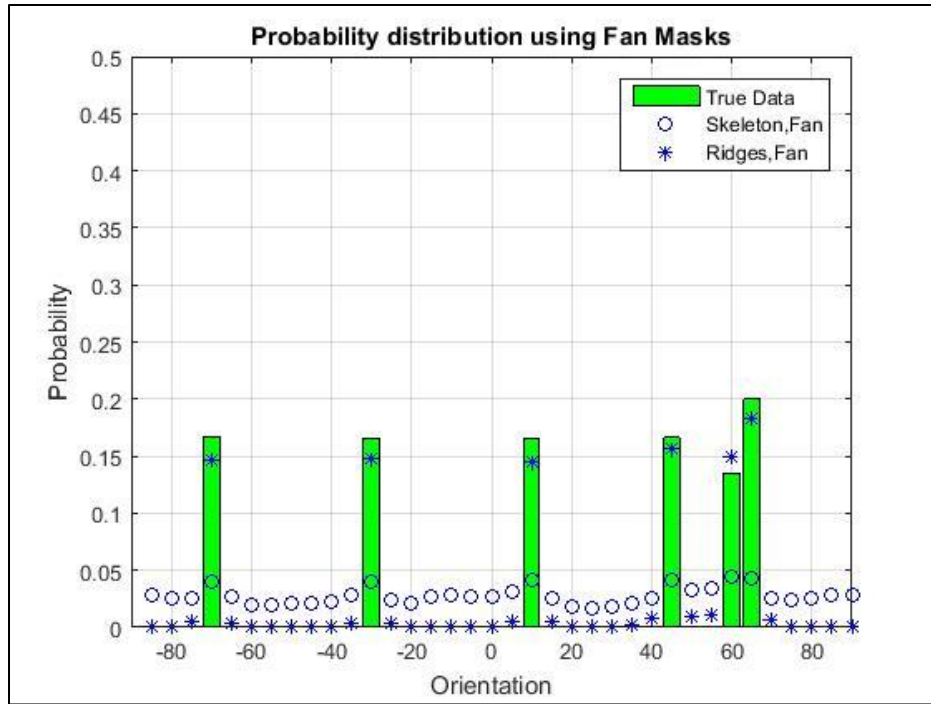


Figure 2.22 2000 nanofibers' simulation image after "Skeleton" operation

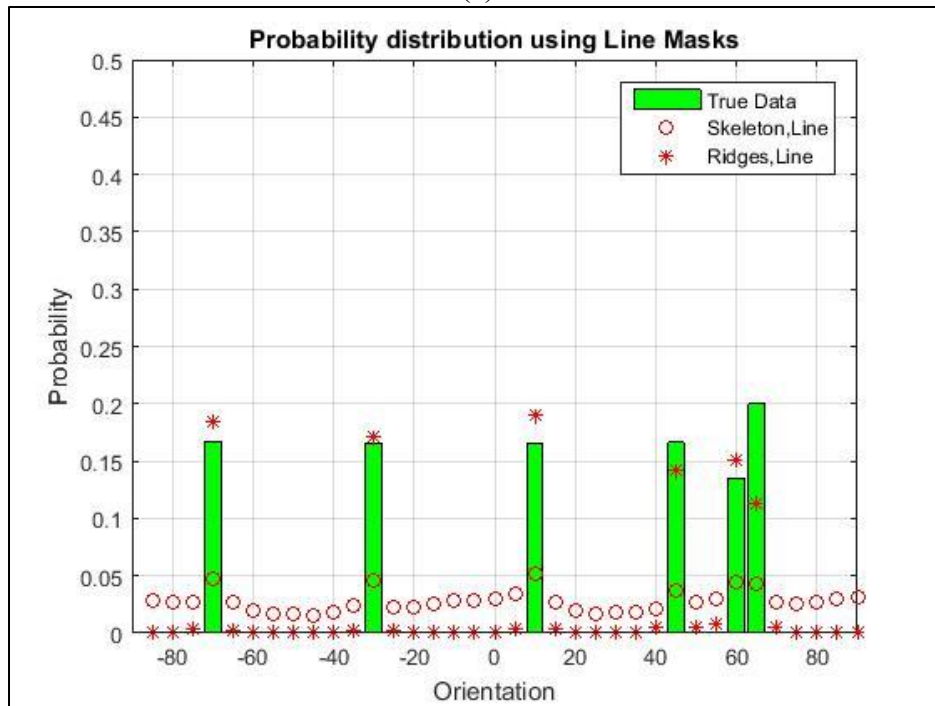
Table 2.5 RSS of methods in case 7

	case 7
RSS SK Fan	0.0086
RSS Ridges Fan	9.9362E-05
RSS SK Line	0.0119
RSS Ridges Line	0.00045136

8) For case 8, there are 6 peaks in $P_8(\theta)$, there are two peaks very close to each other, one is 60° and another is 64° , the interval between these two is only 4° , but the $\Delta\theta$ in this test is 5° , therefore, it could be difficult to distinguish these two peaks with the exactly right angle. In Figure 2.23(a)(b), for “Skeleton” images, the issue discussed previous also effect in this case due to high density of nanofibers. For “Ridges” images, two shape masks can simulate the distribution well, again like the previous test, fan shape is a little ahead of line shape, the difference could be find from RSS table in Table 2.6. If we change the $\Delta\theta$ to 2° , the result is shown in Figure 2.24(a)(b), for smaller $\Delta\theta$, the line shape masks may have the advantage compared with the fan shape masks, also, the two peaks of 60° and 64° could be found correctly, but there’s a tradeoff. The RSS when using $\Delta\theta = 2^\circ$ is higher than using $\Delta\theta = 5^\circ$ ones by the smaller tolerance.



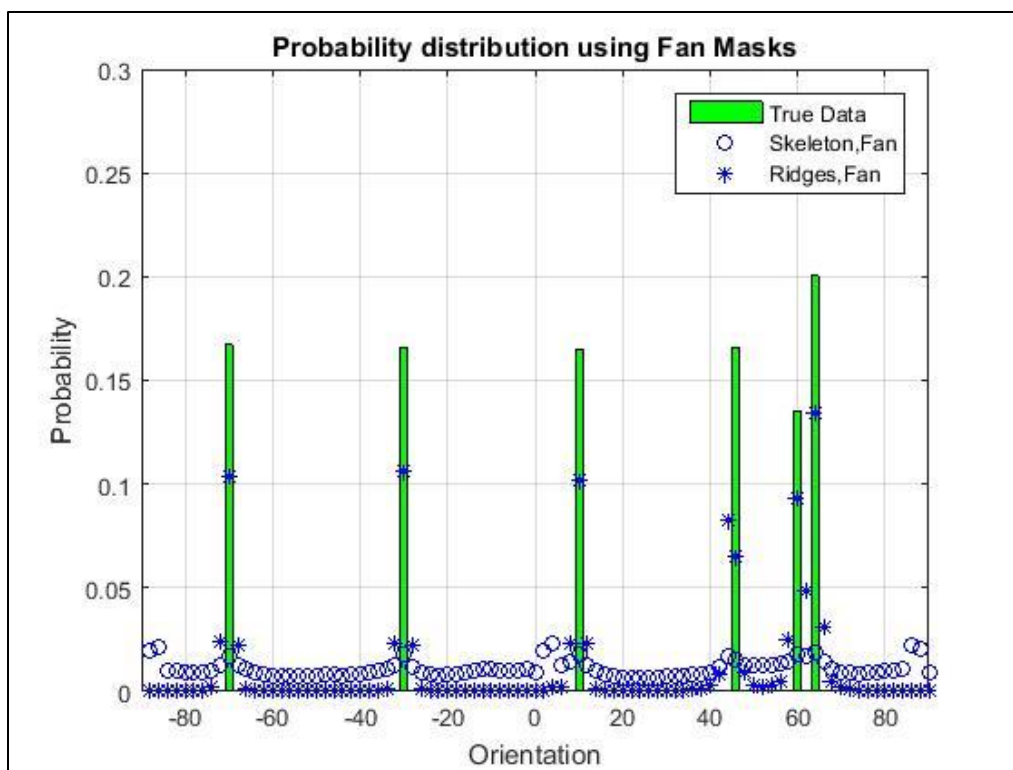
(a)



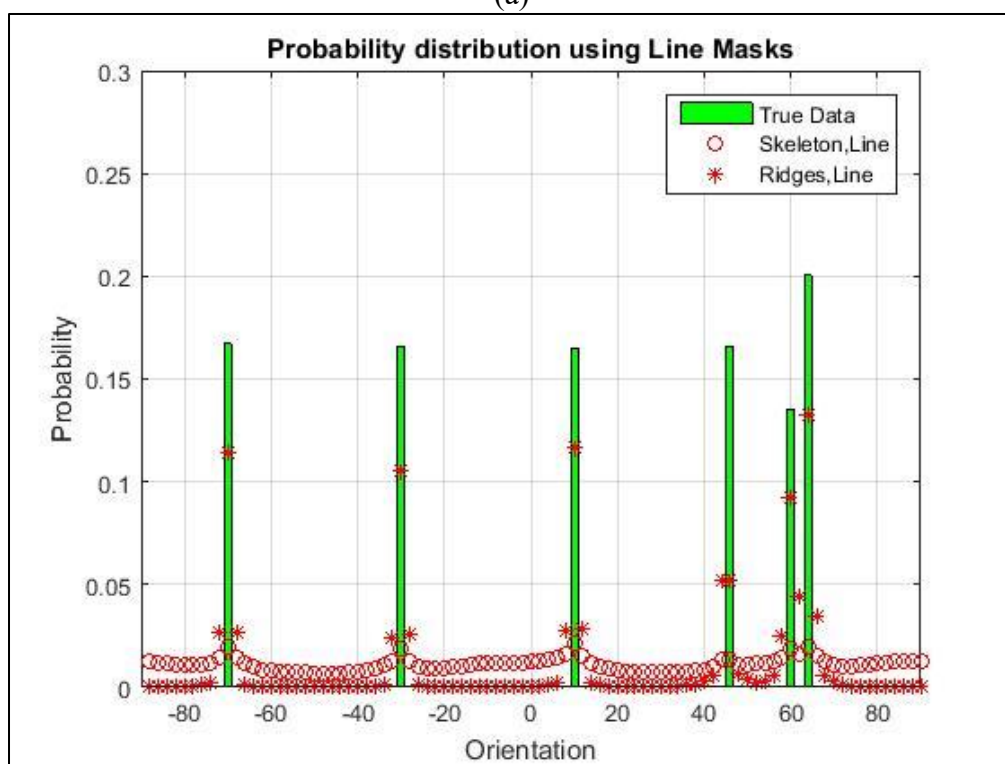
(b)

Figure 2.23 (a) Probability distribution of Fan shape method of simulation case 8 using $\Delta\theta = 5^\circ$

(b) Probability distribution of Line shape method of simulation case 8 using $\Delta\theta = 5^\circ$



(a)



(b)

Figure 2.24 (a) Probability distribution of Fan shape method of simulation case 8 using $\Delta\theta = 2^\circ$

(b) Probability distribution of Line shape method of simulation case 8 using $\Delta\theta = 2^\circ$

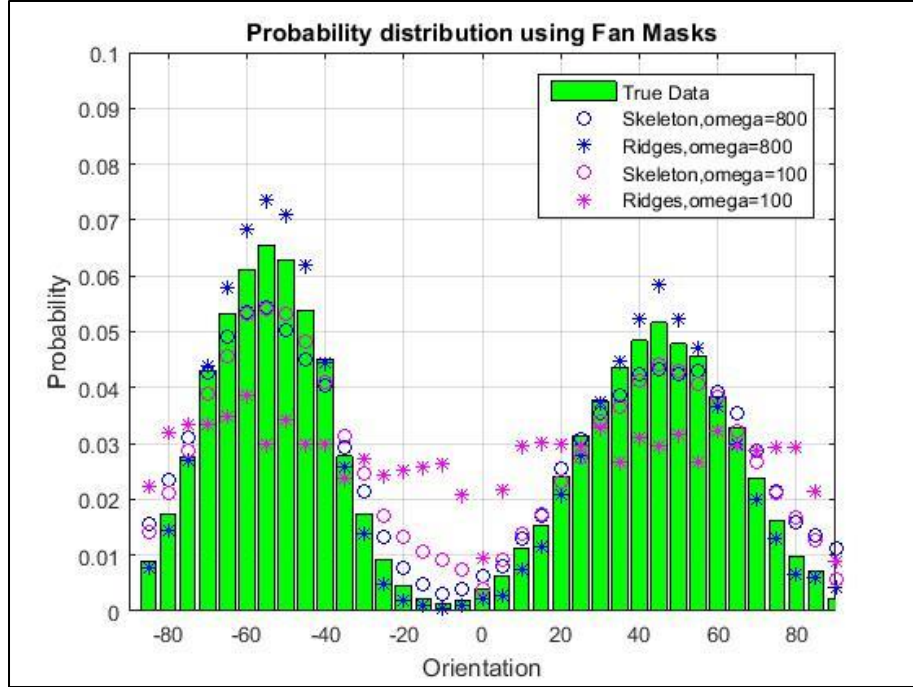
Table 2.6 RSS in Case 8

	case 8 $\Delta\theta = 5^\circ$	case 8 $\Delta\theta = 2^\circ$
RSS SK Fan	0.1154	0.146
RSS Ridges Fan	0.0021	0.042
RSS SK Line	0.11	0.1437
RSS Ridges Line	0.0096	0.0391

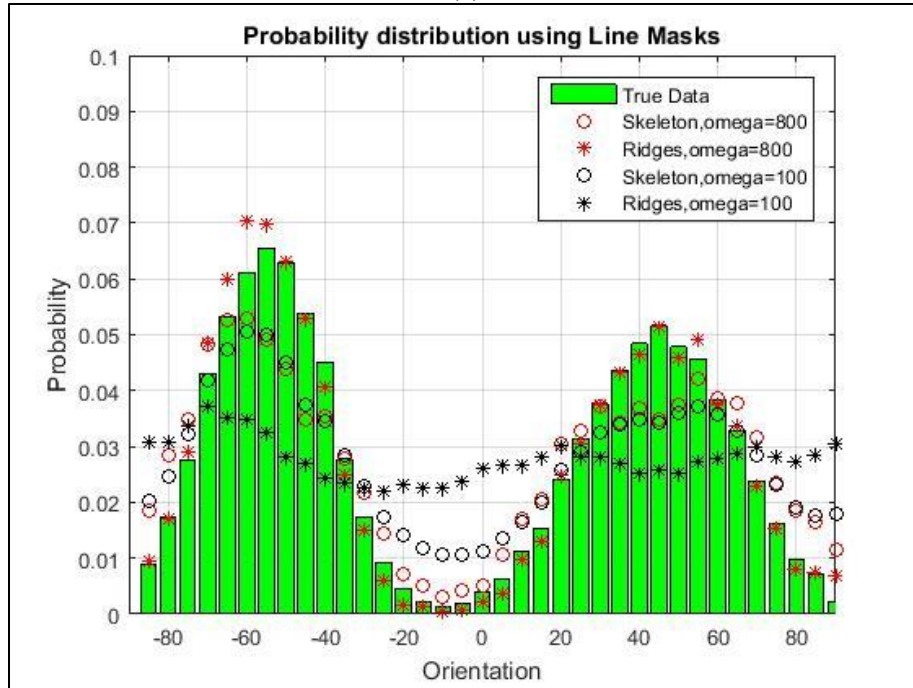
From case 7 and case 8 simulation test, we may conclude that the fan/line shape masks method in “Ridges” images is robustic. These two shapes can be implemented together to have a wiser analysis of the orientation alignment quality control.

2.3.4 Testing with different Omega size

In this case, we use two different size of interrogate window size $\omega = 800$ and $\omega = 100$ to find the difference impacts. Also, we will use the same simulation orientation distribution (case 7) with different density(200 and 2000) to compare the accuracy. Figure 2.25 (a) (b) show the results when using 200 nanofibers’ simulation images and 2.25 (a)(b) using 2000 nanofibers. Table 2.7 and 2.8 shows the RSS in this test.



(a)



(b)

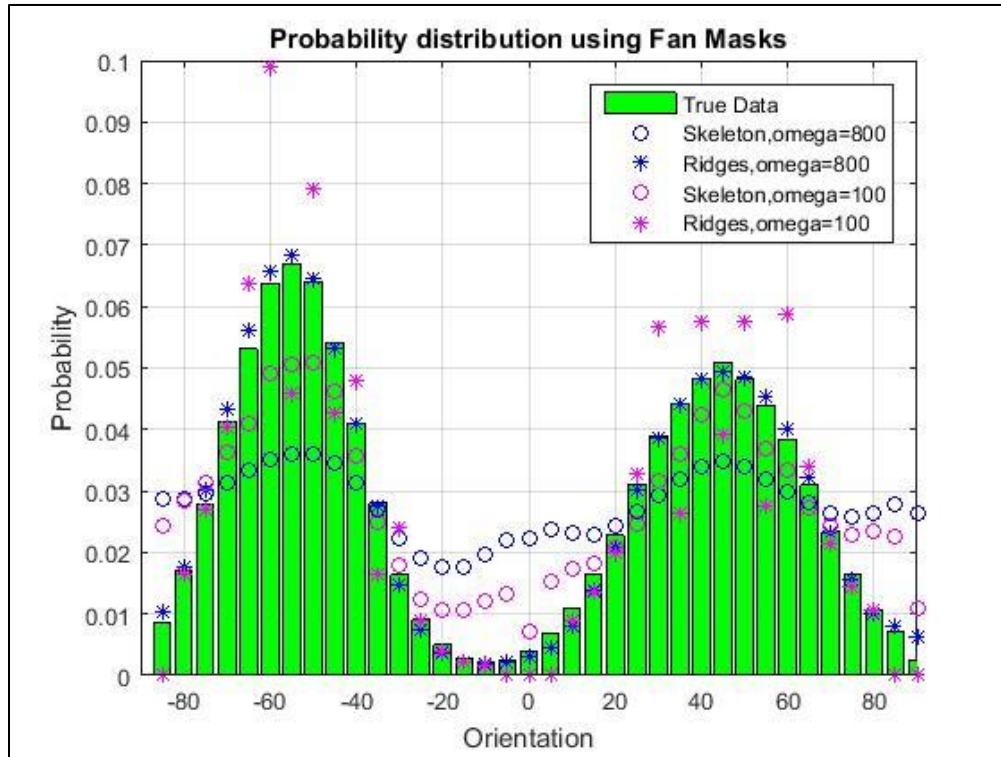
Figure 2.25 (a) Probability distribution of Fan shape method of simulation case 7 using $\omega = 800$ and $\omega = 100$ with 200 nanofibers

(b)) Probability distribution of Line shape method of simulation case 7 using $\omega = 800$ and $\omega = 100$ with 200 nanofibers

Table 2.7 Comparison between different interrogate windows size with 200 nanofibers simulation

	RSS in 200 nanofibers
RSS SK Fan Omega=800	0.001
RSS SK Fan Omega=100	0.0012
RSS Ridges Fan Omega=800	0.00053369
RSS Ridges Fan Omega=100	0.01
RSS SK Line Omega=800	0.0027
RSS SK Line Omega=100	0.0032
RSS Ridges Line Omega=800	0.00031053
RSS Ridges Line Omega=100	0.0128

Through Figure 2.25, we may find that in 200 nanofibers simulation test, the big interrogate window size $\omega = 800$ could provide more accurate track of true data distribution especially in Ridges images. The RSS prove this assumption, we may find for both Fan and Line shape masks methods, the big $\omega = 800$ has much less RSS.



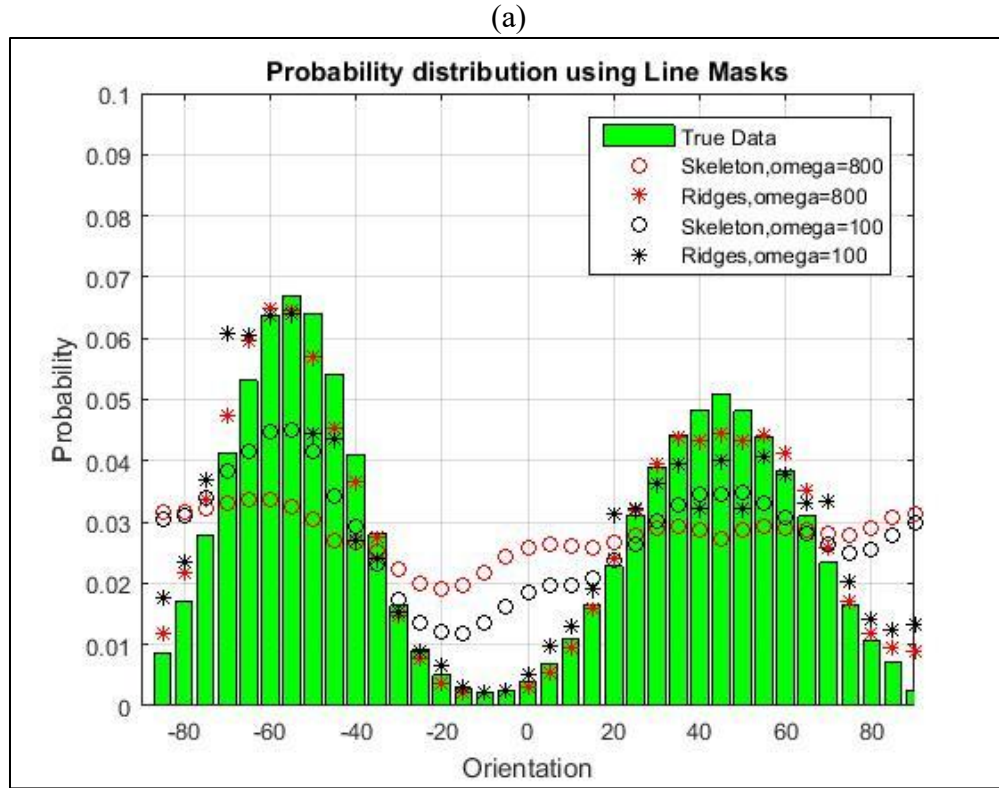


Figure 2.26 (a) Probability distribution of Fan shape method of simulation case 7 using $\omega = 800$ and $\omega = 100$ with 2000 nanofibers

(b)) Probability distribution of Line shape method of simulation case 7 using $\omega = 800$ and $\omega = 100$

With 2000 nanofibers

Table 2.8 Comparison between different interrogate windows size with 2000 nanofibers simulation

	RSS in 2000 nanofibers
RSS SK Fan Omega=800	0.0086
RSS SK Fan Omega=100	0.0026
RSS Ridges Fan Omega=800	0.000087168
RSS Ridges Fan Omega=100	0.0043
RSS SK Line Omega=800	0.0119
RSS SK Line Omega=100	0.0062
RSS Ridges Line Omega=800	0.00047454
RSS Ridges Line Omega=100	0.0024

When the nanofibers increase 10 times, the results is similar when in “Ridges” images. Though for “Skeleton” images, the smaller interrogate windows size $\omega = 100$ could provide better performance, it cannot competitive with the larger interrogate windows size $\omega = 800$ in “Ridges” images.

Through all the validation tests, we may conclude that this pattern mining method could detect the orientation of nanofibers in simulation image effectively, and large interrogate window size could give more accurate results. This method could also handle some condition like extreme high density with 2000 nanofibers, the “Ridges” method is more robust than the “Skeleton” method.

2.4 EXPERIMENTS IN REAL SEM IMAGES

The application of the method have been tested by previous simulation experiments. Here we use the method to detect the alignment in three real images.

1) Figure 2.27 shows the first SEM image, which is a 3D sight SEM image. After some pre image processing, our method has been implemented. Figure 2.28 (a) (b) is the results using fan and line shape masks. It is clear that the detected orientation is around $90^\circ - 95^\circ$.

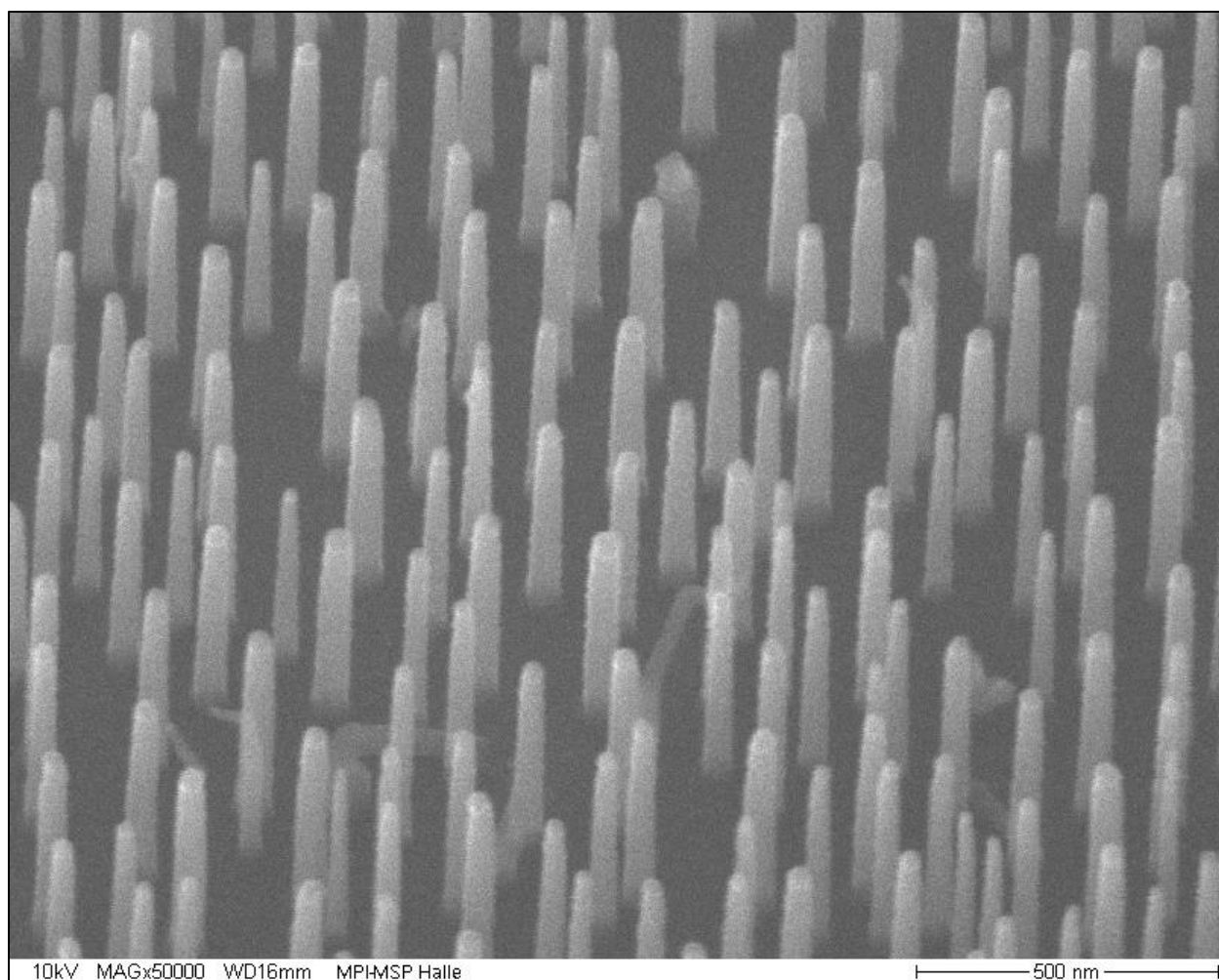
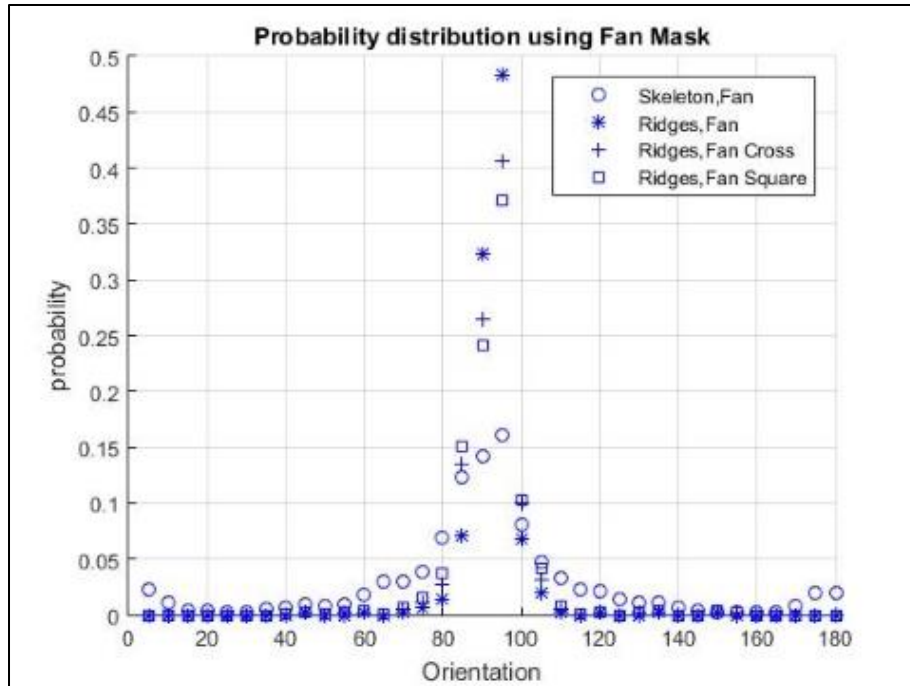
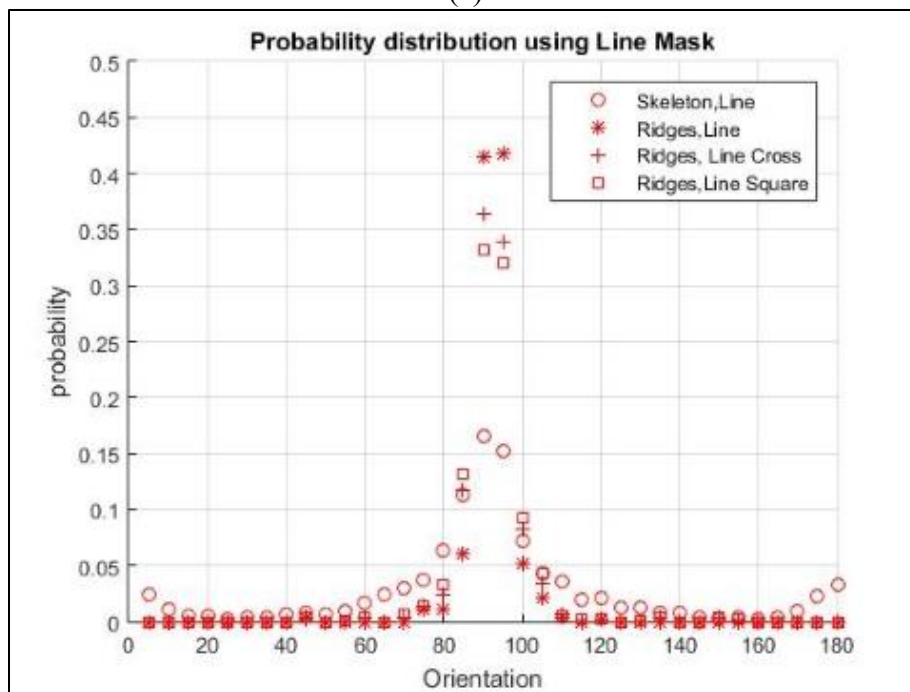


Figure 2.27 SEM image 1



(a)



(b)

Figure 2.28 (a) Probability distribution of Fan shape method of SEM image 1

(b) Probability distribution of Line shape method of SEM image 1

2) Figure 2.29 shows the second SEM image, the orientation is not clear to identify. Still after some pre- image processing steps, we deploy the methods and get the result as shown in Figure 2.30 (a) (b), $55^\circ, 130^\circ, 180^\circ$ may detected as main directions from the results.

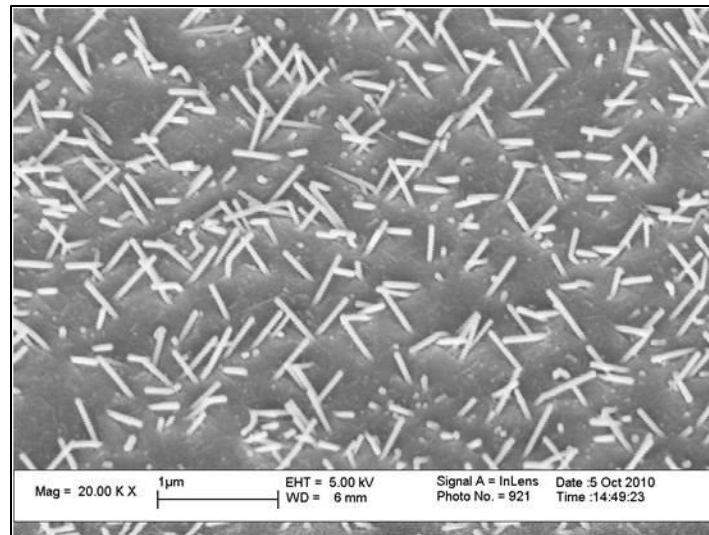
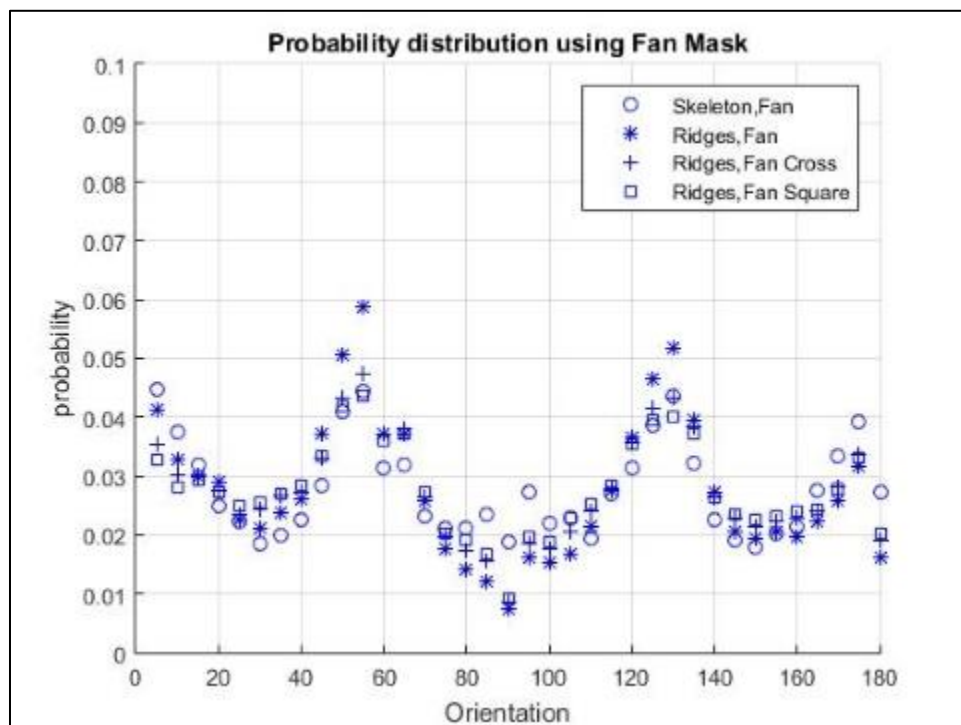
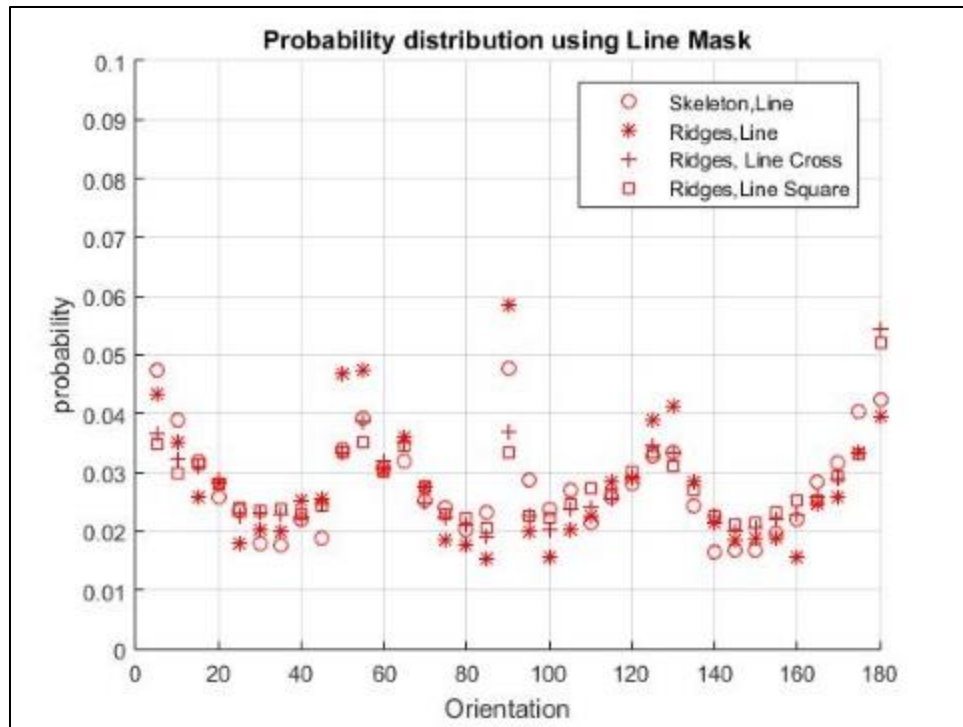


Figure 2.29 SEM image 2



(a)

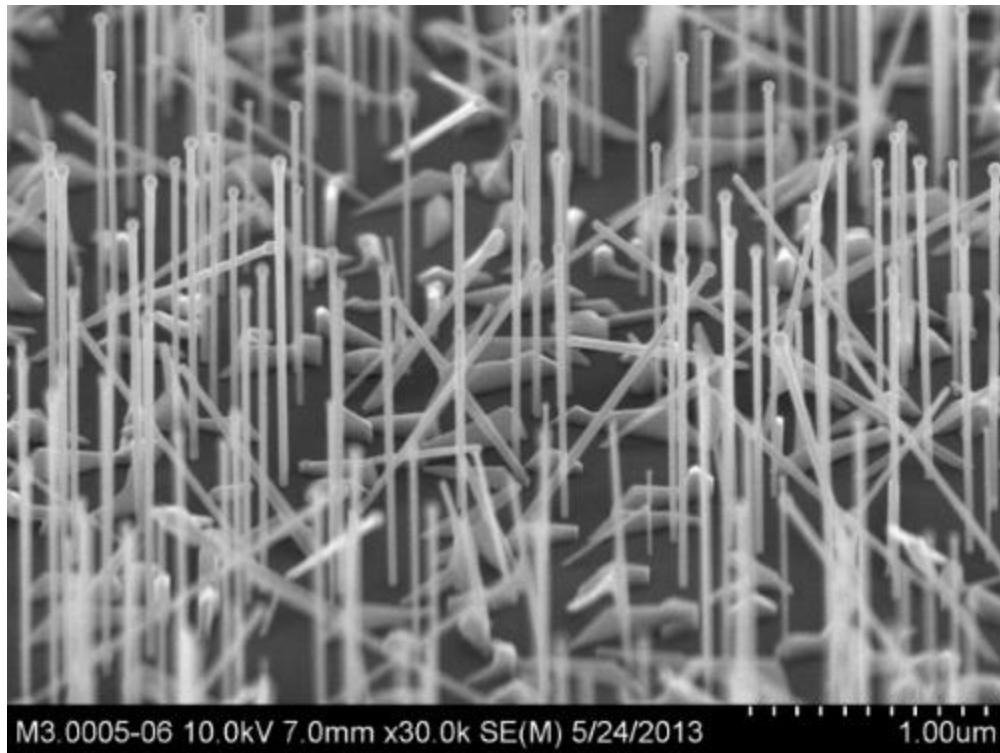


(b)

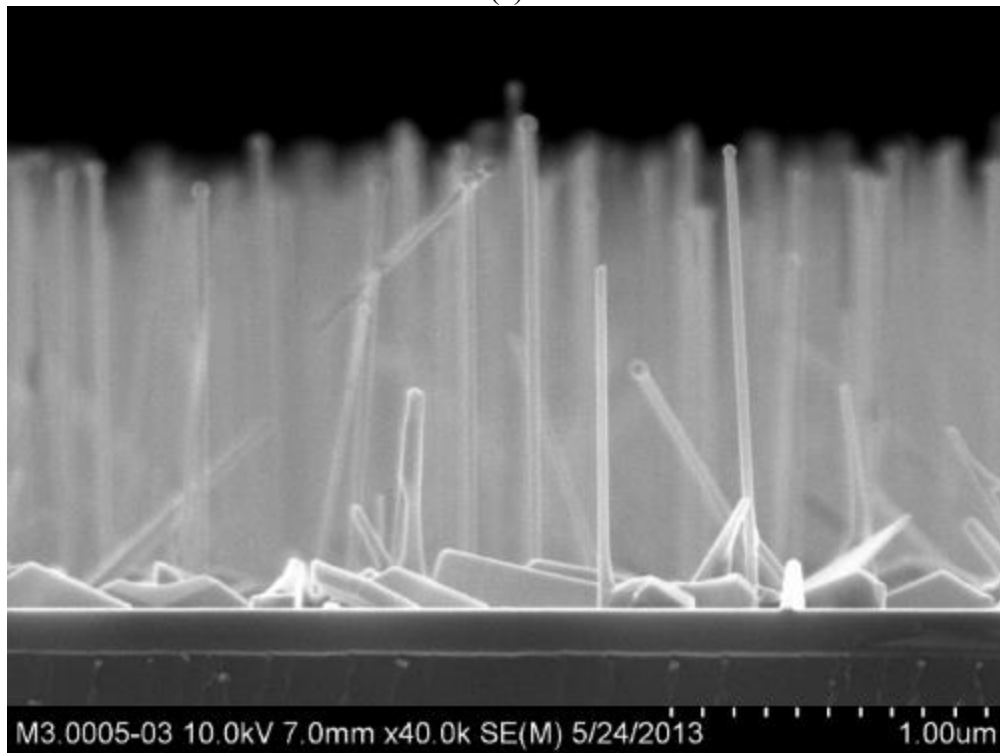
Figure 2.30 (a) Probability distribution of Fan shape method of SEM image 2

(b) Probability distribution of Line shape method of SEM image 2

3) Figure 2.31(a) shows the third SEM image, it is a 3D image with shallow focus (small depth of field), 2.31(b) shows the front view of the 3D SEM image. From the testing results shown in Figure 2.32 (a) (b), the orientation which is 90° match the true SEM image 3.

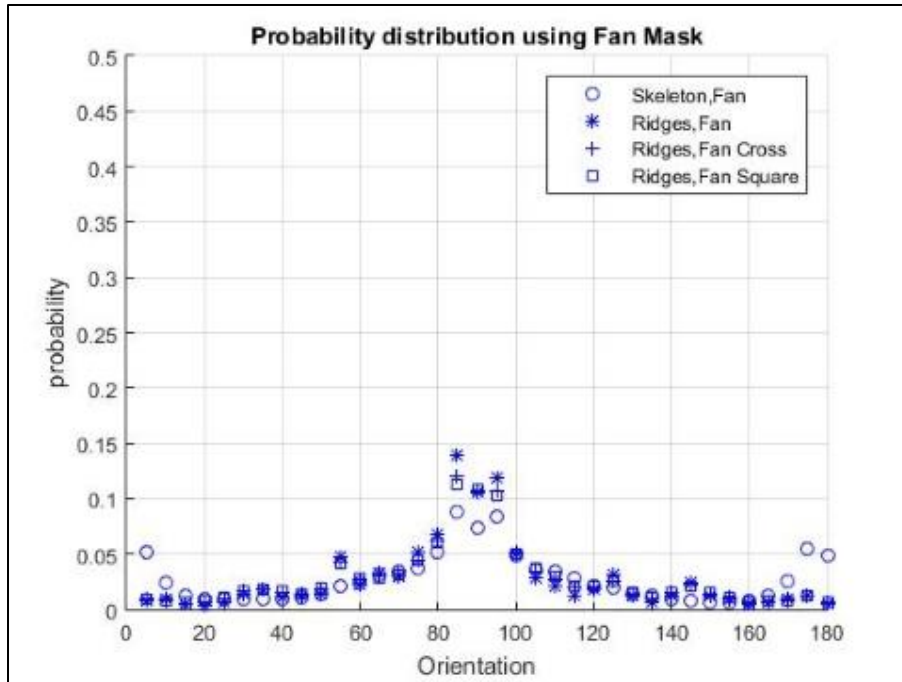


(a)

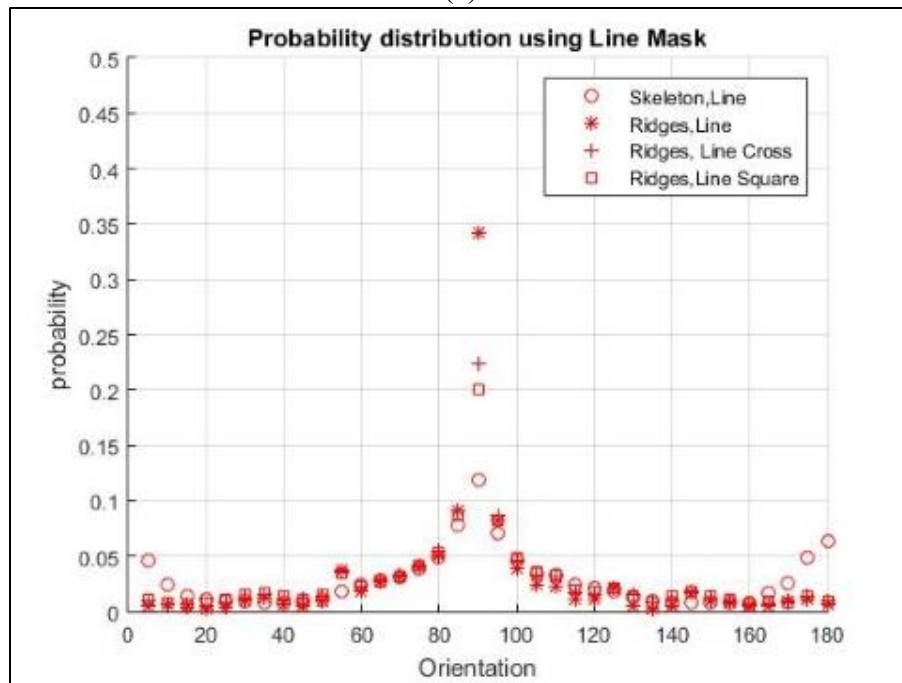


(b)

Figure 2.31 (a) (b) SEM image 3



(a)



(b)

Figure 2.32 (a) Probability distribution of Fan shape method of SEM image 3

(b) Probability distribution of Line shape method of SEM image 3

Chapter 3 Image Segmentation Based on Hough Transform

The image processing requires us to take huge amounts of low level pixel data and extract useful information. The goal of segmentation is to find regions that represent objects or meaningful parts of objects [13].

Segmentation subdivides an image into its constituent regions or objects. The level of detail to which the subdivision is carried depends on the problem being solved. That is, segmentation should stop when the objects or regions of interest in an application have been detected.

Because that in SEM image, the nanofibers' shapes are similar like lines, thus from image processing mining, we may consider the lines represent the nanofibers in the SEM image. In this chapter we will use segmentation method to extract the nanofibers (lines) from the SEM image.

3.1 INTRODUCTION OF HOUGH TRANSFORM

A line is a collection of edge points that are adjacent and have the same direction. For example, an image contains n points, we want to detect all the subsets of these points which locate on straight lines. One simple brute-force approach is first find all lines determined by every pair of points and then find all subsets of points that are lied on or close to particular lines. This approach will generate a $(n) \frac{n(n-1)}{2} \sim n^3$ computation cost in detecting $\frac{n(n-1)}{2} \sim n^2$ lines.

Paul Hough proposed an alternative method to solve the problem and it is common called “Hough Transform”[18]. Actually Richard Duda and Peter Hart invented the “generalized Hough Transform” which is widely used nowadays[19]. It is an important segmentation method to detect edge or a certain class of shapes by voting procedure. The classic Hough transform was used to detect straight lines in the image, and then it extend to arbitrary shapes like circle and ellipse.

Hough Transform algorithm collect n edge points which detected by an edge detector and find all the lines on which these points lie. The advantage of the Hough Transform is that it use parameters to reduce the time for finding lines based on a set of edge points, and these parameters can be adjusted depend on the different requirements.

The basic idea of Hough transform is to transfer a point from the physical domain to a curve in its parameters' domain. A line in x-y plane can be uniquely defined by its distance r from the origin and the orientation angle θ as $x \cos \theta + y \sin \theta = r$, as shown in Figure 3.1. Therefore, this parameterization maps every line in x-y plane to a point (θ, r) in $\theta - r$ domain. For a point $P_0 (x_0, y_0)$, all lines passing through it could be expressed as $x_0 \cos \theta + y_0 \sin \theta = r$ where r is the distance from the origin to the line, and θ is the angle. Therefore, the point P_0 is mapped from (x_0, y_0) to a curve in the $\theta - r$ domain. Based on this transform, for points lying on the same line in x-y plane, their corresponding sinusoidal curves in $\theta - r$ plane will pass through a common point, which is the line in the $\theta - r$ domain.

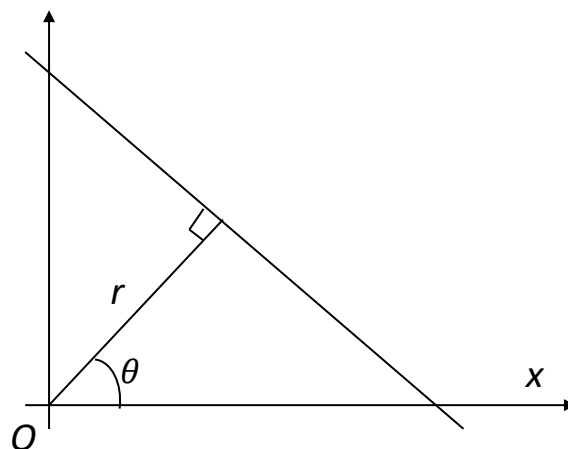


Figure 3.1 Illustration of Hough transform

In the Hough transform algorithm, the θ - r parameter space is divided into accumulator cells to form a two-dimensional matrix. The value of the cell is the number of times a mapping line of points in x-y space intersects that cell. Cells receiving a minimum number of “votes” are assumed to correspond to lines in x-y space. Figure 3.2 illustrate the flowchart of the Hough Transform. The flowchart will end when all $E(r, c)$ have been examined.[20]

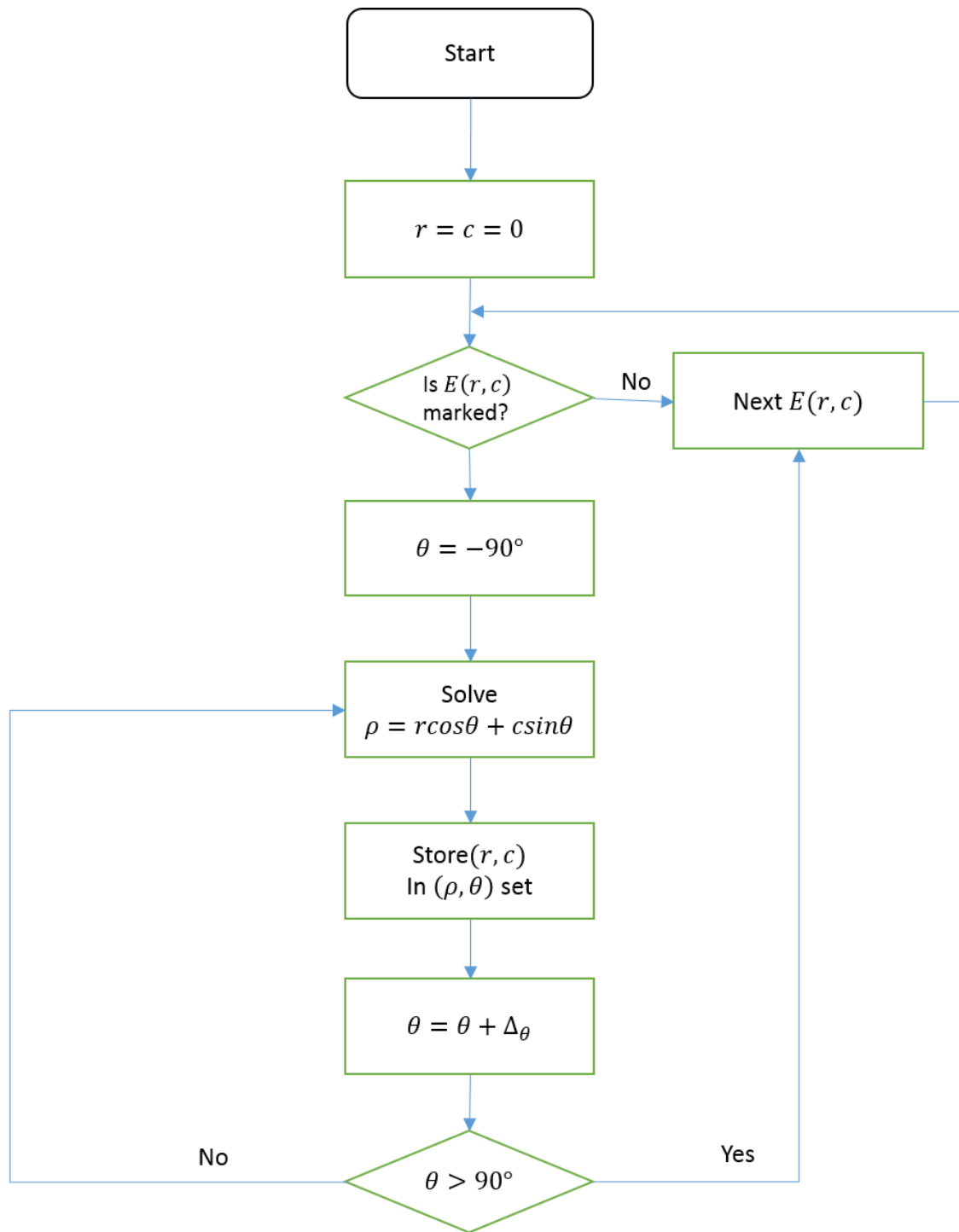


Figure 3.2 Hough Transform flowchart

After these process, the number in each set corresponds to the number of pixels on that straight line which is defined by ρ and θ of that set. The larger quantization set which means

the less processing cost time, but on the other hand, the trade-off is the detection resolution in this certain image is lower. The following step is to use a threshold to check the quantization set whether the set has more points than the threshold. Also, we need to check the continuity of these points, some of the points may not adjacent with each other, this could be done by compute the distance between the pair of points, then check these gaps whether are bigger than the threshold, after that we may decide the line should be divided into two lines which means that one line is on the extension orient of the other line. When all these process completed, we may get all points information(X Y coordinate value) of each detected lines.

3.2 PARTITIONING HOUGH TRANSFORM

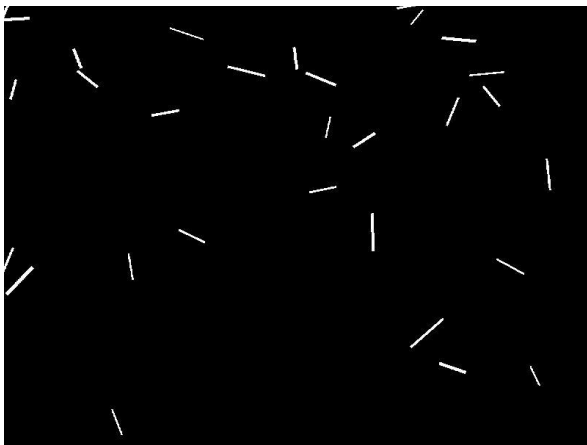
To extract nanofibers, we could use Hough transform to detect lines in SEM images. However, due to the large width, a nanofiber could result in many lines in detection. To overcome this issue, “Skeleton” operation[21, 22] could applied to get the morphological skeleton and then apply Hough transform. However, there is still an issue in the above method. The skeleton of other nanofibers could contribute to the accumulator cell values and may significantly influence the detection accuracy, especially when the density of nanofibers is very large. To overcome this problem, we propose the partitioning Hough transform, where the partitioning is first applied to segment SEM images into multiple images based on connected components and then apply Hough transform to each segmented image. The Partitioning Hough transform is illustrated in Table 3.1 as follows:

Table 3.1. Partitioning Hough transform algorithm for nanofiber segmentation.

-
- Convert the SEM image to a binary image
 - Partition the binary image into n SEM images with each having one connected component
 - For $i=1:n$
 - Extract the morphological “skeleton” of image i
 - Perform Hough transform on image i to get the Hough matrix
 - Identify the peaks from the Hough matrix
 - Detect straight lines based on the peaks and Hough matrix
 - End
-

3.3 SIMULATION CASE STUDY

In the following sections, simulated SEM images of nanofiber reinforced nanocomposites are used to evaluate the proposed method. The image resolution is 2400×1800 pixels. The nanofibers were randomly distributed with a uniform density $\rho = 0.0005$ wires pixel^{-2} . length and width of each nanofiber follow normal distribution with mean of 10 and 120 and standard deviation of 2 and 20 respectively. Figure 2.3 shows simulated SEM images with 30,100,200,400 fibers, which will represent different densities of the nanofibers.



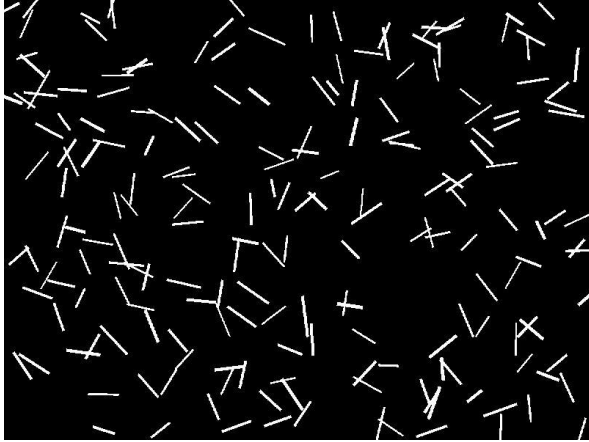
(a)



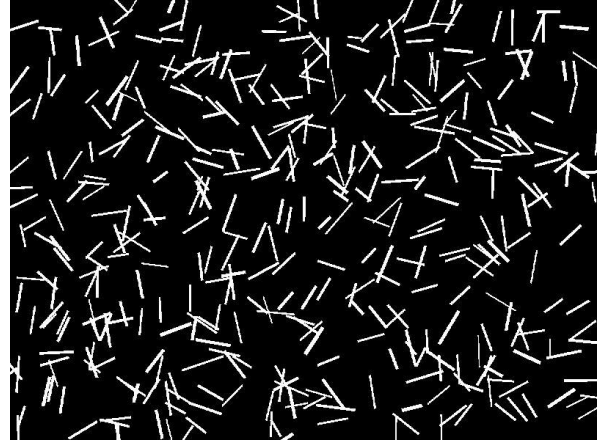
(b)

Figure 3.3 (a) Low density SEM simulation

(b) Median density SEM simulation



(c)

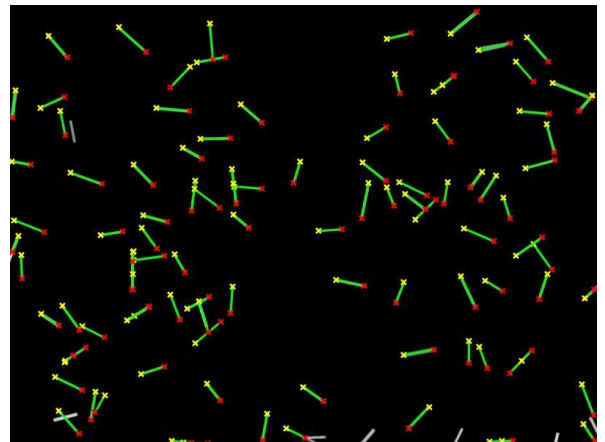
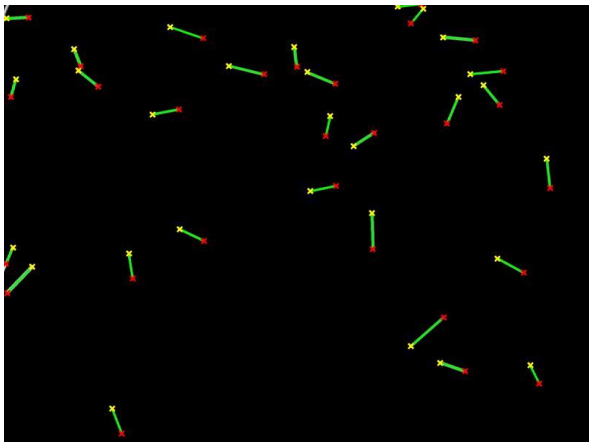


(d)

Figure 3.3 (c) High Density SEM simulation

(d) Extremely High Density SEM simulation

The case study deploys the general simple Hough Transform method and Partitioning Hough Transform method into the simulation Nanofiber SEM images which contain 30 (low density), 100 (medium density), 200 (high density), 400 (extreme high density) nanofibers. After the implementation, we get the result as shown in Figure 3.4 (simple Hough Transform method) and Figure 3.5 (Partitioning Hough Transform method) and the data in Table 3.2.

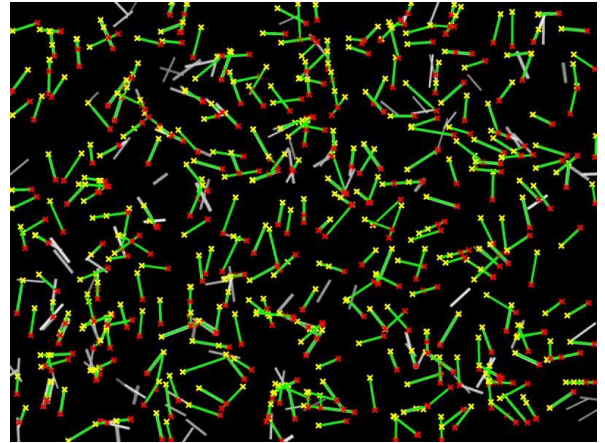
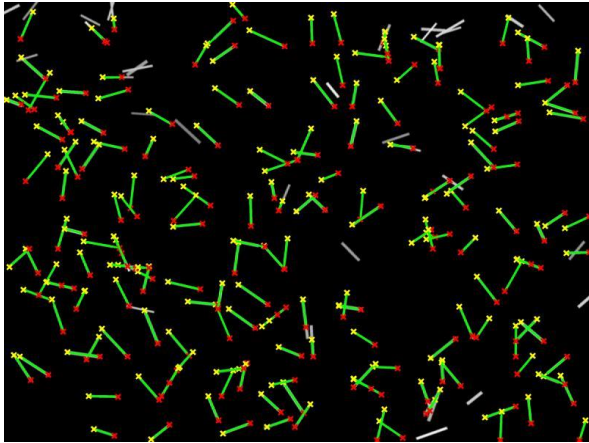


(a)

(b)

Figure 3.4 (a) 30 nanofibers simulation using simple Hough Transform method

(b) 100 nanofibers simulation using simple Hough Transform method

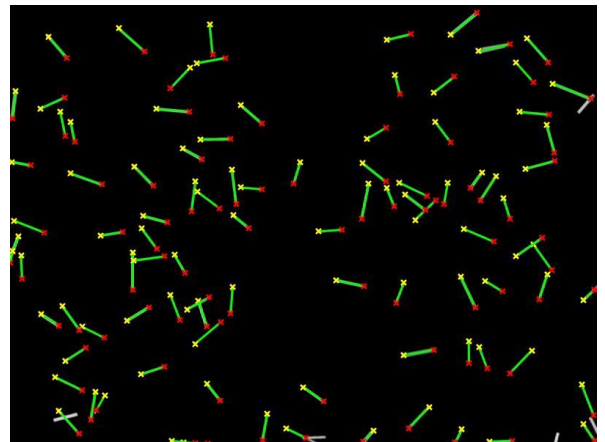
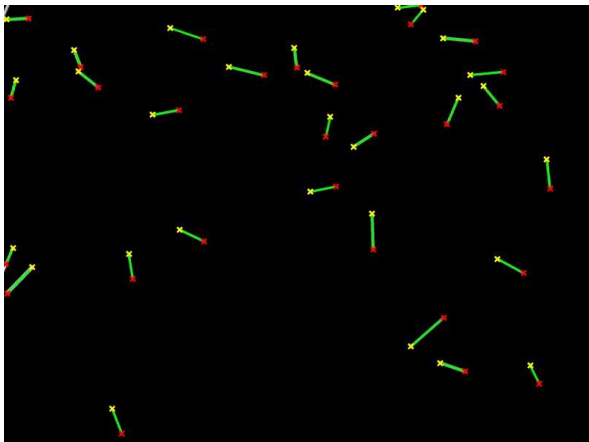


(c)

(d)

(c) 200 nanofibers simulation using simple Hough Transform method

(d) 400 nanofibers simulation using simple Hough Transform method

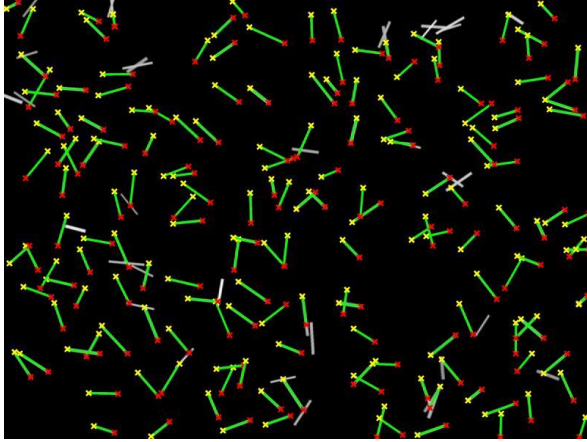


(a)

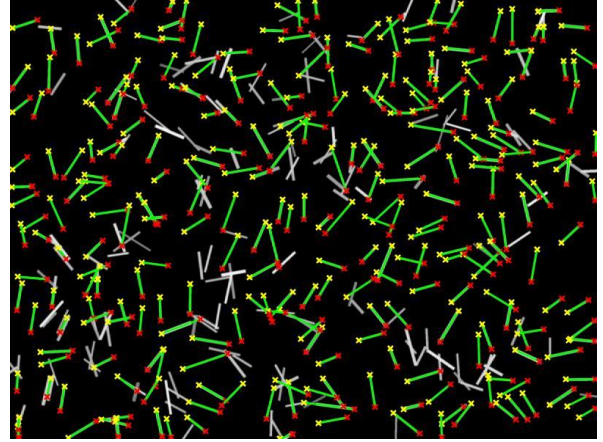
(b)

Figure 3.5 (a) 30 nanofibers simulation using Partitioning Hough Transform method

(b) 100 nanofibers simulation using Partitioning Hough Transform method



(c)



(d)

Figure 3.5 (c) 200 nanofibers simulation using Partitioning Hough Transform method

(d) 400 nanofibers simulation using Partitioning Hough Transform method

Table 3.2 Simulation Case Study Results Data

Numberof nanofibers	Simple Hough Transform Detected	Accuracy of Simple Hough Transform	Partitioning Hough Transform Detected	Accuracy of Partitioning Hough Transform
30	29	0.97	29	0.97
100	90	0.92	95	0.95
200	170	0.85	174	0.87
400	308	0.77	310	0.78

After analysis of the data table above, we plot an accuracy analysis result in the Figure 3.6 as shown in below:

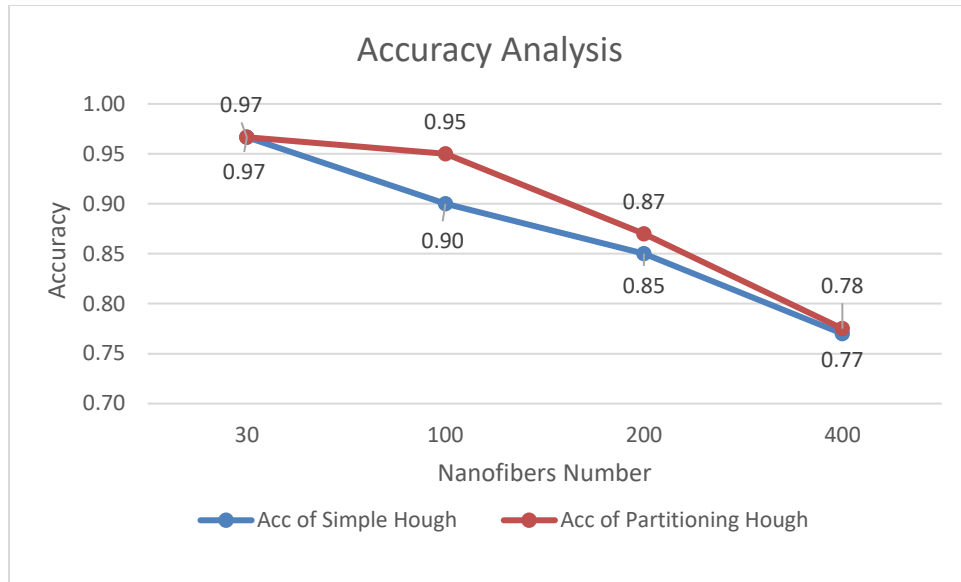


Figure 3.6 Accuracy Analysis of the Segmentation Method

We may find some important discovery and information through the analysis:

- 1, The “Partitioning Hough Transform” method’s performance is better than the “Simple Hough Transform”.
- 2, In high density and extreme high density simulation SEM images’ case studies, the accuracy decreased a lot through both methods.
- 3, In the low density simulation SEM image case study, one same nanofiber is not detected through both methods.

After investigation, we find the reasons for the above things which we discover:

- 1, Simple Hough Transform method would include one disconnected point which is on the extended project lines. This error could happen as shown in Figure 3.7, it is a zoom in the 200 nanofibers simulation case study.

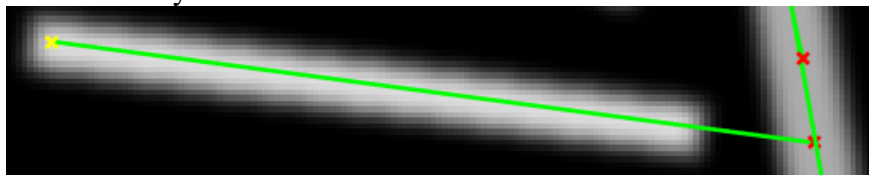


Figure 3.7 Wrong Detection of simple Hough Transform

The root cause of it is that, the “Skeleton” operation will cause some gap between some points which were connected before the operation, Hough Transform function provide a parameter called “FillGap” to solve the problem. This parameter is a trade-off since it would cause some wrong detections.

2, The second problem is why in high density and extreme high density simulation case studies, the accuracy decreased a lot, The reason is that because of the complexity. With the nanofibers increased in one image, there would have more overlapping. Then after the “Skeleton” operation, the shapes of the pattern would change dramatically. This would give a negative impact during the Hough Transform. Figure 3.8 shown one pattern which encounter this problem. It is from the 400 nanofibers simulation SEM image.

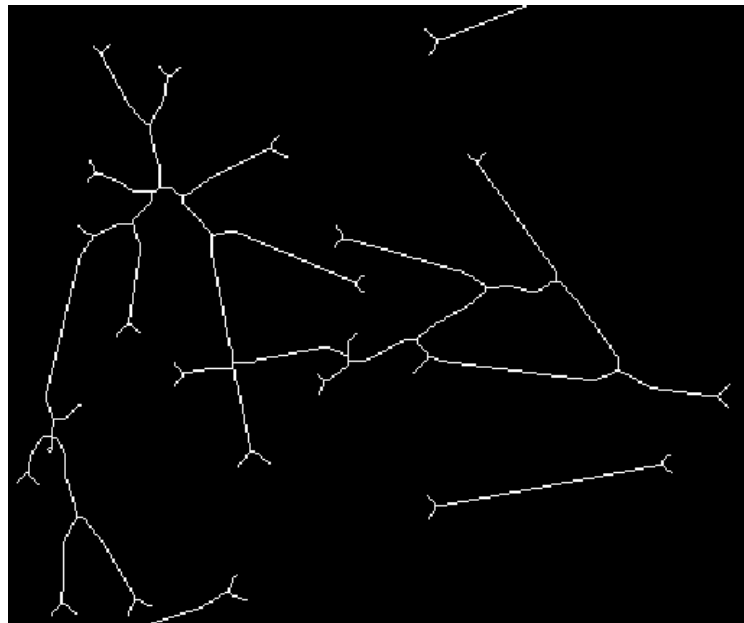


Figure3.8 Problem exist after “Skeleton” operation

3, In 30 nanofibers simulation case study, one miss detected nanofiber is on the left top corner of the image. The length of nanofiber in this image is not long because it is cut by the outer edge of the image. In the Hough Transform function, it provide a parameter called “MinLength” to give a lower limit of the length. Set this parameter could prevent some miss detection but it is also a

trade-off, since some nanofibers may not be detected when they located on the edge of the image. Figure 3.9 display the nanofiber not detected by both methods.

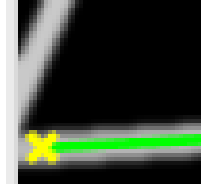


Figure 3.9 corner nanofibers miss detected

3.4 EDGE BASED PARTITIONING HOUGH TRANSFORM

In section 3.2, the partitioning Hough transform separated the image into several parts to prevent some mis-detections from the original image. The partitioning Hough transform did correct some error and improve the detection accuracy compared with the simple Hough transform method, but it still has some issues which list in the analysis part in section 3.3 need to be resolved.

As problem 2 said, the “Skeleton” operation would cause some pattern in the partitioning part distorted or misshapen in the high density situation, this would affect the Hough transform cannot get the right point peaks. Therefore, the root cause is the “Skeleton” operation. But when using Hough Transform, it is better to make the nanofibers as thin as possible. The edge detection/extraction method could help to resolve the problem without “Skeleton” operation.

3.4.1 Introduction of edge detection

In this section, several segmentation methods will be discussed and introduced.

The three type of image features in which we are interested are isolated points, lines, and edges. Edge pixels are pixels at which the intensity of an image function changes abruptly and edges are sets of connected edge pixels. Edge detectors are local image processing methods designed to detect edge pixels. A line may be viewed as an edge segment in which the intensity of the background on either side of the line is either much higher or much lower than the intensity of the line pixels.

Edge detection operators are often discrete approximations to differential operators. From calculus that the derivative measure the rate of change of a line, or the slope of the line, and first-order and second order derivatives will be applied on the image to get the purpose. The first derivative will mark edge points, with steeper gray level changes providing stronger edge points (larger magnitudes), the second derivative returns two impulses, one on either side of the edge.

To obtain the gradient of an image requires computing the partial derivatives $g_x = \partial f / \partial x$ and $g_y = \partial f / \partial y$ at every pixel location in the image.

The gradient of an image f is denoted by ∇f , Equation 3.1 defined as the vector

$$\nabla f = \text{grad}(f) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (3.1)$$

This vector has the important geometrical property that it points in the direction of the greatest rate of change of f at location (x, y) ,

The magnitude (length) of vector ∇f , denoted as $M(x, y)$, where

$$M(x, y) = \text{mag}(\nabla f) = \sqrt{g_x^2 + g_y^2} \quad (3.2)$$

The direction of the gradient vector is given by the angle α measured with respect to the x-axis.

$$\alpha(x, y) = \tan^{-1} \left[\frac{g_y}{g_x} \right] \quad (3.3)$$

We obtain an approximation to be the first derivative at point x of one-dimensional function $f(x)$ by expanding the function $f(x + \Delta x)$ into a Taylor series about x , letting $\Delta x = 1$, and keeping only the linear terms, the result is the digital difference

$$\frac{\partial f}{\partial x} = f'(x) = f(x+1) - f(x) \quad (3.4)$$

The same, we get the second derivative of the $f(x)$ by differentiating

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial f'(x)}{\partial x} = f(x+2) - 2f(x+1) + f(x) \quad (3.5)$$

Here we may use $x = x+1$ to rewrite the Equation 2.5 and get

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial f'(x)}{\partial x} = f(x+1) + f(x-1) - 2f(x) \quad (3.6)$$

The derivation of g_y is similar.

In image processing, we need 2-D mask to scan the all the pixels to detect the different edges in it. Figure 3.10 shown a 3×3 region of an image (the p 's are intensity values) and various masks used to compute the gradient at the point labeled p_5 .

p_1	p_2	p_3
p_4	p_5	p_6
p_7	p_8	p_9

Figure 3.10 2D mask

Roberts cross-gradient operator[23] is a simple approximation of the first derivative and it was the first edge detector proposed by Lawrence Roberts. Equation 3.7 and 3.8 shown the g_x and g_y

$$g_x = \frac{\partial f}{\partial x} = (p_9 - p_5) \quad (3.7)$$

$$g_y = \frac{\partial f}{\partial y} = (p_8 - p_6) \quad (3.8)$$

Hence, the Roberts operators' masks can be shown in Figure 3.11

-1	0	0	-1
0	1	1	0

Figure 3.11 Roberts masks

Both Prewitt [16] operator and Sobel operator [17] use the second derivative, the masks are shown in Figure 3.12 and Figure 3.13.

-1	-1	-1	-1	0	1
0	0	0	-1	0	1
1	1	1	-1	0	1

Figure 3.12 Prewitt masks

-1	-2	-1	-1	0	1
0	0	0	-2	0	2
1	2	1	-1	0	1

Figure 3.13 Sobel masks

The difference between these two operators is Sobel operator use a weight of 2 in the center coefficient. Thus, Prewitt masks is easier to compute and implement than the Sobel operator with the same hardware resource. But the Sobel masks have a higher better noise-suppression.

Similar with the Prewitt and Sobel operators, Laplacian operators[24] are another approximations of the second derivative. In Figure 2.5 shows below represented three different approximations of the Laplacian operators. Unlike the Prewitt and Sobel masks, the Laplacian masks are rotationally symmetric, which means edges at all orientations contribute to the result. Hence, only one mask needed to be selected to apply with the image.

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

-2	1	-2
1	4	1
-2	1	-2

Figure 3.14 Laplacian masks

Except the Robert, Prewitt, Sobel and Laplacian operators, there are some other advance edge detectors have been developed these several years. Table 3.3 lists some methods of edge detectors.

Table 3.3 Some advanced edge detector algorithms

Method Name	Method Procedure	Ref
Laplacian of Gaussian(LoG)	<ol style="list-style-type: none"> 1. Use a Gaussian filter to smooth the image 2. Use a Laplacian mask convolve the image 	[25, 26]
Canny algorithm	<ol style="list-style-type: none"> 1. Use a Gaussian filter to smooth the image. 2. Get the gradient magnitude and angle image 3. Apply nonmaxima suppression 4. Use two thresholds to detect and link edges. 	[27]
Boie Cox algorithm	Generalization of the Canny algorithm, use Wiener filters to allow for a more generalized edge model	[28]
Marr-Hildreth edge detector	<ol style="list-style-type: none"> 1. Apply the LoG filter with the image 2. Find the zero crossing of the image 	[26, 29]

3.4.2 Algorithm of Edge based partitioning Hough Transform

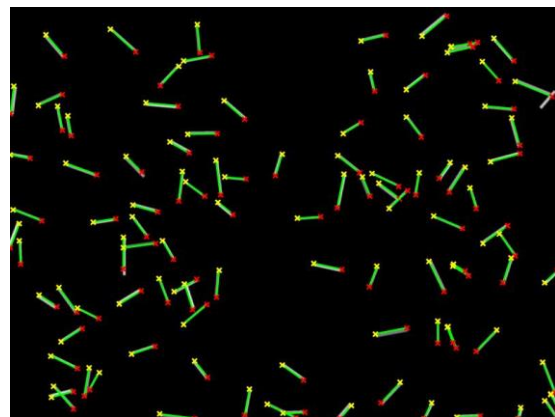
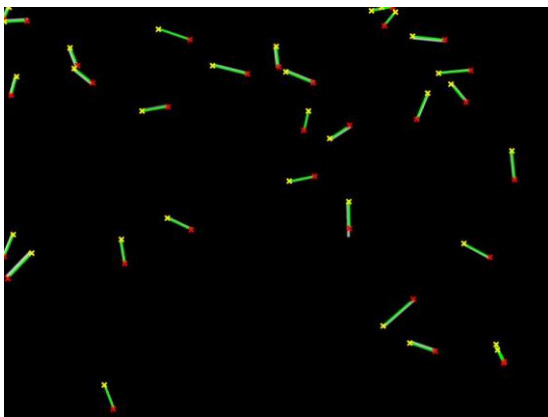
Here we use Laplacian masks to detect the edges of the partition part instead of the “Skeleton” operation, thus, we may get a or several pairs of parallel lines from the edges, check the length of them and remove the short one in each pair, then we may extract the lines from the boundary. The algorithm of Edge based partitioning Hough Transform is shown in following Table 3.4.

Table 3.4 Table 3.4. Edge based partitioning Hough transform algorithm for nanofiber segmentation.

Step 1	Convert the SEM image to a binary image
Step 2	Partition the binary image into n SEM images with each having one connected component
Step 3	<p>For $i=1:n$</p> <ul style="list-style-type: none"> • Extract edges using Laplacian masks • Perform Hough transform on image i to get the Hough matrix • Identify the peaks from the Hough matrix • Detect straight lines based on the peaks and Hough matrix • Check the parallel lines based on closed ρ value, remove the short one. <p>End</p>

3.4.3 Case study and analysis

We apply the new method-edge based partitioning Hough Transform on the same simulation SEM images we used previous. In Figure 3.15 shows the result.

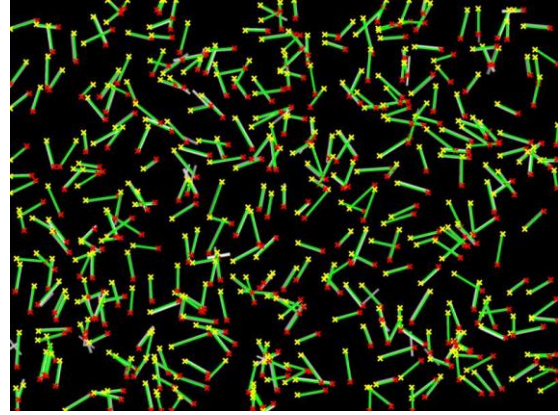
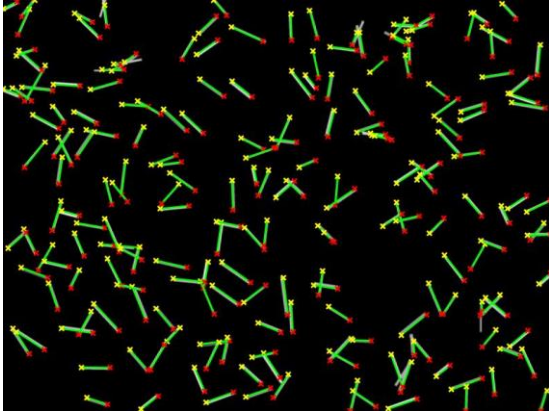


(a)

(b)

Figure 3.15 (a) 30 nanofibers simulation using Edge based Partitioning Hough Transform method

(b) 100 nanofibers simulation using Edge based Partitioning Hough Transform method



(c)

(d)

Figure 3.15 (c) 200 nanofibers simulation using Edge based Partitioning Hough Transform method

(d) 400 nanofibers simulation using Edge based Partitioning Hough Transform method

We may find the detection error has been greatly improved from the simulation case study result data illustrated in Table 3.5. Figure 3.16 shows the comparison between three method using Hough Transform.

In high density situation, it can still maintain a high detection rate.

Table 3.5 Simulation case study results

Number of Nanofibers	New method detected	Accuracy of Edge based partitioning Hough Transform
30	30	1
100	100	1
200	196	0.98
400	394	0.97

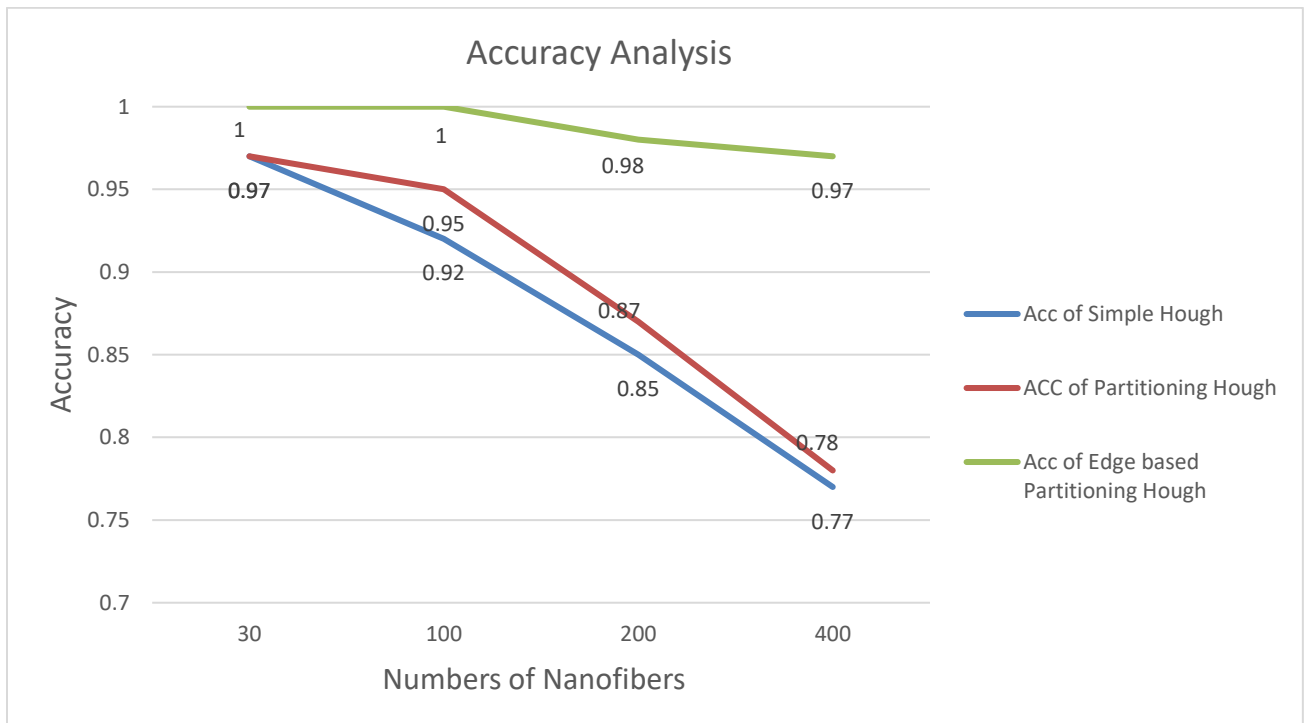


Figure 3.16 Accuracy Analysis of simulation case study

In conclusion, the developed method based on Hough transform is effective and efficient in the nanofiber segmentation. With this method, nanofibers morphology analysis can be automatically implemented. But we may not forget, there are still some problem and defects need to be solved. It could be improved by some other methods.

Chapter 4 Nanofiber segmentation based on Boundary detection and multiple Changepoints models

Chapter 3 describe using Hough Transform based method to extract the lines' information in the images. In this chapter, we will use another innovatory method to solve the problem. This method change the problem from a spatial image processing problem to a statistical problem. The detection process adapt a data mining approach to get the critical feature (change point), after that through this change points, we may extract the lines in some specific ways.

The statistical method including three major procedures which is illustrated in the Table 4.1.

Table 4.1 Major procedures in the statistical method in this chapter

-
- Use a certain boundary method to extract the boundary data for every non-connected components in the giving image.
 - Run the change point algorithm to detect the change points for each boundary data.
 - Find the right lines by some geometric mathematic mechanism based on the detected change points' information.
-

4.1 BOUNDARY METHOD

Tracing boundary or called contour/border is a method of digital image processing[13]. It is performed by detecting boundaries between objects, then defining the objects indirectly. This method usually begin with marking points which may be a part of an edge. After, these points are then combine to line segments, and the line segments combine to the boundaries. In Chapter 3, we have introduced several edge detectors, these detectors most used to mark points of rapid change, which indicating the possibility of an edge. These edge points represent local discontinuities in specific features, for instance, the color, brightness and texture. Besides these segmentation method, some morphology method can be used to get the boundaries of certain patterns in a given

image. Using the operation in Chapter 2 like erosion or dilation, we may get the edges easily by compute the difference between before and after the operation.

Also, there are several other representation approaches can be used to extract the boundary, such as the “Freeman Chain Code” [30-34], “Minimum-Perimeter Polygon(MPP)” [35, 36] and “Moore Boundary Tracking Algorithm” which will be applied in this section. “Moore Boundary Tracking Algorithm” whose output is an ordered sequence of points. These type of data can be utilize in our statistical change point algorithm.

4.1.1 Moore Boundary Tracking Algorithm

In Moore Boundary Tracking Algorithm, the Moore neighborhood of a pixel is the basis of this algorithm. In a word the Moore neighborhood of a pixel P is the set of 8 pixels for 8-neighbors or 4 pixels for 4-neighbors, which surround that certain pixel. These pixels are namely pixels P_1 to P_8 for 8-neighbors or P_1 to P_4 for 4-neighbors which shown in Figure 4.1.

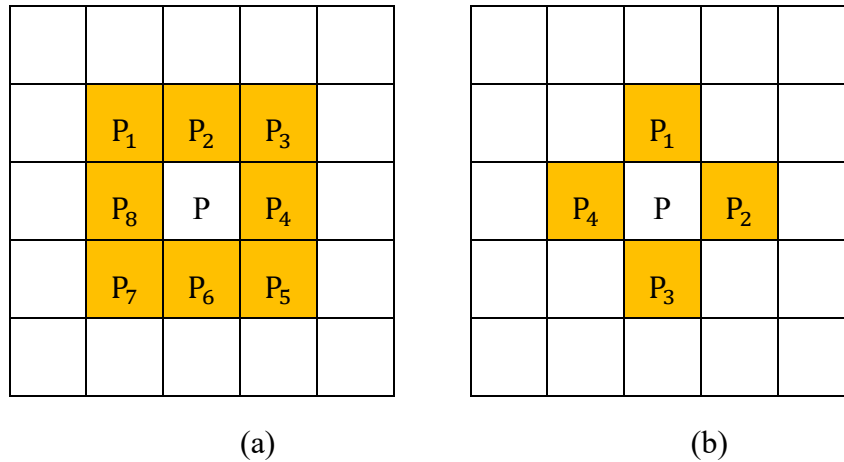


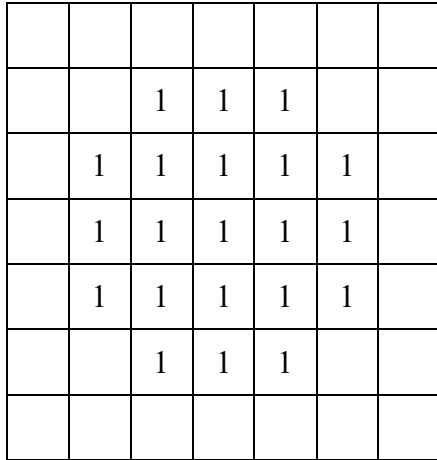
Figure 4.1 (a) 8-neighbors Moore Neighborhood

(b) 4-neighbors Moore Neighborhood

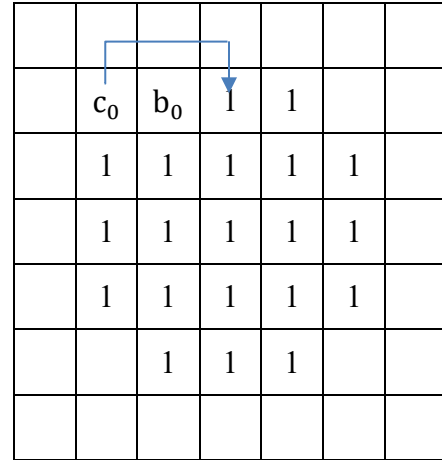
After, given a binary region R , the Moore boundary tracking algorithm for this region consists of the following steps in table 4.2, the tracking process are illustrated in Figure 4.2(a)-(e), and the correct boundary is shown in Figure 4.2(f) whose points are a clockwise-ordered sequence.

Table 4.2 Moore boundary tracking algorithm working process

-
1. Let the starting point, b_0 be the uppermost, leftmost point in the image. Denote by c_0 the west neighbor of b_0 Figure 4.2 (b). Clearly, c_0 always is a background point. Examine the 4- or 8-neighbors of b_0 , starting at c_0 and proceeding in a clockwise direction. Let b_1 denote the first neighbor encountered whose value is 1, and let c_1 be the (background) point immediately preceding b_1 in the sequence. Store the location of b_0 and b_1 for use in Step5.
 2. Let $b = b_1$ and $c = c_1$. Figure 4.2(c)
 3. Let the 4- or 8- neighbors of b , starting at c and proceeding in a clockwise direction, be denoted by n_1, n_2, \dots, n_x . Find the first n_k labeled 1.
 4. Let $b = n_k$ and $c = n_{k-1}$
 5. Repeat Step 3 and 4 until $b = b_0$ and the next boundary point found is b_1 . The sequence of b points found when the algorithm stops constitutes the set of ordered boundary point.
-



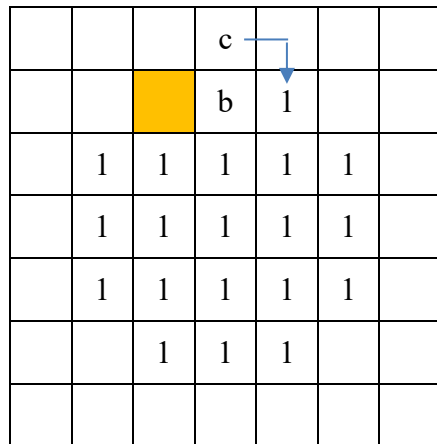
(a)



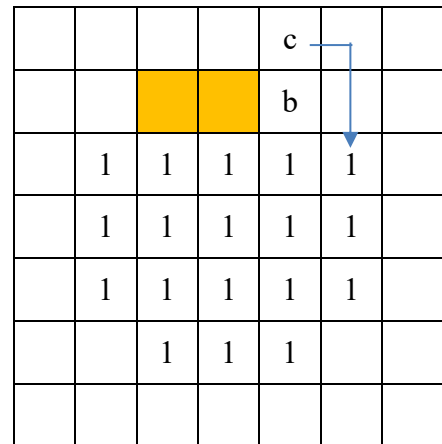
(b)

Figure 4.2 (a) given binary region R

(b) Initiate the start point and find the first b_1



(c)



(d)

Figure 4.2 (c) Set $b = b_1, c = c_1$

(d) repeat the step 3 and 4

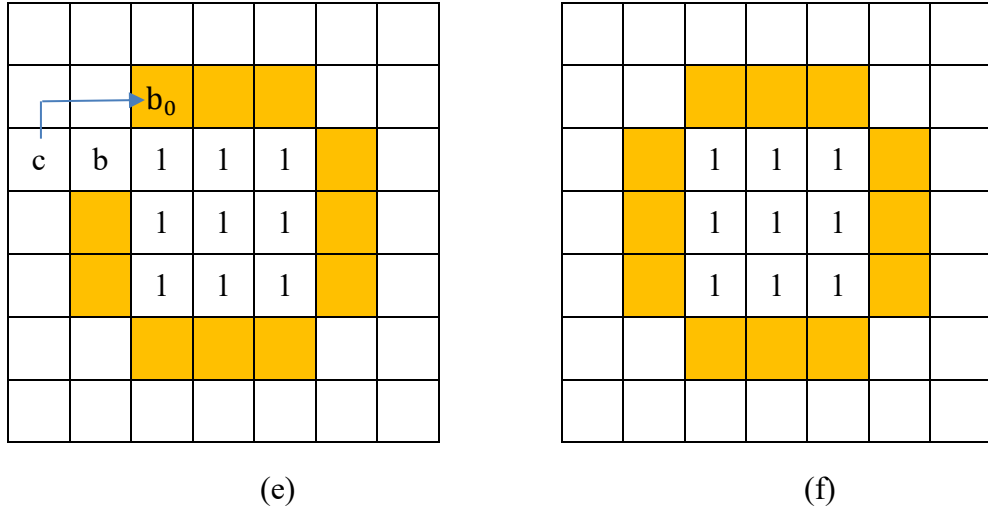


Figure 4.2 (e) meet the stopping criterion in step 5

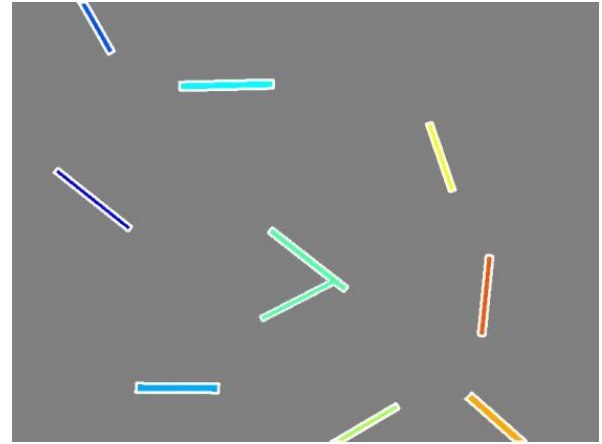
(f) Full boundary points of region R

In matlab image processing toolbox, $B = bwboundaries(BW)$ function is used to traces the exterior boundaries of objects, as well as boundaries of holes inside these objects of the binary image BW and then store each boundaries' points' location value in the certain data set which is a cell array contain the X, Y coordinates. This function also descends into the outermost objects (parents) and traces their children (objects completely enclosed by the parents). As the function utilize the “Moore Boundary Tracking Algorithm”, the points' location of each boundary is in a clockwise-ordered.

Figure 4.3(a) is a simulation of SEM image including 10 nanofibers. Figure 4.3(b) illustrate the boundaries of each non-connected components, pay attention to the two nanofibers in the center of the image, as there are a little portion of these two nanofibers overlapped, both of them are displayed as the same color which indicate these two nanofibers constitute one connected components. Therefore, the total non-connected components' number in the Figure 4.3(b) is 9.



(a)



(b)

Figure 4.3(a) a simulation SEM image with 10 nanofibers

(b) Result after using the bwboundaries function in matlab

Once a boundary data has been generated, we got a series of pair data, in each pair data contains X and Y coordinates values of one boundary point. Figure 4.4 illustrate an example of the boundary image shown in matlab contains three nanofibers be piled up.

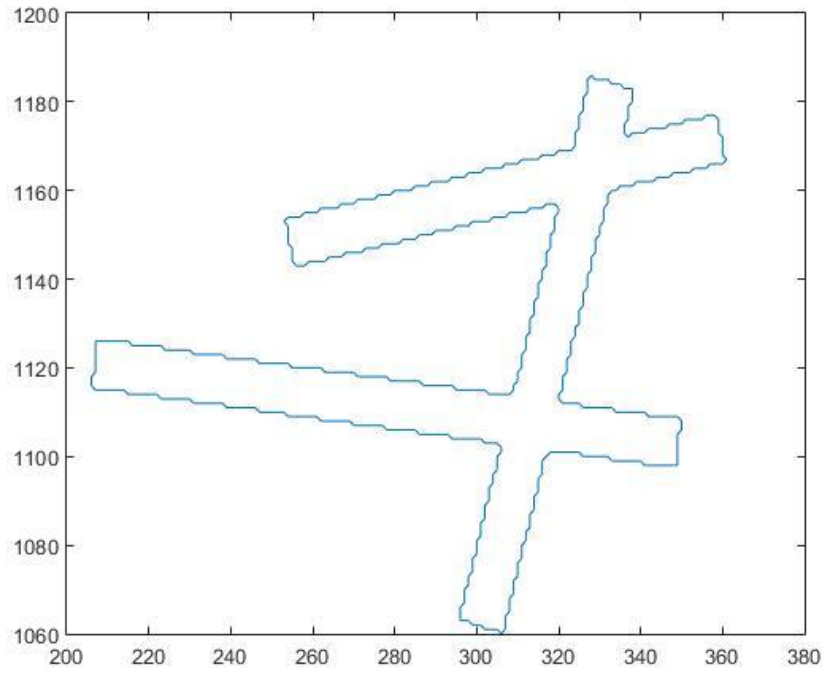


Figure 4.4 Boundary of 3 connected nanofibers

In order to extract right lines through these boundary data, change points detection is applied in the next section.

4.2 CHANGE POINT METHOD

4.2.1 Introduction of Change Point Algorithm

In statistical analysis, change detection or change point detection tries to identify times when the probability distribution of a stochastic process or time series changes. In general the problem concerns both detecting whether or not a change has occurred, or whether several changes might have occurred, and identifying the times of any such changes.

Since in some complex SEM image, especially for some high density Nanofiber SEM image, there could be some intersections or overlapped by several Nanofibers. The change points in the boundary data could provide some critical information, for example the corners of one line

or the intersection of two or more lines. Each change point indicate an abrupt steep trend change in this series of points.

The change point algorithms are usually divided into two type: “online” and “offline”. The term

online sometimes refers to “concurrently” which means “fast enough to keep up”, in the real cases, the online algorithms can predict the change point through the current and previous data before the true change happens, or detect a change point as soon as possible after it occurs. Some other times, online it means “incremental”, which means the algorithm update a new change point based on the original/previous stored result instead of re-computing from scratch. Compared with the online algorithms, the offline algorithms usually run without the real-time constraints. An offline algorithm derive the results using the entire data set at once. Hence, “Online” or “Offline” is just a description of the implementation, in most of time, it is not related to the intrinsic characteristic of the algorithms. Actually, some algorithms can be deployed in both online and offline scenarios.

The change point algorithms have been researched and developed for quit a long time, there are huge literatures introduced all kinds of algorithms, there are several typical algorithms, like “Shewhart control charts”, “Geometric Moving Average(GMA) control charts”, “Finite Moving Average(FMA) charts”, “CUSUM type algorithm”, “Generalized Likelihood ratio(GLR) test” and “Bayes type algorithms”. Table 4.3 illustrate paper, books or documents of these algorithms’ theories, properties and some applications.

Table 4.3 Some Online change point detection method list

Online detection algorithms	Theory Documents	Application
Shewhart control charts	[37]	[38, 39]
GMA control charts	[40]	[41, 42]
FMA charts	[43, 44]	[45]
CUSUM type algorithm	[46-48]	[49-51]
GLR test	[48, 52]	[53-55]
Bayes type algorithms	[56-58]	[59-61]

Offline detection methods usually deal with the whole data set to detect the single or multiple change points. In some time, the number of change points are unknown, many methods use the penalized version method to get the best change points' number. Also, in most cases, the methods were built based on dynamic programming. Therefore, a matrix need to be computed, it may require more time and other resource cost. Especially nowadays, the big data era is coming. We may face a new challenge, whether the capacity and efficiency of our systems or methods is enough. In the next section we will introduce some offline change point methods focusing on improving the performance of the detection methods.

4.2.2 Introduction of PELT Algorithm

For unknown number of change points, there are two widely used methods which support multiple change point detection. One is “Binary Segmentation”(BS) [62, 63] and another is “Segment Neighborhood”(SN) [64]. These two popular methods are the basis of the PELT (Pruned Exact Linear Time) algorithm.

The PELT algorithm was first introduced in the paper “Optimal detection of change points with a linear computational cost”.

The paper proposes a new algorithm PELT that scales linearly in running time with the size of the input time series to detect exact locations of change points. The paper aims to replace both

an approximate binary partitioning algorithm (BS), and an optimal segmentation algorithm (OP) which didn't involve a pruning mechanism to speed up the running time. The paper uses an MLE algorithm at the heart of its dynamic partitioning in order to locate change points. Next we will look for detail in this PELT algorithm.

In this paper, an ordered sequence of data, $y_{1:n} = (y_1, \dots, y_n)$ has been used for analysis. Our boundary data is also an ordered sequence of pairs data. The model will have m change points with their location address $\tau_{1:m} = (\tau_1, \dots, \tau_m)$. Each change point's location is a positive integer between 1 and $n - 1$ inclusive. The τ and $\tau_{m+1} = n$ are assigned as 0 and n . The change points' order is assume that $\tau_i < \tau_j$ if and only if $i < j$. Then, the m change points will split the data into $m + 1$ segments, with the i th segment which including $y_{(\tau_{i-1}+1):\tau_i}$. The common way to search multiple change points is to minimize the Equation (4.1)

$$\sum_{i=1}^{m+1} [C(y_{(\tau_{i-1}+1):\tau_i})] + \beta f(m) \quad (4.1)$$

Then in the Binary Segmentation method is an extension application by the single change point detection method, it start with applying one start change point for the entire data set and check if there is a τ exists that satisfies

$$C(y_{1:\tau}) + C(y_{(\tau+1):n}) + \beta < C(y_{1:n}) \quad (4.2)$$

If equation (4.2) is false, this indicate that there is no more change pint exist in the data segment $y_{1:n}$, and the method stops here. Otherwise, there should be a change point detected and these method will continue be applied in these two segments $y_{1:\tau}$ and $y_{(\tau+1):n}$ which have been split by the change point τ . Then check the equation again in each segment, repeat the previous steps until no more equation satisfied which means no further change point detected. The recursive procedure will get an $O(n \ln n)$ computation cost. Compared with the normal method of equation (4.1), it is more efficient. However, this method will have an accuracy issue, the change points detected is the local change point in each separated data segment which may cause that they are not the globe optimal change points which found by in equation (4.1).

Another Optimal Partitioning (OP) method[65, 66] is similar to the Segment Neighborhood(SN).

The method utilize the equation 4.1, and just let $f(m) = m$, then equation 4.1 can be rewrote as following

$$\sum_{i=1}^{m+1} [C(y_{(\tau_{i-1}+1):\tau_i} + \beta)] \quad (4.3)$$

In the paper of “An algorithm for optimal partitioning of data on an interval,”, Jackson use this OP method to decide the last change point at the beginning. After that, compute the optimal value of the cost function to the cost cost for the optimal partition of the data prior to the last change point plus the cost for the last segment which is thre rest of the segment starts with the last change point. Here, in paper, he use $F(s)$ to denote as the minimization from equation 4.3 for the data set $y_{1:s}$ and $T_s = \{\tau: 0 = \tau_0 < \tau_1 < \tau_2 \dots < \tau_m < \tau_{m+1} = s\}$ be the set of change points in the data set. Also, let $F(0) = -\beta$ as the initial status.

Hence, the $F(s)$ can be derived as following equation

$$\begin{aligned} F(s) &= \min_{\tau \in T_s} \left\{ \sum_{i=1}^{m+1} [C(y_{(\tau_{i-1}+1):\tau_i}) + \beta] \right\} \\ &= \min_t \left\{ \min_{\tau \in T_t} \sum_{i=1}^m [C(y_{(\tau_{i-1}+1):\tau_i}) + \beta] + C(y_{(t+1):n}) \right. \\ &\quad \left. + \beta \right\} \\ &= \min_t \{F(t) + C(y_{(t+1):n}) + \beta\} \end{aligned} \quad (4.4)$$

From the equation 4.4, actually it contain a recursion to get the minimal cost for data set $y_{1:s}$ in terms of the minimal cost for data set $y_{1:t}$ when $t < s$. The recursion be solved in an order of $s = 1, 2, \dots, n$. Then the overall computation time of the solution of $F(n)$ is $O(n^2)$.

Following Table 4.4 list the process steps of the OP method.

Table 4.4 Pseudocode of Optimal Partitioning algorithm

Input:
A set of data of the form, (y_1, y_2, \dots, y_n) where $y_i \in \mathbb{R}$.
A measure of fit $C(.)$ dependent on the data.
A penalty constant β .
Initialize:
Let $n = \text{length of data}$ and set
$F(0) = -\beta, cp(0) = NULL, R_1 = \{0\}$.
Iterate for $\tau^* = 1, \dots, n$
1. Calculate $F(\tau^*) = \min_{0 \leq \tau \leq \tau^*} [F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta]$.
2. Let $\tau' = \arg\{\min_{0 \leq \tau \leq \tau^*} [F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta]\}$
3. Set $cp(\tau^*) = (cp(\tau'), \tau')$
Output
$cp(n)$ is the change point set.

As aforementioned, the PELT method is based on the “Binary Segmentation”(BS) method and “Optimal Partitioning”(OP) method, actually, the main part of change point detection algorithm of PELT is the OP method, but PETL use the equation 4.2 in Binary Segmentation method to “pruning” the iteration procedure in the Optitam Partitioning method to improve the computing effieciency. In PELT method, it assume there exists a constant K and for all $t < s < T$. There could be equation 4.5 which similar with equation (4.2).

$$C(y_{(t+1):s}) + C(y_{(s+1):T}) + K \leq C(y_{(t+1):T}) \quad (4.5)$$

If

$$F(t) + C(y_{(t+1):s}) + K \geq F(s) \quad (4.6)$$

Then, for $t < s < T$, t cannot be the optimal last change point prior to T forever.

With this theorem the PELT method can reduce the number of points which need to be calculated their cost function. Obviously, if K is bigger, the less change point will be detected. Note here, if $K = 0$, the pruning operation will only related to the penalty β . After this ”pruning”

operation, the total computation cost could be reduced to $O(n)$. In Table 4.5 illustrate the steps of PELT algorithm.

Table 4.5 Pseudocode of PELT algorithm

<p>Input:</p> <p>A set of data of the form, (y_1, y_2, \dots, y_n) where $y_i \in \mathbb{R}$.</p> <p>A measure of fit $C(\cdot)$ dependent on the data.</p> <p>A penalty constant β.</p> <p>A constant K that satisfies following equation</p> $C(y_{(t+1):s}) + C(y_{(s+1):T}) + K \leq C(y_{(t+1):T})$ <p>Initialize:</p> <p>Let $n = \text{length of data}$ and set</p> $F(0) = -\beta, cp(0) = NULL, R_1 = \{0\}.$ <p>Iterate for $\tau^* = 1, \dots, n$</p> <ol style="list-style-type: none"> 1. Calculate $F(\tau^*) = \min_{\tau \in R_{\tau^*}} [F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta]$ 2. Let $\tau^1 = \arg \left\{ \min_{\tau \in R_{\tau^*}} [F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta] \right\}$ 3. Set $cp(\tau^*) = [cp(\tau^1), \tau^1]$ 4. Set $R_{\tau^*+1} = \{\tau^* \cap \{\tau \in R_{\tau^*} : F(\tau) + C(y_{\tau+1:\tau^*}) + K < F(\tau^*)\}\}$ <p>Output:</p> <p>The change points recorded in $cp(n)$.</p>
--

4.2.3 Cost Function in PELT algorithm

To detect the change point, both OP Method and PELT need to minimize the cost function equation (4.1), so it comes a question, how to define the cost function C . For example, $C(y_{(\tau_{i-1}+1):\tau_i})$ stand for the cost between $(\tau_{i-1} + 1)th$ point and $\tau_i th$ point in y data set. In this PELT paper[67], the author use the MSE(mean square error) to define the cost function.

In our project, the goal is to find the lines (nanofiber) in the SEM image. In the boundary data we extracted in the previous section 4.1, there are a series of points' data. The points are

adjacent by 4- 8- moore neighborhood. Hence, we may think that these series of points are somewhat continuous.

Since we need to find the lines through these random points, linear regression model usually been adapted to fit into these points. In statistics, linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X . The case of one explanatory variable is called simple linear regression. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. The following equation is a linear regression models, in this equation, y_i is the output variable, variable x_i is the independent predictor, β_0 is the intercept, β_1 is the slope and ε_i is the error or we called it residuals. A residual is defined as the difference between the actual value of the dependent variable and the value predicted by the model. The simple linear regression model is shown in equation 4.7.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (4.7)$$

Here we may use the residuals as the cost function. For example, in cost function $C(y_{(\tau_{i-1}+1):\tau_i})$, calculate the linear regression parameters slope, intercept and the prediction residual using the points' data series. The mean square error (MSE) of this data set is shown in equation 4.8.

$$MSE = \sum_{j=\tau_{i-1}+1}^{\tau_i} \varepsilon_j^2 / (\tau_i - (\tau_{i-1} + 1)) \quad (4.8)$$

Hence, the MSE is the cost of between $(\tau_{i-1} + 1)th$ point and $\tau_i th$ point.

However, there comes to several issue if we use adapt equation 4.7 directly into our points data. First, different from the normal regression data, our data is from the location of the points of computer display system. In computer's hardware, this point is called pixel. The pixels like a matrix on the display facilities. Hence, in 1965, Bresenham develop the line represent mechanism

on these discrete pixel based system.[68]. Figure 4.5 shows an example how the compute display a line representation which from point (0,0) to point (5,8) using Bresenham method.

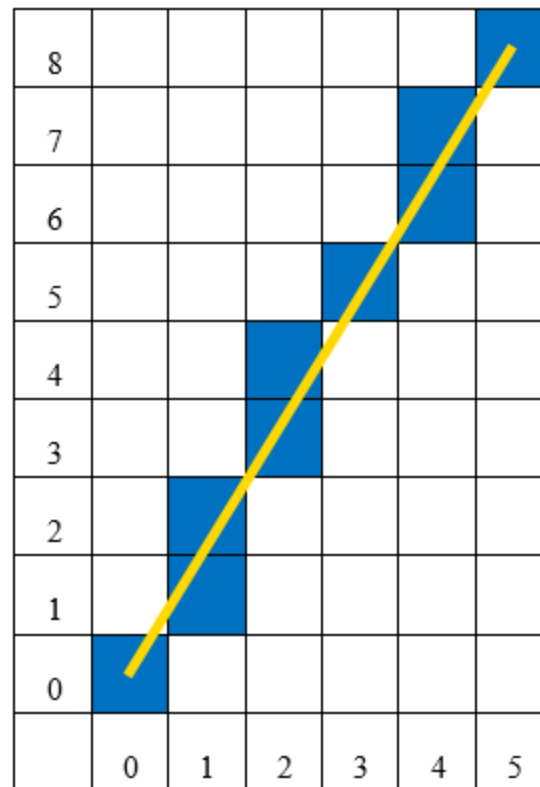


Figure 4.5 Example of line representation on the computer display system

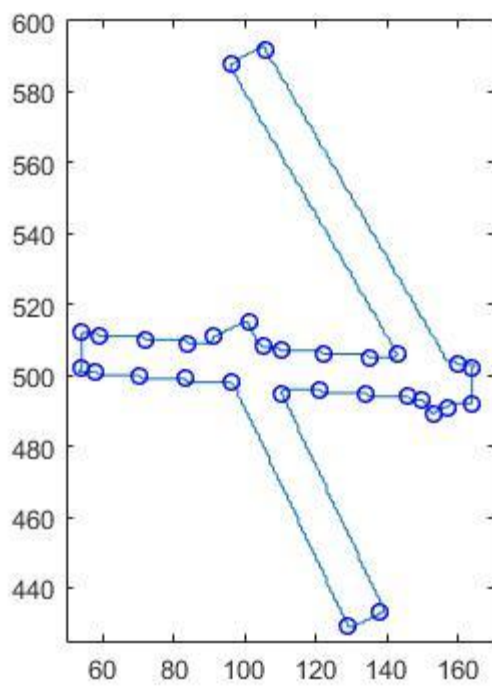
In Figure 4.5, we may find that different from the normal line connect two points, the shapes of these line pixels look like the steps, therefore, in some section of the pixels, the slope of these pixels are 90 degree or $\frac{\pi}{2}$ grad. Especially the original line's slope is large than 45 degree, these pixels' section of 90 degree slope would be more and longer. This would cause to the slope parameter come to infinite, then fail to get the correct regression model which means the residual would be wrong, also the incorrect cost function would lead to the inaccurate change points detection.

The second problem is, the linear regression model in equation 4.7 only contain the residual between y and y predicted. It does not including the error of x. Actually, the points data including the x and y corrdinate value which are independent value which means there is no relation between

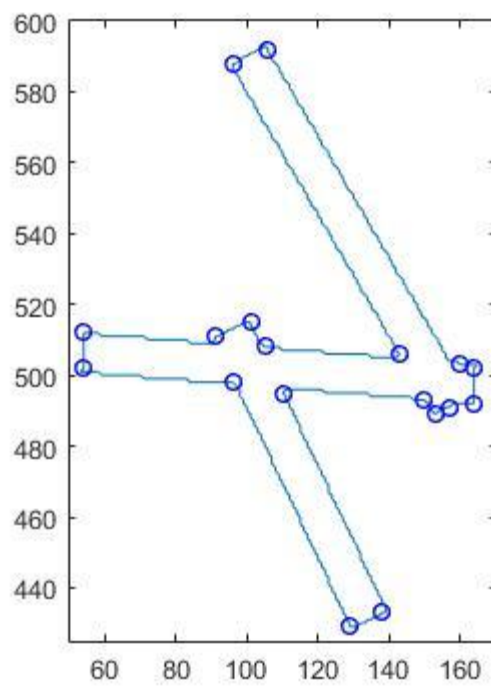
these two values, hence we cannot use the linear regression model here. To overcome this problem, we proposed a workaround solution, which we named it “Synthesized 2D” linear regression model. This solution may also resolve the problem of “infinite slope” which mentioned before .

Synthesized 2D residual method compute the linear regression twice which one for X-axis coordinate value data set and another for Y-axis coordinate value data set. This method transfer the original 2D linear regression model(Deming Regression) into two 1D simple linear regression model. In detail, using the serious number of each point in the boundary data set as the regressor, the X or Y axis coordinate value as the dependent variable to build two simple linear regression model. With these two separated regression model, we may get each MSE based on X and Y data set. This method can detect the residual on both X and Y side. Also, since in each point, there is only one X or Y value, that prevent the problem of “infinite slope”. We just need to sum up the two MSE to get the synthesized cost which could be used to calculate the cost function $C(y_{(\tau_{i-1}+1):\tau_i})$.

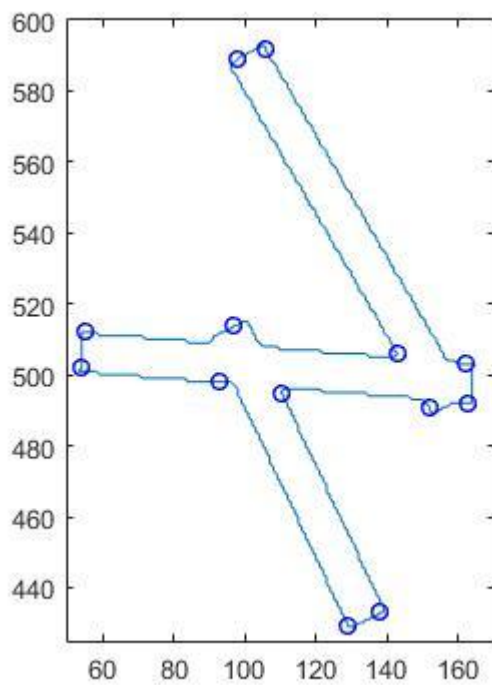
Utilize the synthesized 2D residual method, we may find the change points of the boundary file, notice here, different penalty choices would conduct different results. We may tuning the penalty to get the higher accuracy. Figure 4.6 (a-d) shows the different result based on different penalty β choices(from $\beta = 0.01, 0.1, 1, 2$). After comparism and tuning, the penalty $\beta = 0.1$ would be a better choice.



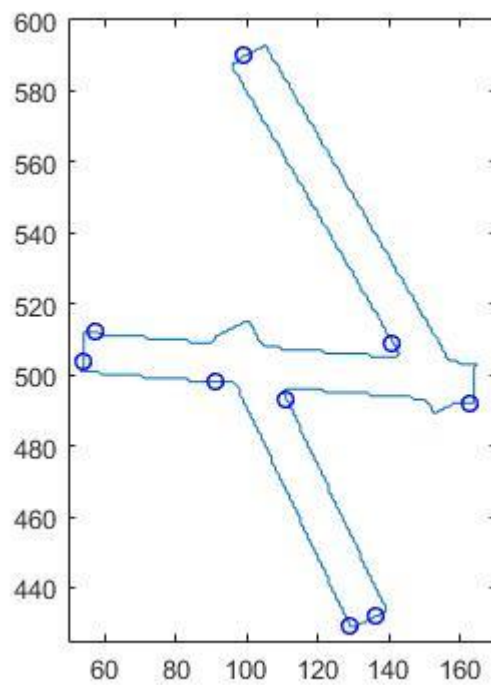
(a)



(b)



(c)



(d)

Figure 4.6 Change points detection when penalty (a) $\beta = 0.01$ (b) $\beta = 0.1$ (c) $\beta = 1$ (d) $\beta = 2$

4.2.4 Comparison using simulation SEM image

Here in this section, first we use 5 different connected components to test the performance of OP with PELT algorithm. The testing computer we use is a desktop PC which CPU is i7-4770(4 core3.4GHz with hyperthreads technology and RAM is 16 Gigabytes).Figure 4.7 shows the time cost difference between OP method and PELT method when using the points of 5 connected components' boundaries' data. Figure 4.8 shows when apply OP method and PELT method, difference of the ratio of cost time divided the length of boundary data for each connect components. We may find that PELT method maintain a stable ratio around 0.001 while the OP method's ratio increased when the length of boundary data increased. This could prove that the PELT's computation cost is $O(n)$ which indicate that the time cost of PELT is linear. Figure 4.9 shows the performance analysis, in the chart, the blue line shows the contrast of the time cost, and from the red line we may derive that the speed ratio is around $n/2$, the n is the number of the points in each boundary data set.

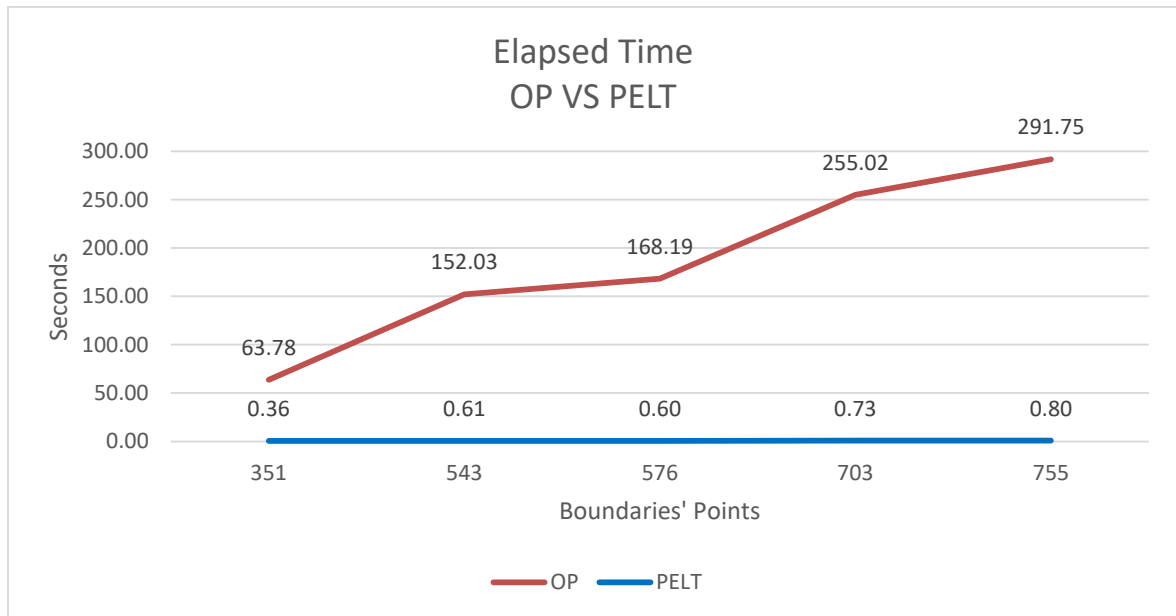


Figure 4.7 The time cost of OP method and PELT method

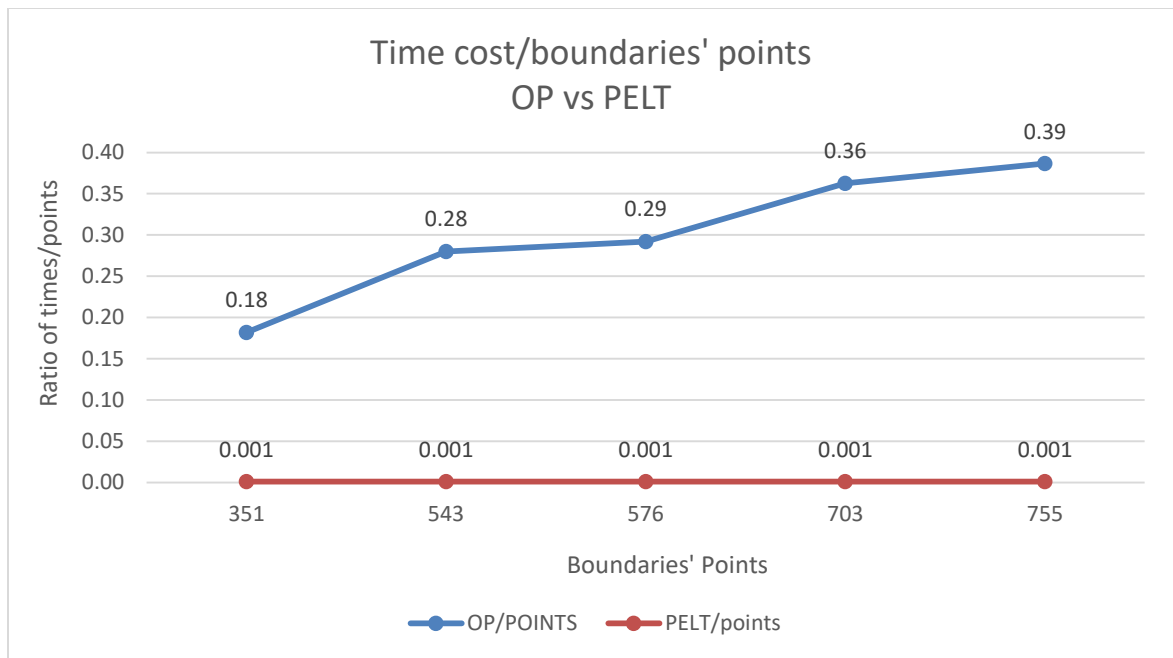


Figure 4.8 The time cost/the length of boundary data

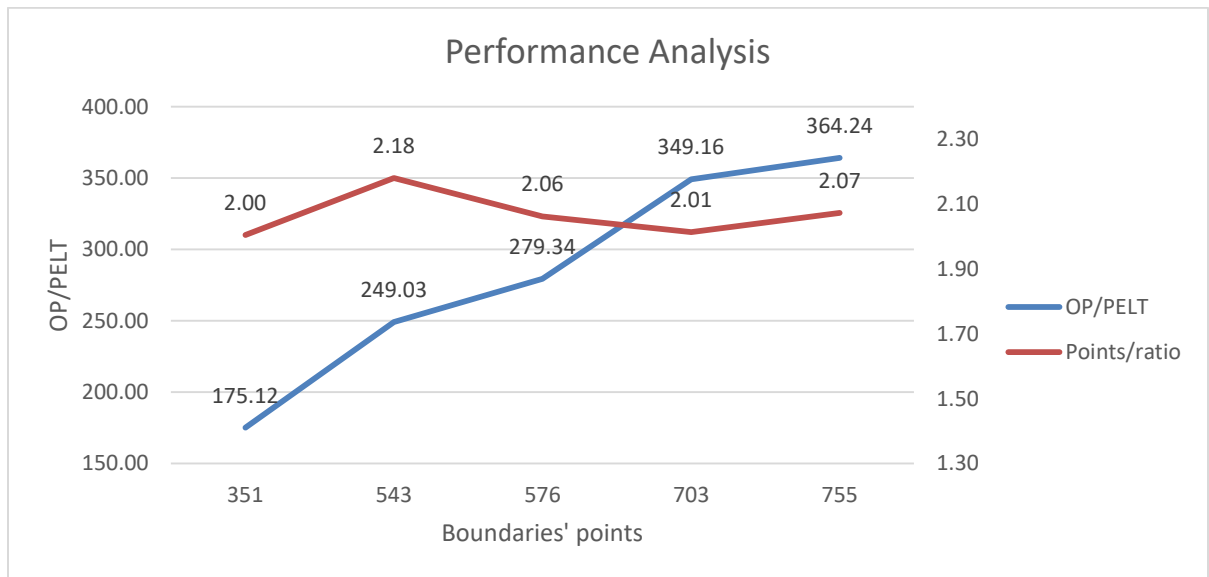


Figure 4.9 Performance Analysis

Above performance analysis is based on the 5 single connected component, after we will continue to test these two change point algorithms in 4 different densities SEM simulation image which contain 30(low density), 100(medium density), 200(high density) and 400(extreme high density) nanofibers. To accelerate the computing speed, we will also use the parallel computing technology. Here, 8 workers in the matlab which utilize a maximum of 8 logical processors has been set. We use four simulation SEM image with different density to compare the performance between OP method and PELT method. Table 4.6 illustrates the computing results.

Table 4.6 OP and PELT method computing time with/out Parallel Computing in seconds

Number of Nanofibers	OP Method	PELT Method	OP Method with Parallel Computing	PELT Method with Parallel Computing
30	663	7	209	2
100	2204	23	641	5
200	8133	46	2037	10
400	21602	89	5138	20

Through the data table, we may find even using Parallel Computing to reduce time about 3.5 to 4 times, the OP method's cost is not acceptable. With Parallel computing, the time spend on different simulation images are also linear, the accelerate ratio is round 4.5 when have more nanofibers. With the help of Parallel Computing, the PELT method could provide a better performance which can be applied into the case study more efficient.

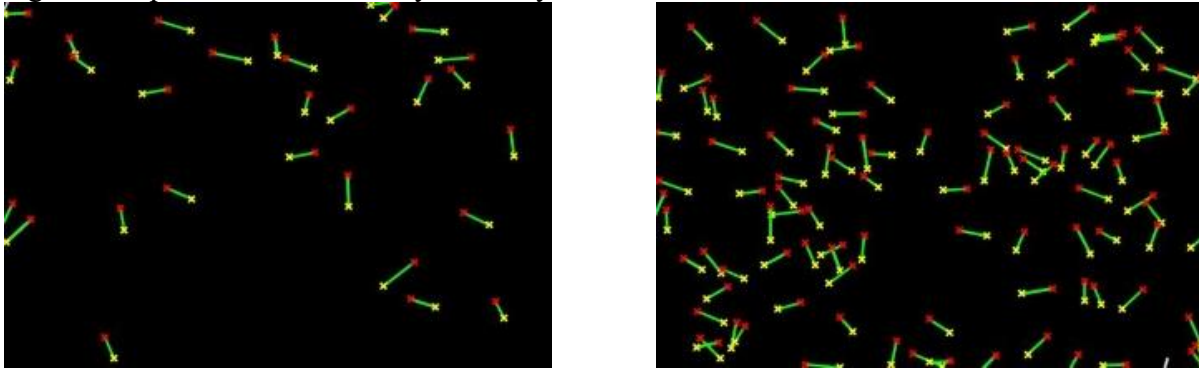
4.3 MATHEMATICS GEOMETRY METHOD

The change points of each connected components have been detected in the last section 4.2, in this section, we would utilize some basic mathematics geometry method to extract the lines through these change points.

As we all know that, in Euclidean space, a straight line is the shortest distance between two points. If three points are not located on one straight line, they would determine a triangle. There's a basic theorem that any summation of two edges is big than the rest edge in this triangle. Hence, we may check the distance between two neighbor change points first, and find the distances

between these two change points with another third change points. If the summation of the two small distances is close to the greatest distance, we may say there may be a line through these three points. For the ideal situation, we may find even pair lines, each pair lines are parallel or having closed orientation. Then we may find the center of these two lines and utilize the longer line's length as the nanofiber's length.

Figure 4.10 shows the case study based on this algorithm using the same simulation SEM image in Chapter 3. The case study accuracy results is in Table 4.7.

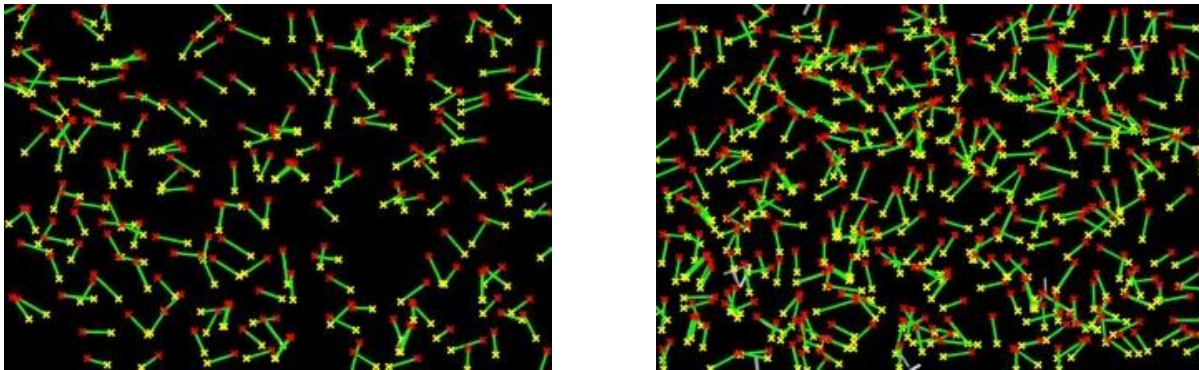


(a)

(b)

Figure 4.10 (a) 30 nanofibers simulation using Boundary based multiple changepoints method

(b) 100 nanofibers simulation using Boundary based multiple changepoints method



(c)

(d)

Figure 4.10 (c) 200 nanofibers simulation using Boundary based multiple changepoints method

(d) 400 nanofibers simulation using Boundary based multiple changepoints method

Table 4.7 Accuracy of simulation case study using Boundary based multiple changepoints method

Number of Nanofibers	Detected nanofibers	Accuracy of the method
30	30	1
100	99	0.99
200	194	0.97
400	384	0.92

The advantage of this algorithm is easy to deploy, nevertheless there could be a severe issue. The algorithm need several parameters to set the thresholds like the angle differences and the distance difference. These paramethers would greatly effect the detection accuracy, especially in the high and extreme high density simulation image. Figure 4.11 shows a typical wrong detected line.

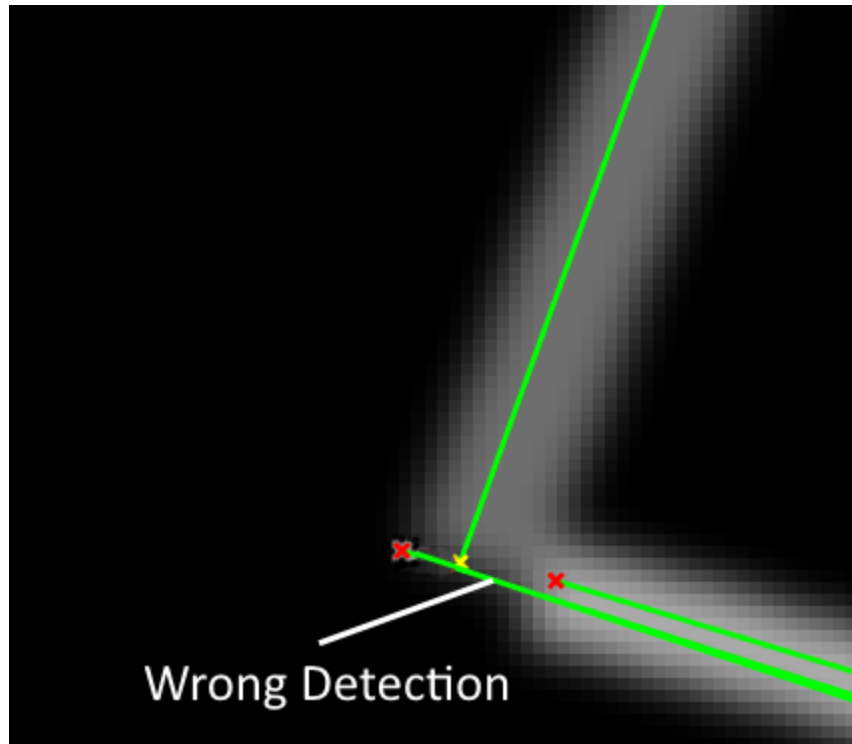


Figure 4.11 An example of one redundant wrong detected line

Chapter 5 Discussion and Conclusion

5.1 RESEARCH WORK TO DATE

In this paper, we propose some methods to extract information like the aligned orientation and lines' information through SEM image of nanofibers. The current research accomplished in this research topic can be summarized as follows:

1) Pattern Mining of aligned nanofibers, because of the importance of the orientation alignment in the piezoelectric nanocomposites, the orientation of the nanofibers is critical in quality control part, this method contain several image processing method to get a quick and easy way to identify the dominant direction of the nanofibers in a SEM image. This method could maintain a high accuracy and robustness in the validation of the simulation test. In the real case, it also provide good performances.

2) Segmentation method based on Partitioning Hough Transform method improve the performance compared with the general simple Hough Transform method in detecting lines in a SEM image. And the Edge based Hough Transform method solves the problems of Partitioning Hough Transform method, it use Edges detection instead of the "Skeleton" operation to get better segmentation results in high density condition of the nanofibers SEM iamges.

3) Statistical Method base on the PELT change point algorithm, it provide a good performance in finding change points. With the "Synthesized-2D" regression model, we overcome the problem in computing the cost function, also with the help of "Parallel computing" technique, it would save huge of waiting time. The next step is to optimize the algorithm in extract the lines throught the change points effectively and meanwhile with high accuracy.

5.2 FUTURE WORK

In this period of research, although several methods aims different problems have been developed, we still facing some problems. The future work will focus on several parts.

First, some methods could be fused to take both advantages to improve the accuracy and efficiency. For example, based on the changepoints detected in each components describe in

Chapter 4 If tune the penalty β , let the PELT method generate more changepoints, is there any way to apply the Hough Transform method to get better results? It should be theoretically.

Second, extend the methods' application. Like Hough Transform could not only detect the simple lines, but also could segment other curves shape. It could be the next step to segment different shapes of nanofibers.

And the last is most important. In the real cases, SEM image of nanocomposites is variaty, it is hard to use one exact method to mining all the information. The nanocomposites may have more different shapes other than the line/rectangle shapes, also, the SEM image may have several complex type, like 3D image or perspective image, these kind of the chanllenges require more strong/effective method to overcome. With the help of machin learning and deep learning technology, it provide a wider way to find the solution we want. Therefore, in future time, the goal of research work will focus on utilize the big data to build flexible intelligence artificial algorithm to fulfillment the needs.

In the meanwhile, new preprocessing of the real case study will be investigated in the future time, a user friendly graphic user interface program is under developing.

References

- [1] Z.-M. Huang, Y.-Z. Zhang, M. Kotaki *et al.*, "A review on polymer nanofibers by electrospinning and their applications in nanocomposites," *Composites science and technology*, vol. 63, no. 15, pp. 2223-2253, 2003.
- [2] Y. Cao, P. C. Irwin, and K. Younsi, "The future of nanodielectrics in the electrical power industry," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 11, no. 5, pp. 797-807, 2004.
- [3] E. Karden, S. Ploumen, B. Fricke *et al.*, "Energy storage devices for future hybrid electric vehicles," *Journal of Power Sources*, vol. 168, no. 1, pp. 2-11, 2007.
- [4] A. Gheibi, R. Bagherzadeh, A. A. Merati *et al.*, "Electrical power generation from piezoelectric electrospun nanofibers membranes: electrospinning parameters optimization and effect of membranes thickness on output electrical voltage," *Journal of Polymer Research*, vol. 21, no. 11, pp. 1-14, 2014.
- [5] L. Persano, C. Dagdeviren, Y. Su *et al.*, "High performance piezoelectric devices based on aligned arrays of nanofibers of poly (vinylidene fluoride-co-trifluoroethylene)," *Nature communications*, vol. 4, pp. 1633, 2013.
- [6] J. Chang, M. Dommer, C. Chang *et al.*, "Piezoelectric nanofibers for energy scavenging applications," *Nano Energy*, vol. 1, no. 3, pp. 356-371, 2012.
- [7] V. Tomer, and C. Randall, "High field dielectric properties of anisotropic polymer-ceramic composites," *Journal of Applied Physics*, vol. 104, no. 7, pp. 074106, 2008.
- [8] C. Bowen, R. Newnham, and C. Randall, "Dielectric properties of dielectrophoretically assembled particulate-polymer composites," *Journal of materials research*, vol. 13, no. 01, pp. 205-210, 1998.
- [9] J. Li, L. Zhang, and S. Ducharme, "Electric energy density of dielectric nanocomposites," *Applied physics letters*, vol. 90, no. 13, pp. 132901, 2007.
- [10] H. Tang, Y. Lin, and H. A. Sodano, "Enhanced energy storage in nanocomposite capacitors through aligned PZT nanowires by uniaxial strain assembly," *Advanced Energy Materials*, vol. 2, no. 4, pp. 469-476, 2012.
- [11] V. Tomer, G. Polizos, C. Randall *et al.*, "Polyethylene nanocomposite dielectrics: implications of nanofiller orientation on high field properties and energy storage," *Journal of Applied Physics*, vol. 109, no. 7, pp. 074113, 2011.
- [12] Y. J. Jeon, H. W. Kang, S. H. Ko *et al.*, "Pattern analysis of aligned nanowires in a microchannel," *Measurement Science and Technology*, vol. 24, no. 3, pp. 035303, 2013.
- [13] R. C. Gonzalez, and R. E. Woods, "Digital image processing," Prentice hall Upper Saddle River, NJ, USA:, 2008.
- [14] J. Serra, *Image analysis and mathematical morphology, v. 1*: Academic press, 1982.
- [15] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection." pp. 465-470.
- [16] J. M. Prewitt, "Object enhancement and extraction," *Picture processing and Psychopictorics*, vol. 10, no. 1, pp. 15-19, 1970.
- [17] I. Sobel, *Camera models and machine perception*, DTIC Document, 1970.
- [18] H. P. VC, "Method and means for recognizing complex patterns," Google Patents, 1962.
- [19] R. O. Duda, and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11-15, 1972.

- [20] S. E. Umbaugh, *Computer imaging: digital image analysis and processing*: CRC press, 2005.
- [21] T. Zhang, and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236-239, 1984.
- [22] H. Lü, and P. S.-P. Wang, "A comment on "a fast parallel algorithm for thinning digital patterns"," *Communications of the ACM*, vol. 29, no. 3, pp. 239-242, 1986.
- [23] A. Grasselli, "Automatic interpretation and classification of images," *New York: Academic Press*, | c1969, edited by Grasselli, A., vol. 1, 1969.
- [24] R. Jain, R. Kasturi, and B. G. Schunck, *Machine vision*: McGraw-Hill New York, 1995.
- [25] F. Neyenssac, "Contrast enhancement using the Laplacian-of-a-Gaussian filter," *CVGIP: Graphical Models and Image Processing*, vol. 55, no. 6, pp. 447-463, 1993.
- [26] D. Marr, and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 207, no. 1167, pp. 187-217, 1980.
- [27] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679-698, 1986.
- [28] I. Cox, R. Boie, and D. Wallach, "Line recognition." pp. 639-645.
- [29] D. Marr, T. Poggio, E. C. Hildreth *et al.*, "A computational theory of human stereo vision," *From the Retina to the Neocortex*, pp. 263-295: Springer, 1991.
- [30] H. Freeman, and R. Shapira, "Determining the minimum-area encasing rectangle for an arbitrary closed curve," *Communications of the ACM*, vol. 18, no. 7, pp. 409-413, 1975.
- [31] P. Annapurna, S. Kothuri, and S. Lukka, "Digit Recognition Using Freeman Chain Code," *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, vol. 2, no. 8, pp. 362-365, 2013.
- [32] H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Transactions on Electronic Computers*, no. 2, pp. 260-268, 1961.
- [33] H. Freeman, "Computer processing of line-drawing images," *ACM Computing Surveys (CSUR)*, vol. 6, no. 1, pp. 57-97, 1974.
- [34] E. Bribiesca, "A new chain code," *Pattern Recognition*, vol. 32, no. 2, pp. 235-251, 1999.
- [35] E. Bribiesca, and A. Guzman, "How to describe pure form and how to measure differences in shapes using shape numbers," *Pattern Recognition*, vol. 12, no. 2, pp. 101-112, 1980.
- [36] E. Bribiesca, "Arithmetic operations among shapes using shape numbers," *Pattern Recognition*, vol. 13, no. 2, pp. 123-137, 1981.
- [37] E. Page, "Control charts for the mean of a normal population," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 131-135, 1954.
- [38] L. S. Nelson, "Column: Technical aids: The Shewhart control chart—tests for special causes," *Journal of quality technology*, vol. 16, no. 4, 1984.
- [39] C. W. Champ, and W. H. Woodall, "Exact results for Shewhart control charts with supplementary runs rules," *Technometrics*, vol. 29, no. 4, pp. 393-399, 1987.
- [40] S. Roberts, "A comparison of some control chart procedures," *Technometrics*, vol. 8, no. 3, pp. 411-430, 1966.
- [41] S. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 42, no. 1, pp. 97-101, 2000.
- [42] J. M. Lucas, and M. S. Saccucci, "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, vol. 32, no. 1, pp. 1-12, 1990.

- [43] W. Bohm, and P. Hackl, "Improved bounds for the average run length of control charts based on finite weighted sums," *The Annals of Statistics*, vol. 18, no. 4, pp. 1895-1899, 1990.
- [44] T. L. Lai, "Control charts based on weighted sums," *The Annals of Statistics*, pp. 134-147, 1974.
- [45] A. G. Gray, and P. J. Thomson, "On a family of finite moving - average trend filters for the ends of series," *Journal of Forecasting*, vol. 21, no. 2, pp. 125-149, 2002.
- [46] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100-115, 1954.
- [47] C. S. v. D. De Bruyn, *Cumulative sum tests: theory and practice*: Griffin, 1968.
- [48] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, pp. 1897-1908, 1971.
- [49] V. M. De Oca, D. R. Jeske, Q. Zhang *et al.*, "A cusum change-point detection algorithm for non-stationary sequences with application to data network surveillance," *Journal of Systems and Software*, vol. 83, no. 7, pp. 1288-1297, 2010.
- [50] S. Lee, J. Ha, O. Na *et al.*, "The cusum test for parameter change in time series models," *Scandinavian Journal of Statistics*, vol. 30, no. 4, pp. 781-796, 2003.
- [51] A. Pettitt, "A non-parametric approach to the change-point problem," *Applied statistics*, pp. 126-135, 1979.
- [52] G. Lorden, "Open-ended tests for Koopman-Darmois families," *The Annals of Statistics*, pp. 633-643, 1973.
- [53] D. Siegmund, and E. Venkatraman, "Using the generalized likelihood ratio statistic for sequential detection of a change-point," *The Annals of Statistics*, pp. 255-271, 1995.
- [54] A. Willsky, and H. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *IEEE Transactions on Automatic control*, vol. 21, no. 1, pp. 108-112, 1976.
- [55] A. S. Willsky, and H. L. Jones, "A generalized likelihood ratio approach to state estimation in linear systems subjects to abrupt changes." pp. 846-853.
- [56] A. Shiryaev, "The problem of the most rapid detection of a disturbance in a stationary process."
- [57] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22-46, 1963.
- [58] A. Shiryaev, "Some exact formulas in a "disorder" problem," *Theory of Probability & Its Applications*, vol. 10, no. 2, pp. 348-354, 1965.
- [59] R. P. Adams, and D. J. MacKay, "Bayesian online changepoint detection," *arXiv preprint arXiv:0710.3742*, 2007.
- [60] C. Erdman, and J. W. Emerson, "A fast Bayesian change point analysis for the segmentation of microarray data," *Bioinformatics*, vol. 24, no. 19, pp. 2143-2148, 2008.
- [61] D. Barry, and J. A. Hartigan, "A Bayesian analysis for change point problems," *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 309-319, 1993.
- [62] A. J. Scott, and M. Knott, "A cluster analysis method for grouping means in the analysis of variance," *Biometrics*, pp. 507-512, 1974.
- [63] A. Sen, and M. S. Srivastava, "On tests for detecting change in mean," *The Annals of Statistics*, pp. 98-108, 1975.
- [64] I. E. Auger, and C. E. Lawrence, "Algorithms for the optimal identification of segment neighborhoods," *Bulletin of mathematical biology*, vol. 51, no. 1, pp. 39-54, 1989.

- [65] B. Jackson, J. D. Scargle, D. Barnes *et al.*, “An algorithm for optimal partitioning of data on an interval,” *IEEE Signal Processing Letters*, vol. 12, no. 2, pp. 105-108, 2005.
- [66] Y.-C. Yao, “Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches,” *The Annals of Statistics*, pp. 1434-1447, 1984.
- [67] R. Killick, P. Fearnhead, and I. Eckley, “Optimal detection of changepoints with a linear computational cost,” *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590-1598, 2012.
- [68] J. E. Bresenham, “Algorithm for computer control of a digital plotter,” *IBM Systems journal*, vol. 4, no. 1, pp. 25-30, 1965.

Vita

Zhonghua Hu was born on April 2, 1983 in Shanghai, China. He got his bachelor degree in Mechanical Engineering on the summer of 2005 from Tongji University, Shanghai, China. In 2012 he received his Master of Science degree in Industrial Engineering from The University of Texas at El Paso. In 2013 he joined the doctoral program in Electrical Computer Engineering at The University of Texas at El Paso.

While pursuing his degree, Dr. Hu worked as a research assistant for the Department of Industrial Manufacturing and System Engineering.

Permanent Address: 5600 Star View Dr, El Paso, Texas, 79912