

Study of Morphology Based Cell Assay by Diffraction Imaging Flow Cytometry

by

Wenhuan Jiang

June 1, 2015

Director of Dissertation: Dr. Xin-Hua Hu

Major Department: Physics

Development of an accurate and label-free method for single cell assay attracts intensive research efforts for its importance to cell biology research and clinical applications. Flow cytometry is one of the most widely used technologies for rapid assay of single cells but existing approaches provide very limited information on cell morphology and require the fluorescence staining. In this dissertation research, we focus our efforts on the quantitative analysis of cell morphology using confocal microscopy based three-dimensional (3D) reconstruction and the exploration of a new approach of flow cytometry through imaging of highly coherent scattered light. The goal of the dissertation research is to develop a new and morphology based approach for rapid cell assay and phenotyping with the polarization diffraction imaging flow cytometry (p-DIFC) platform through investigation of the structure-function relations at the cell level. To achieve this goal, cross-polarized diffraction image pairs have been acquired from single cells excited by a linearly polarized laser beam. Image texture and intensity parameters are extracted with a gray level co-occurrence matrix (GLCM) algorithm to obtain a set of image parameters to quantify the diffraction patterns.

An automated cell classification method has been developed using a Support Vector Machine (SVM) algorithm in the feature space formed by the training data of the cross-polarized diffraction image pairs. We have investigated different types of human lymphocytes and prostate epithelial cells with the confocal imaging and p-DIFC measurements and conducted cell morphology and classification studies. The analysis of 3D morphology parameters among the six types of cells provides, for the first time, the ability to quantitatively evaluate the morphologic differences among these phenotypes and to gain insights on the morphology based classification. It has been further shown that the diffraction image parameters can be mapped into a high-dimensional feature space with the SVM algorithm to obtain the optimized model and yield accurate classifications between Jurkat T cells and Ramos B cells and between the normal and cancerous prostate epithelial cells. Based on these results we conclude that the p-DIFC method has significant potentials to be developed into a rapid and label-free method for cell assay and morphology based classification to discriminate cells of high similarity in their morphology.

Study of Morphology Based Cell Assay by Diffraction Imaging Flow Cytometry

A Dissertation

Presented To the Faculty of the Department of Physics

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy in Biomedical Physics

by

Wenhuan Jiang

June, 2015

© Wenhuan Jiang, 2015

Study of Morphology Based Cell Assay by Diffraction Imaging Flow Cytometry

by

Wenhuan Jiang

APPROVED BY:

DIRECTOR OF
DISSERTATION:

COMMITTEE MEMBER: Xin-Hua Hu, PhD

COMMITTEE MEMBER: John M. Kenney, PhD

COMMITTEE MEMBER: Jun Qing Lu, PhD

COMMITTEE MEMBER: Mark D. Mannie, PhD

CHAIR OF THE DEPARTMENT
OF PHYSICS: Li V. Yang, PhD

DEAN OF THE
GRADUATE SCHOOL: John C. Sutherland, PhD

Paul J. Gemperline, PhD

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Xin-Hua Hu for his tireless guidance and tremendous help during my 5-year study. His strict attitude and the focus on the science lead my way forever. I would like to thank my committee members Dr. Jun Lu, Dr. John Kenneth, Dr. Mark Mannie, and Dr. Li Yang for taking the time to help me and providing feedbacks. I am very thankful for Dr. Li Yang and Ms. Lixue Dong for their extensive helps on providing cell samples and teaching me cell techniques. I thank Dr. Kenneth Jacobs for building the Flow Cytometry System that I can work with and his unlimited support in spirit. I am very thankful for Dr. Douglas Weidner for his helps on confocal imaging and cell sorting.

Finally I want to thank my parents who are 7107.2437 miles away from me and still love me every day. And most gratefully I want to thank my lovely wife for her company in my toughest time and believing I could accomplish this.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	IV
LIST OF TABLES	VIII
LIST OF FIGURES	IX
LIST OF ABBREVIATIONS	XII
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 BACKGROUND	4
2.1 Investigation of cell morphology	4
2.2 Light scattering	5
2.3 Flow cytometry	7
2.4 Diffraction imaging flow cytometry	10
2.5 Cell types in this study	11
2.5.1 PC3 and PCS	11
2.5.2 Jurkat and Ramos	12
2.5.3 CD4+ T lymphocytes and CD8+ T lymphocytes	13
CHAPTER 3 IMAGE ANALYSIS ALGORITHMS	15
3.1 Image preprocessing	15
3.1.1 Different image types in the raw data	15
3.1.2 Image preprocessing by pixel intensity	16

3.1.3	Image preprocessing by stripe patterns	17
3.1.4	Image preprocessing by speckle size	18
3.1.5	Image normalization	20
3.2	The GLCM algorithm for texture analysis.....	21
3.2.1	Analysis of image texture	21
3.2.2	Definition of the GLCM algorithm.....	22
3.2.3	Definition of image parameters	24
3.3	Support Vector Machine algorithm	26
3.3.1	General description of the algorithm	26
3.3.2	LIBSVM	31
CHAPTER 4 EXPERIMENTAL METHODS		34
4.1	Confocal imaging and 3D reconstruction	34
4.1.1	Cell preparation.....	34
4.1.2	Confocal imaging.....	36
4.1.3	3D reconstruction and parameter calculation	43
4.2	Diffraction imaging measurement	45
4.2.1	Optical design	45
4.2.2	Camera control and data acquisition.....	47

4.2.3	Data analysis	50
CHAPTER 5	RESULTS AND DISCUSSION.....	54
5.1	Confocal imaging and 3D reconstruction	54
5.1.1	Morphology study of the Jurkat T and Ramos B cell lines.....	54
5.1.2	Morphology study of the normal and cancerous prostate epithelial cells.....	58
5.1.3	Morphology study of the primary CD4+ and CD8+ T lymphocytes.....	61
5.2	Results of p-DIFC measurement.....	66
5.2.1	p-DIFC measurement of Jurkat T and Ramos B cell lines	68
5.2.2	p-DIFC measurement of the normal and cancerous prostate epithelial cells.....	71
5.3	Discussion.....	78
CHAPTER 6	SUMMARY.....	81
REFERENCES	87
APPENDIX A	DIFFRACTION IMAGE PARAMETERS DEFINATION.....	94
APPENDIX B	PROTOCOL OF DETACHING CELLS	96
APPENDIX C	SPLENOCYTES ISOLATION PROCEDURE	98
APPENDIX D	IMMUNOFLUORESCENCE STAINING PROCEDURE	99
APPENDIX E	PROTOCOL OF STAINING WITH SYTO 61 AND MITOTRACKER ORANGE.....	100
APPENDIX F	CELL COUNTING PROTOCOL.....	102

LIST OF TABLES

Table 2-1: Possible refractive index range of cell components	7
Table 5-1: Morphological parameters of Jurkat and Ramos cell lines	56
Table 5-2: Morphological parameters of PC3 and PCS cells	59
Table 5-3: Morphological parameters of CD4+ and CD8+ T lymphocytes	62
Table 5-4: Experimental parameters and classification results of Jurkat and Ramos cell lines	69
Table 5-5: Experimental parameters and classification results of PC3 and PCS cells	72

LIST OF FIGURES

Fig. 3-1: Pixel values of a 4x4 image	23
Fig. 3-2: A linearly separable hyperplane.....	28
Fig. 3-3: (A). A nonlinearly separable dataset; (B). New dataset after mapping	29
Fig. 4-1: Confocal image stack acquired from a PC3 cell	38
Fig. 4-2: Confocal image stack acquired from a PCS cell.....	38
Fig. 4-3: Confocal image stack acquired from a Jurkat cell	39
Fig. 4-4: Confocal image stack acquired from a Ramos cell.....	39
Fig. 4-5: Confocal image stack acquired from a CD4+ T lymphocyte.....	40
Fig. 4-6: Confocal image stack acquired from a CD8+ T lymphocyte.....	40
Fig. 4-7: Confocal Slice Image acquired from a normal prostate cell in different channels: (A) Image slice in the red channel; (B) Image slice in the red channel; (C) Image slice in red and green channels.	41
Fig. 4-8: User interface of the 3D reconstruction software	43
Fig. 4-9: Top view of the p-DIFC system for acquisition of s- and p-polarized diffraction images	47
Fig. 4-10: Flow diagram of the image acquisition software	48
Fig. 4-11: Flow diagram of data processing in p-DIFC method	51
Fig. 4-12: User interface of parameter analysis software	52
Fig. 4-13: User interface of SVM classification software	53

Fig. 5-1: Perspective views of reconstructed 3D structures of three Ramos (A) and three Jurkat (B) cells	55
Fig. 5-2: Histograms of nucleus-to-cell volume ratio for the Jurkat and Ramos cells	57
Fig. 5-3: Perspective views of reconstructed 3D structures of three PC3 (A) and PCS (B) cells	58
Fig. 5-4: The scatter plots of PC3 and PCS with 4 combinations of 3D parameters: (a) V_c vs S_c ; (b) V_n vs S_n ; (c) $\langle R_c \rangle$ vs SVr_c ; (d) SIi_m vs SVr_n	60
Fig. 5-5: Perspective views of reconstructed 3D structures of three CD4+ (A) and three CD8+ (B) T lymphocytes.	63
Fig. 5-6: Scatter plot of V_c versus S_c for the 6 cell lines or types	64
Fig. 5-7: Scatter plot of V_n versus S_n for the 6 cell lines or types.....	65
Fig. 5-8: Scatter plot of V_m versus S_m for the 6 cell lines or types	65
Fig. 5-9: Selected pairs of polarization-resolved diffraction images acquired from Jurkat and Ramos cells, and PC3 and PCS cells.....	67
Fig. 5-10: 2 D scatter plot and histograms of selected parameters for Ramos and Jurkat cell lines from sequence #1	69
Fig. 5-11: 3D scatter plot of selected parameters for Ramos and Jurkat cell lines from sequence #1	70
Fig. 5-12: The scatter plots of training data with values of classifier F versus the top two ranked GLCM parameters used by the best SVM model	71

Fig. 5-13: 2D scatter plot and histograms of selected parameters for PC3 and PCS cells from sequence #1.	73
Fig. 5-14: 3D scatter plot of selected parameters for PC3 and PCS cells from sequence #1. The polarization of incident beam is vertical.....	74
Fig. 5-15: The scatter plots of training data with values of classifier F versus the top two ranked GLCM parameters used by the best SVM model	75
Fig. 5-16: The averaged accuracy A_{av} versus the maximum number of image parameters in a feature vector N_{max} for Jurkat and Ramos classification	76
Fig. 5-17: The averaged accuracy A_{av} versus the maximum number of image parameters in a feature vector N_{max} for PC3 and PCS classification	77

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
ACC	Accuracy
API	Application Program Interface
ASM	Angular Second Moment
CIP	Intact cell image pair
CLP	Cluster prominence
CLS	Cluster shade
CON	Contrast
COR	Correlation
DDA	Discrete-dipole-approximation
DIFC	Diffraction Imaging Flow Cytometry
EM	Electromagnetic
ENT	Entropy
FCM	Flow Cytometry
FDTD	Finite-Difference-Time-Domain
FSC	Forward scatter
GLCM	Gray level co-occurrence matrix
IDM	Inverse Difference Moment

IMAX	Normalized maximum pixel intensity
IMIN	Normalized minimum pixel intensity
LIP	Large speckle image pair
LSCM	Laser scanning confocal microscope
OIP	Overexposed image pair
OS	Operating system
PMT	Photomultiplier
p-DIFC	polarization Diffraction Imaging Flow Cytometry
SCC	Side scatter
SIP	Stripe image pair
SVM	Support Vector Machine
UIP	Underexposed image pair
VAR	Variance
WBCs	White blood cells

CHAPTER 1 INTRODUCTION

Flow Cytometry (FCM) is a widely used method for rapid assay of single cells by measurement of the physical and chemical characteristics of the cells. With the FCM method, cellular components or organelles are labeled with fluorescent reagents which emit light signals at wavelengths longer than those of excitation. Both of the scattered and fluorescent light signals from the cells are measured in FCM in the form of pulsed signals [1], which are used to analyze and classify cells. The scattered light signals, including the forward scatter (FSC) signal and side scatter (SSC) signal, indicate the cell volume and degree of heterogeneity in intracellular distribution of the refractive index. The two light scatter signals provide useful information for cell assay according to their morphology but cannot yield sufficient details on the complex cellular morphology. As a result, multiple fluorescent probes or labels have to be used to obtain molecular information of the measured cells for various applications such as classifying subtypes of white blood cells for immunotherapy of cancers [2] and detection of circulating tumor cells in blood [3]. Furthermore, cells in different conditions undergo significant and characteristic changes in the structures of intracellular organelles such as cytoplasm and nucleus. Therefore, the investigations of cellular morphology with minimum or without extraneous interferences by fluorescent staining [4, 5] are highly desired. It is thus important to develop label-free and morphology

based cell assay approaches that are complementary to the existing FCM methods and useful for cell biology research and clinical development.

This dissertation research project focuses on the analysis of different cell types including human lymphocytes and epithelial cells using the methods of polarization diffraction imaging flow cytometry (p-DIFC) and confocal microscopy based three-dimensional (3D) morphology methods developed in the Biomedical Laser Laboratory. The long-term goal of the research project is to develop a new and morphology based approach for rapid cell assay and phenotyping using the p-DIFC platform by exploring the structure-function relation at the cell level. To achieve this goal, the existing p-DIFC system has been improved with in-house developed software on diffraction image acquisition, processing and extraction of feature parameters. Different types of primary and cultured cells have been investigated quantitatively on their 3D morphology to quantify the structural differences in major intracellular organelles of cytoplasm, nucleus, and mitochondria that play significant roles in the light scattering. The cross-polarized diffraction images have been acquired from the above cell samples with the improved p-DIFC system. We have applied the gray level co-occurrence matrix (GLCM) algorithm to quantify the diffraction patterns of the imaged cells in terms of the texture parameters extracted from the diffraction image data. Automated cell classification has been performed using a Support Vector Machine (SVM) algorithm in the parameter space formed by the diffraction image parameters.

Through this dissertation study, we have investigated two types of primary prostate epithelial and cancer cell lines, the Jurkat T and Ramos B cancer cell lines and primary

lymphocytes extracted from human spleen tissues. The 3D morphology of these cell types have been quantified with 29 parameters for comparison and analysis of the statistical significance in their differences. The cross-polarized diffraction image data have been acquired and analyzed to examine the capability of the p-DIFC method for label-free cell assay and classification. The results of this study demonstrate the capabilities of p-DIFC method to rapidly measure and analyze the diffraction image data for accurate label-free classification of cells of high similarity in morphology.

This dissertation is organized into the following chapters. Chapter 2 provides the background information of light scattering by single biological cells, cell morphology, phenotyping, and the FCM methods. Chapter 3 presents the various algorithms employed for processing the acquired diffraction image data, analysis of diffraction image texture, and the machine learning tools for cell classification. Chapter 4 describes the experimental methods for cell extraction, preparation, confocal imaging measurement, and p-DIFC measurements. In Chapter 5 the result of 3D morphology and p-DIFC based classification studies are presented on different cell lines and primary cells. We discuss in Chapter 6 the significance of the dissertation study and directions for future research.

CHAPTER 2 BACKGROUND

This chapter provides the background information on the theoretical and experimental methods to study single biological cells and cell morphology from the literature. In particular, we present a method comparison of the conventional flow cytometry and diffraction imaging flow cytometry which has been employed as a major experimental means in this dissertation study.

2.1 Investigation of cell morphology

Biological cells were first discovered through the structural examination by Robert Hooke with a primitive microscope of low magnification more than three centuries ago [6] . Since then it has been well established that cells are the basic units of life whose functions are closely related to their structures. An organism can contain numerous and specialized types of cells that differ both in their structures or morphology and functions. These cells can be classified either by phenotyping through their morphology based structural analysis or genotyping. Genotyping is the process of determining differences in the genetic makeup or genotype [7] of concerned cells by examining their DNA and RNA sequences using various assays and comparing with other cells' sequences. However, the identification of cell types through genotyping provides only molecular aspects of cell information, which fundamentally relates to the structural or morphologic aspect but not equivalent. Furthermore, genotyping requires analysis of DNA and RNA sequences and expression levels of proteins,

which can be time-consuming and difficult to be implemented at the single cell level. Unlike genotyping, phenotyping through morphology distinguishes cell types based on structural characteristics or traits to support specific cellular functions and/or developmental behaviors. Since morphological information cannot be extracted in general from the genotyping data, it is highly desired to develop rapid morphology based phenotyping tools that allow rapid and accurate extraction of the structural information of single cells on the basis of the well-recognized structure-function relations.

2.2 Light scattering

Light scattering by single particles occurs as a result of the variations of refractive index from the host medium and within the particle. When a light beam in a form of electromagnetic (EM) waves strikes a particle, the excited molecules inside the particle can be seen as a phased array of many induced electric dipoles. If the incident light beam is highly coherent, the induced dipoles within the illuminated particle radiate EM waves of scattered light that are highly coherent as well. The spatial distribution of the scattered light in this case thus presents characteristic diffraction patterns as the results of coherent superposition of the highly coherent EM fields of scattered light by the induced molecular dipoles within the particle. The intensity of scattered light along a particular direction of θ_s as the polar scattering and ϕ_s as the azimuthal scattering angle can be measured by one or multiple imaging sensors in space, and can be simulated by solving the Maxwell equations with the intracellular distribution of refractive index $n(\mathbf{r})$ with \mathbf{r} as the position vector.

The coherent distribution of light scatter is a direct consequence of superposition among the EM fields emitted by the induced molecular dipoles inside the illuminated particle. The characteristic patterns in space of the scattered light are thus highly correlated to the structure of $n(\mathbf{r})$ of the scatter's morphology. This correlation could thus provide a foundation to extract morphology information of the scatter and establish a rapid method of cell assay based on morphology through angle-resolved imaging of the scattered light. The measurement of the coherent distribution of scattered light with imaging sensors has been developed in our lab and is termed as diffraction imaging in this dissertation research. Compared with traditional non-coherent imaging approaches for the morphology measurement such as microscopy, diffraction imaging yields a new and different means to characterize morphology. As the new method, diffraction imaging of the scattered light allows probing of the 3D structure of the coherently illuminated particle or cell.

Biological cells are the typical form of the 3D particles that composed of cytoplasm and various organelles such as nucleus and mitochondria. They scatter light because of their heterogeneous distribution of the refractive index inside the cells that is different from the index of the host medium. We can find a wide range of values for refractive indices of cell components in the literature [8,9,10] as shown in Table 2-1. The range of cell membranes is totally different from the cytoplasm and the nucleus. The nucleus and the mitochondria are expected to be in the following ranges even though the exact values remain unknown for almost all cell types.

Table 2-1 Possible refractive index range of cell components

Cell Component	Range of refractive index
Cell membranes	1.46-1.54
Cytoplasm	1.35-1.37
Nucleus	1.38-1.41
Mitochondria	1.38-1.41

Elastically scattered light from a cell illuminated with a highly coherent laser beam presents characteristic and intriguing diffraction patterns in space as a result of the index heterogeneity and high coherence among the wave fields of the scattered light. The correlation between the spatial distribution of the coherent light scattered by a cell and the intracellular distribution of the refractive index shows the possibility for 3D morphology based analysis and phenotyping of single cells without fluorescent staining, which can provide a powerful tool for cell biology applications in many fields of the life science research.

2.3 Flow cytometry

The study of single cells often requires measurements on a large number of cells to obtain statistically meaningful data and therefore FCM is the ideal candidate to measure a great amount of cells' information in a high speed. FCM technology analyzes the physical and chemical characteristics of single particles such as biological cells in a fluid as they pass through the laser beam. The speed of light scatters and fluorescence detections is very fast with throughput reaching up to 10,000 cells/s. Components in cells are labeled fluorescently

and emit light signals at varying wavelengths after exciting by the laser. The detectors of FCM measure scattered and fluorescence light signals in the form of pulsed signals [1], which are then sorted by a multichannel analyzer, displaying histograms of the number of cells having a specific pulse height versus the values of the height. Main components of FCM include: a fluidic device, an optical system, and a data analysis system. A fluidic device converts a fast moving cell suspension into only one-cell wide. An optical system is composed of one or more laser beams for the excitation of the cells in the flow and detectors to acquire the scattered and fluorescence light signals. The data analysis system analyzes the data in histograms or scatter plots for cell classification even cell sorting.

Historically, the first automatic measurement and analysis of single cells were carried out with the Coulter counter in the early 1940's [11]. As blood cells pass through a capillary, they changed the electric impedance between intercepting electrodes which can be measured for counting the number of cells. In 1950s cell population on a microscope slide were interrogated with a micro spectrophotometer at UV wavelengths, which was a precursor to fluorescence detectors of a cell stream [12]. Some researchers tried to identify cells automatically based on their nuclear sizes and indices of refractions as biomarkers for cell malignancy [13]. As the development of computing technology, further efforts were made to process a great amount of cell data. At the same time, Acridine Orange was firstly used for fluorescence staining to generate and analyze RNA concentration in the cell [14]. In 1960s, some researchers attempted to use a laser scanning microscopy to extract cellular features automatically such as the size of cell and nuclear as well as the texture pattern of the acquired

images by the pattern recognition analysis. After that, vibrating mirrors were firstly used to scan a laser beam and extract cell features such as the size of the cell and the size of the nucleus according to the variation of dye's concentration on the image texture. In addition, this provided an insight of extracting image patterns by the intracellular variation in the refractive index. In these capacities, it usually took two minutes or more time to scan and analyze a single cell with a laser scanning cytometer [1]. Then another group made an effort to analyze large population of cells through the flow cytometer by the forward light scatter signals to differentiate white blood cells [11]. In the early 1970s a research group in the Los Alamos National Laboratory applied the ink jet principle for cell sorting [15]. They charged droplets containing cells and steered them based on results of the optical interrogation. After 1970s, the major development of FCM has been the detection of fluorescence signals for analysis of cellular functions. Most FCM systems now acquire both fluorescent and scatter signals for cell assay with very limited morphology information [16]. The forward scatter (FSC) signal and side scatter (SC) signal only indicate the cell volume and the degree of heterogeneity in intracellular distribution of the refractive index, which provide limited information for cell classification and quantifying the morphology changes. Incorporation of cell imaging into the fast FCM technology platform attracts long-standing interests for its potential to peek into and extract morphological features in additional to the molecular information [17]. Non-coherent imaging FCM is commercially available now [18]. But several shortcomings exist such as the requirement of fluorescent staining and the difficulty to automate the image processing in real time because of the localized, highly varied and

convoluted structural elements in the fluorescence images. Several groups have investigated imaging of coherently scattered light with flow cytometric or microfluidic devices. But the images they have reported are of poor contrast that are difficult to be analyzed accurately [19,20].

2.4 Diffraction imaging flow cytometry

Since 2005 the Biomedical Laser Laboratory at East Carolina University has developed a method of diffraction imaging flow cytometry (DIFC) to record the scattered light as high-contrast images from single cells excited by a coherent laser beam[21,22,23,24]. As the result of coherent superposition of scattered wavefields in space, diffraction images from single cells present non-local patterns and rarely contain segmental structures, which enable automated image processing in real time. Furthermore, coherent scattering by cells are modeled accurately by the classical electrodynamic or wave theory and thus provide opportunity to compare the theoretical modeling results of diffraction imaging with experimental data to gain insights. Since 2013, the DIFC method has been improved and renamed as polarized- Diffraction Imaging Flow Cytometry (p-DIFC). The new method can simultaneously acquire two polarized diffraction images by two CCD cameras for each cell, which enables for the first time extraction of information from polarized light scattering patterns for detailed analysis of the cells in an imaging flow cytometer.

2.5 Cell types in this study

We investigate the p-DIFC method with different types of cells including three cell lines and three primary cell types. A cell line consists of transformed and immortal cells derived typically from a single cancer cell that can be maintained in culture medium. Therefore the cells in a cell line have a uniform genetic makeup and a low degree of heterogeneity in morphology variations. After the cells are isolated from the tissue, they proliferate under appropriate conditions and occupy all of the available substrate. This stage is called primary culture and cells have to be subcultured by transferring them to a new vessel with fresh growth medium to provide more room for continued growth. After the subculture, the primary culture becomes a cell line or subclone. Cell lines usually have a limited life span and as they are passaged, cells with the highest growth capacity predominate, resulting in a high degree of genotypic and phenotypic uniformity in the population, which provides a perfect starting point to test our p-DIFC method. In our study, the three cell lines of PC3, Jurkat, and Ramos cells were utilized. On the other hand, primary cells are directly from an animal or plant without transformation to cell lines. Primary cells reveal the physiological state of cells in vivo and provide more relevant data representing the living system. The primary cells in this dissertation are CD4⁺ T lymphocytes and CD8⁺ T lymphocytes from human spleen tissue.

2.5.1 PC3 and PCS

Prostate cells are typical epithelial cells from the prostate, which is a walnut-sized gland located between the bladder and the penis and produces the seminal fluid that nourishes and

transports sperm for males. Prostate cancer is one of the most common types of cancer in US. It usually grows slowly and initially remains confined to the prostate gland, where it may not cause serious harm. While some types of prostate cancer grow slowly and may need minimal or no treatment, other types are aggressive and can spread quickly. Prostate cancer that is detected early, when it's still confined to the prostate gland, has a better chance of a successful treatment.

Rapidly quantification and measurement of the morphological differences in normal prostate cells and their cancerous counterparts is a long-term challenge, which can be developed as a rapid method of cell assay for early detection of prostate cancers through testing blood or urine samples. We select PCS as normal prostate cells and PC3 as prostate cancer cells to test our p-DIFC method. PCS (PCS-440-010, ATCC) cells are prostate epithelial cells and usually used as a control group for the study of prostate cancer. PC3 (CRL-1435, ATCC) cells are cancer cell lines that established in 1979 from bone metastasis of grade IV of prostate cancer in a 62-year-old male [25]. A morphology based classification of PCS and PC3 cells may yield a significant insight on the morphological differences between malignant and “normal” cells of common prostate lineages for study of prostate cancers.

2.5.2 Jurkat and Ramos

White blood cells (WBCs) are the cells of the immune system that protect the body against both infectious diseases and foreign invaders. All WBCs are produced and derived from the multipotential hematopoietic stem cells from bone marrow . WBCs are found throughout

the body including the lymphatic system and the blood [26]. There are five different types of WBCs: Neutrophils, Eosinophils, Basophils, Lymphocytes, and Monocytes. And they can be distinguished by their physical and functional characteristics. The number of WBCs in the blood is often an indicator of disease. The normal white cell count is usually between 4 and $11 \times 10^9/L$. In the US this is usually expressed as four thousands to eleven thousands white blood cells per μl of blood [27]. They make up approximately 1% of the total blood volume in a healthy adult.

We analyze WBCs by investigating the differences in morphology and light scattering among various lymphocytes including subtypes of T and B cells that have been widely deemed as morphologically indistinguishable [16]. At the beginning of my dissertation study, we have studied cancer cell lines Jurkat and Ramos. The Jurkat cell line (TIB-152, ATTC) was established in the late 1970s from the peripheral blood of a 14-year-old boy with T cell leukemia [28]. The Ramos cell line (CRL-1596, ATTC) was derived from the B lymphocytes of a 3-year-old boy who had Burkitt's lymphoma.

2.5.3 CD4+ T lymphocytes and CD8+ T lymphocytes

T lymphocytes are very important cell types of mammalian immune responses to pathogens and abnormal cells. T lymphocytes can be further divided into subtypes through immunophenotyping with fluorochrome-conjugated CD surface markers which have been shown to play different and critical roles in immune responses. For example, CD4+ T lymphocytes can be activated to become T regulatory (T_{Reg}) cells. CD8+ T lymphocytes can become T cytotoxic (T_C) cells, which kills infected and cancer cells carrying antigens or

mutations. In contrast, the T_{Reg} cells perform regulatory roles by suppressing immune responses to self-antigens. Clinical studies of various T cell subtypes in cancer patients have shown that cancer patients tend to have higher ratios of CD8+/CD4+ T cells [29, 30]. Despite the fact that T lymphocytes can be routinely classified into subtypes using CD markers, development of label-free and rapid method to distinguish these primary cells can yield a powerful tool for study of lymphocytes and other white blood cells in immunology and other fields such as immunotherapy of cancer patients [31]. Besides its clinical implications, the investigation of cellular structures among the different T lymphocyte subtypes provides insights on the fundamental relations between structure and function at the cell level.

CHAPTER 3 IMAGE ANALYSIS ALGORITHMS

The experimental data acquired through this dissertation research are presented in the form of confocal microscopy and diffraction images. Analysis of these image data and subsequent cell classification study require multiple image processing and data mining algorithms which have been developed over the last few decades. In this chapter, we present the principles of these algorithms and their applications in our investigations.

3.1 Image preprocessing

3.1.1 Different image types in the raw data

A cell suspension sample contains various types of particles other than intact cells such as cellular debris and solid particles or aggregated particles formed inside the cell culture medium. Furthermore, due to the variations of experimental conditions such as the core fluid position relative to the focus of the incident laser beam, the raw diffraction images acquired by our p-DIFC system can become underexposed or overexposed. Thus the diffraction image data acquired by the p-DIFC system requires preprocessing before analysis of the image textures for cell assay. For this purpose, an image preprocessing software has been developed to classify the raw image pair data into five types of underexposed image pair (UIP), overexposed image pair (OIP), large speckle image pair (LIP) , stripe image pair (SIP) and intact cell image pair (CIP). Through previous experimental and modeling studies, we found

that the diffraction patterns in CIP can be attributed to the coherent superposition of EM fields of light scattered from intact biological cells of sizes 5 μm or larger [21,22,23].

Therefore, the CIPs provide the image data we need for subsequent image analysis and cell classification. Contrary to CIPs, LIPs are the results of coherent light scattering by small-sized particles or cell debris with dimensions typically less than 3 μm while the SIPs are formed by the coherent light scattered by particles of highly symmetric morphology such as microspheres. The other two types of image pairs, UIPs and OIPs, are dominated by noises which need to be removed together with LIPs and SIPs. The four types of images other than the CIP are filtered out by a preprocessing software based on the following algorithms.

3.1.2 Image preprocessing by pixel intensity

The image preprocessing software reads the raw diffraction images $J(z, y)$ as the input data with the z-axis representing the horizontal direction or the incident beam direction and y-axis representing the vertical or the flow direction of the imaged particle. Multiple pixel intensity parameters are obtained as the values of minimum pixel intensity J_{min} , maximum pixel intensity J_{max} and average pixel intensity J_{av} from each $J(z, y)$ in the 12-bit cross-polarized diffraction image pair data. An image pair is marked as UIP if the values of J_{av} for both images in the pair are smaller than 80 or less than 2% of the 12-bit pixel saturation value J_{sat} at 4095. If the total number of saturated pixels J_{sat} of one image in an image pair is more than 3000 or more than 1% of the total number of pixels in the image, the image pair is marked as OIP.

3.1.3 Image preprocessing by stripe patterns

Different from the identification of UIP or OIP, it is much more complex to select and remove the SIP by single or aggregated spheres or spherical particles with a homogeneous internal structure in terms of the refractive index distribution [21, 24] and LIP by small sized cellular objects such as debris [45]. The details of the algorithm are provided below.

The image preprocessing software [45] calculates firstly the average pixel intensity of an image in a pair as the threshold to convert the normalized diffraction image $J(z, y)$ into a binary one as $B(z, y)$. After that, four Sobel operators shown below for edge detection are applied to obtain the speckle borderlines or edges along four directions of horizontal (h), vertical (v), left diagonal (l) and right diagonal (r) respectively [47]:

$$G_h = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}, G_v = \begin{pmatrix} -1 & 0 & -1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, G_l = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix}, G_r = \begin{pmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

After convoluting the four operators with $B(z, y)$, a set of four directional edge images $E_a(z, y)$ can be derived, where $a=(h, v, l, r)$. A complete edge image $E_T(z, y)$ can be obtained by summing the four directional images. The maximum intensity for borderlines is set to 1 and all other pixels are set to 0. We can sum quickly for a set of 5 length parameters as $[C_h, C_v, C_l, C_r]$ and C_T , which help us separate those images of the stripe patterns from the other two types of speckle patterns by comparing $[C_h, C_v]$ or $[C_l, C_r]$ in a pair of edge images. $[C_h, C_v]$ or $[C_l, C_r]$ are mutually perpendicular. If $C_l < C_{th}$ and $C_l < 0.3C_2$ (C_l is the lesser of the two C parameters in the pair and $C_{th}=2500$ is the threshold), the diffraction image is marked as the

stripe with the direction along (or approximately along) the direction of the edge image with C_2 .

3.1.4 Image preprocessing by speckle size

To identify the image pairs by cell debris or small particles, the image preprocessing software extracts additional parameters from $J(z, y)$ in frequency space for accurate grouping the rest of the diffraction images according to the speckle sizes in the real space. $J(z, y)$ is mapped in to the frequency space (u, v) with a 2D fast Fourier transform (FFT) as

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J(z, y) \exp\{-2\pi i (uz + vy)\} dz dy. \quad (1)$$

A power spectrum image can be obtained as $P(u, v) = |F(u, v)|$. Analysis of our diffraction images from previous experiments led to the conclusion that the images of large speckle patterns often contain bright spots of linear sizes of 150 or more pixels [45]. Therefore we choose the frequency threshold as

$$f_{th} = \frac{1}{150(\Delta)} = 0.00667 \left(\frac{1}{\Delta}\right), \quad (2)$$

where Δ is the inter-pixel distance. After getting f_{th} , we can derive a histogram $N(f)$ of high frequency pixels in $P(u, v)$ with $f = (u^2 + v^2)^{1/2}$. N is the number of pixels with $f > f_{th}$ and $P(u, v) > 0.02P(0,0)$. The sum of $N(f)$ yields the number of pixels with high power and frequency N_P in the power spectrum image $P(u, v)$.

In our previous study, we also found that diffraction images with the normal speckle pattern tend to have C_T and N_P values larger than images with large speckles [45]. Moreover the

presence of the noise due to spurious light and the significant morphological variations of the imaged objects can lead to fluctuations in the absolute values of C_T and N_P as well as their relative differences among samples of different cells. The software successfully calibrates and minimizes the effect of these fluctuations in an automated procedure by using k-means clustering technique to rank and scale the extracted parameters of the given data set. The k-means clustering technique separates all dots into k groups according to their distances to k centers under appropriate conditions [48]. If a dot is closer to one center than others, it is assigned into the group represented by that center. The software utilizes two 2D scatter plots of $A_s(N_P, C_T)$ and $A_p(N_P, C_T)$ on the N_P - C_T plane. The global patterns as characterized by N_P and C_T correlate strongly with the categorical sizes of the imaged objects.

To divide the image pairs into two patterns with $k=2$, we initially assign two centers as (N_{Pai}, C_{Tai}) , where $a=s$ or $a=p$ for each of the two plots. The initial centers are updated by the averaged values of N_P and C_T in each plot and it iterates until the two centers converged to the final values of (N_{Paf}, C_{Taf}) . Minor changes [45] have been made for the standard k-means clustering technique to reduce the effect of fluctuation in N_P and C_T . And the SVM analysis for each imaged object using a classification vector of four parameters $(N_{P1}, C_{T1}, N_{P2}, C_{T2})$ is also performed after correct ranking. We will introduce the SVM algorithm in section 3.3. A scaling method is used in the software to further eliminate the training repeatedly for different data sets. Four parameters extracted from a cross-polarized diffraction image pair in the new dataset are scaled as follows

$$N_{P_i}' = \frac{N_{P_{if}} |_{ref}}{N_{P_{if}}} N_{P_i} \quad \text{and} \quad C_{T_i}' = \frac{C_{T_{if}} |_{ref}}{C_{T_{if}}} C_{T_i}, \quad (3)$$

where $N_{P_{if}}$ and $C_{T_{if}}$ are the averaged parameters obtained by the k-means clustering analysis on the new data set, $N_{P_{if}} |_{ref}$ and $C_{T_{if}} |_{ref}$ are values of the reference data set, and $i=1$ or 2 . At last, the software automatically classifies the image objects into two types of speckle patterns after removal of those with the stripe patterns using the calibrated parameters of $(N_{P_1}', C_{T_1}', N_{P_2}', C_{T_2}')$. The averaged value of accuracy for this preprocessing with our software was found to be 97.1% [45].

3.1.5 Image normalization

After identification and removal of OIP, UIP, SIP and LIP images, the rest of the raw 12-bit images $J(z, y)$ are normalized into 8-bit images $I(z, y)$ as the input data to a GLCM algorithm based image processing software for the subsequent texture analysis. A linear transformation of the pixel intensity given below is used to obtain the normalized images:

$$I(z, y) = \frac{J(z, y) - J_{\min}}{J_{\max} - J_{\min}} \times 255. \quad (4)$$

With this transformation the minimum and maximum pixel intensities in the 12-bit image $J(z, y)$ are set to 0 and 255 in the 8-bit image $I(z, y)$. The normalization and bit reduction are necessary to remove the dependence of the image textures on the power of the incident laser beam and speed up the calculation of texture parameters with the GLCM algorithm without significant loss of the dynamic range.

3.2 The GLCM algorithm for texture analysis

3.2.1 Analysis of image texture

Diffraction images present diffraction patterns of coherent light as image texture. It is difficult to define a set of meaningful features to quantify image texture information from the block of resolution pixels. As long as these features are defined, image blocks can be categorized using pattern-recognition techniques. When we search for meaningful features for describing pictorial information, there is an instinct to focus on types of features that can interpret pictorial information. Spectral, textural, and contextual are three fundamental pattern elements among features used in interpretation of images.

- a. **Spectral features** describe the average tonal variations in various bands of the visible and infrared portions in an electromagnetic spectrum.
- b. **Contextual features** contain information derived from blocks of image data surrounding the area being analyzed.
- c. **Textural features** contain information of the spatial distribution of tonal (denoted here as pixel values or as gray-levels) variations within a band, where tone is based on the varying shades of gray of resolution cells in a photographic [32]. Texture and tone are not independent features and they have an inextricable relationship to one another. The texture is the only one concerned with the spatial distribution (statistical) of gray tones. Texture is also an innate property of a diffraction image and contains vital information about the structure of the scatters and their response to the wavefields of the incident light in our case. Since the textural properties of images carry useful information especially for discrimination of the biological cells, we focus our image analysis effort on image texture for our study on cell classification.

3.2.2 Definition of the GLCM algorithm

Haralick introduced a method of gray level co-occurrence matrix (GLCM) in 1970s to quantify image texture [33]. Since then, the GLCM algorithm has been widely used as a texture characterization tool in various fields such as analysis of satellite images [49] and identification of abnormal tissues in clinical images [50]. The GLCM algorithm can be described as a method to extract second order statistical parameters from a monochromatic or gray-level image. In GLCM, $L_x = \{1, 2, \dots, N_x\}$ is the horizontal spatial domain, $L_y = \{1, 2, \dots, N_y\}$ is the vertical spatial domain, and $G = \{1, 2, \dots, N_g\}$ is the set of quantized gray levels N_g ($G=255$ for the 8-bit gray level image). The set $L_y \times L_x$ is the set of pixels of the image ordered by their row-column designations. An input image I can be regarded as a function that assigns some gray level in G to each pixel in $L_y \times L_x$. It is assumed that the texture information in an image I is contained in the overall or average spatial relationship. Let's denote $p(i, j, \mathbf{d})$ as the "co-occurrence" frequency of two neighboring pixels that are separated by the displacement vector $\mathbf{d} = (d, \theta)$ with one pixel of gray level i and the other of gray level j . The frequencies of 4 angles ($\theta=0, 45, 90,$ and 135 degree) are defined by

$$p(i, j, d, 0^\circ) = \#\{((k, l), (m, n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid k - m = 0, |l - n| = d, I(k, l) = i, I(m, n) = j\}$$

$$p(i, j, d, 45^\circ) = \#\{((k, l), (m, n)) \in (L_y \times L_x) \times (L_y \times L_x) \mid \begin{array}{l} (k - m = d, l - n = -d) \\ \text{or} (k - m = -d, l - n = d) \\ I(k, l) = i, I(m, n) = j \end{array}\}$$

$$p(i, j, d, 90^\circ) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid |k - m| = d, l - n = 0, I(k, l) = i, I(m, n) = j\}$$

$$p(i, j, d, 135^\circ) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid \begin{array}{l} (k - m = d, l - n = -d) \\ \text{or } (k - m = -d, l - n = d) \end{array} \mid I(k, l) = i, I(m, n) = j\} \quad , \quad (5)$$

where # is the number of pixels in the set. A gray level co-occurrence matrix P can be obtained with the elements as frequency $p(i, j, \mathbf{d})$. It is easy to show that the matrix is symmetric since $p(i, j, \mathbf{d}) = p(j, i, \mathbf{d})$ and depends on the choice of \mathbf{d} .

For example, if we have a 4x4 image I with pixel values ranging from 0 to 3 as shown in Fig. 3-1:

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Fig.3-1 Pixel values of a 4x4 image

The frequencies of the gray-level co-occurrence can be found for the horizontal direction ($\theta=0$) and $d=1$:

$$p(0,0,1,0^\circ) = 4, \quad p(0,1,1,0^\circ) = 2, \quad p(0,2,1,0^\circ) = 1, \quad p(0,3,1,0^\circ) = 0,$$

$$p(1,0,1,0^\circ) = 2, \quad p(1,1,1,0^\circ) = 4, \quad p(1,2,1,0^\circ) = 0, \quad p(1,3,1,0^\circ) = 0,$$

$$p(2,0,1,0^\circ) = 1, \quad p(2,1,1,0^\circ) = 0, \quad p(2,2,1,0^\circ) = 6, \quad p(2,3,1,0^\circ) = 1,$$

$$p(3,0,1,0^\circ) = 0, \quad p(3,1,1,0^\circ) = 0, \quad p(3,2,1,0^\circ) = 1, \quad p(3,3,1,0^\circ) = 2.$$

With those frequencies, the GLCM of the input image I can be represented by a matrix P_H for the horizontal direction as:

$$P_H = \begin{pmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

The GLCM matrix for 45-degree direction P_{RD} , vertical direction P_V , and 135-degree direction P_{LD} are obtained as well:

$$P_{RD} = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix};$$

$$P_V = \begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix};$$

$$P_{LD} = \begin{pmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}.$$

3.2.3 Definition of image parameters

The GLCM is usually expressed with normalized frequencies as $p(i,j)=P(i,j)/R$, where R is the total pair of the neighboring pixels for calculating the matrix P . The GLCM based image processing software reads the normalized 8-bit diffraction image $I(z, y)$ and calculates the corresponding GLCM as $p(i, j)$ of $I(z, y)$, where $i=\{0,1,2,\dots, G-1\}$, $j=\{0,1,2,\dots,G-1\}$, and $G=255$ for the 8-bit gray level images. In our study, a total of 19 parameters have been extracted for each of the cross-polarized diffraction image pair which include 17 texture parameters defined through $p(i, j)$ and 2 parameters of maximum pixel and minimum pixels from the 12-bit image $J(z, y)$. Therefore, each cross-polarized image pair yields 38 parameters to represent each imaged cell by the p-DIFC method. The definitions of the 19 parameters

extracted from a diffraction image are provided in Appendix A. We select a few parameters here to discuss their geometric implications.

Correlation (COR) is a measure of the linear dependency of gray levels on those of neighboring pixels or specified points, and it is an indicator of local gray-level dependency on the texture image. Higher values can be obtained for similar gray-level regions.

Contrast (CON) is a measure of the local variations presented in an image. It is highly correlated with the difference between the highest and the lowest values of a continuous set of pixels particularly when the value of the displacement vector is 1[35]. The contrast will be very high if there is a large amount of variation in an image.

Inverse difference moment (IDM) measures the local homogeneity of an image. The incidence of co-occurrence of pixel pairs is enhanced when they are close in gray-scale value. The weight factor $(1+(i-j)^2)^{-1}$ makes IDM have small contributions from inhomogeneous areas and $i \neq j$. For an inhomogeneous image, the IDM value is low.

Entropy (ENT) measures the randomness of the image texture according to its intensity distribution. We get the highest entropy when all probabilities are equal. Thus, a lower entropy value indicates a homogeneous image, while an inhomogeneous region has a higher entropy value.

Angular Second Moment (ASM) also known as “Energy”, is a measure of homogeneity of an image [36]. A homogeneous area only has a few gray levels. So GLCM matrix will have a

few but relatively high $p(i, j)$. Meanwhile the sum of the squares will be high. Higher values indicate that the textural is uniform or homogeneous.

Variance (VAR) has a relatively high weight on the elements that differ from the average value of $p(i, j)$. It refers to the gray-level variability of pixel pairs and measures heterogeneity. When the gray-scale value is different from the mean, the variance will increase. The variance, unlike contrast, has no spatial frequency. A high variance indicates a high contrast value; however the converse relationship does not apply [37].

Cluster shade (CLS) measures the skewness of the matrix. It is believed that the CLS predicts the uniformity [38]. When CLS is high, the image is asymmetric.

Cluster prominence (CLP) is also a measure of asymmetry [38]. When CLP is high, the image is less symmetric. Also, when CLP is low, there is a peak in the GLCM matrix around the mean value.

Normalized minimum pixel intensity (IMIN) and normalized maximum pixel intensity (IMAX) are used to remove the effect of different power of the incident laser beam.

3.3 Support Vector Machine algorithm

3.3.1 General description of the algorithm

Support Vector Machine (SVM) is a supervised machine learning method that has been used widely for data mining research such as classification, regression, and other learning

tasks [34]. In general, a classification algorithm uses an object's characteristic parameters to identify which type it belongs to by making a classification decision based on the value of characteristic parameters, which can be expressed in the form of a parameter vector \mathbf{x}_i with i as the identification index of the object. Given a vector $\mathbf{x}_i \in R^J$, $i=1, \dots, N$ in a training data set of N objects and an indicator vector \mathbf{y} with its component y_i of value either 1 or -1, indicating one of the two classes to which the point \mathbf{x}_i belongs to, and $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{iJ})$ can be considered as a position vector in a parameter space of J dimension, where J is the total number of parameters. The set of parameter vectors of \mathbf{x}_i and \mathbf{y}_i can thus represent a training data set $D = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_i, y_i), \dots, (\mathbf{x}_N, y_N)\}$.

In SVM, classification is achieved with a decision function defined as

$$F(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad (6)$$

where $\mathbf{w} = \sum_{j=1}^J w_j \mathbf{u}_j$ is a weighting vector and b is the bias parameter. The goal of SVM algorithm is to determine the decision function F from a given training data set D , which can be used to determine the value of y for an unknown object \mathbf{x} using following criteria:

$$\begin{cases} F(\mathbf{x}) > 0 \Leftrightarrow y = 1 \\ F(\mathbf{x}) < 0 \Leftrightarrow y = -1 \end{cases} \quad (7)$$

At first we consider a linearly separable training data set D . The goal of SVM algorithm is to obtain the decision function which defines an optimal hyperplane of $F=0$ and uniquely separates the objects in a training data with the maximal margins or distances between the hyperplane and nearest objects in each type.

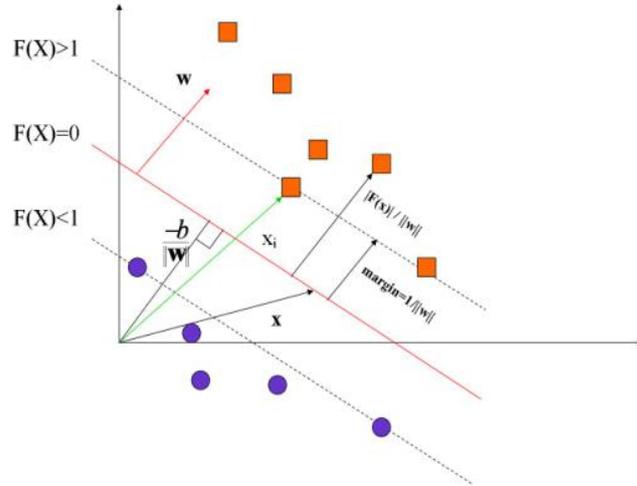


Fig. 3-2 A linearly separable hyperplane

As shown in Fig. 3-2, the distance from the origin to the hyperplane is given by $-\frac{b}{\|\mathbf{w}\|}$ and the margin from distances of the plane $F(x)>1$ and $F(x)<1$ defined by the nearest objects are equal to each other and given by $\frac{2F(\mathbf{x}_i)}{\|\mathbf{w}\|} = 2\|\mathbf{w}\|^{-1}$. With the hyperplane $F=0$, the SVM algorithm separates the two types of objects by finding the maximum margin. This problem is solved in SVM as an optimization problem by minimizing the value of $\|\mathbf{w}\|$ under certain constrains.

Since a data set may not be linearly separable, the accuracy of classification is not always 100%. For these cases, the SVM algorithm applies a soft-margin classification scheme that allows classification errors and solves the following problem to maximize the margin $\|\mathbf{w}\|^{-1}$:

$$\begin{cases} \text{minimize} & Q_1(\mathbf{w}, b, \xi_i) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{subject to} & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall (\mathbf{x}_i, y_i) \in D \text{ and } \xi_i \geq 0 \end{cases}, \quad (8)$$

where ξ_i is the penalty parameters induced by the object I and $C > 0$ is the regularization parameters. Despite the use of the soft-margin scheme, classification of the objects in a practical data set may not yield satisfactory results. To enhance the accuracy, SVM further employs a mapping kernel function to perform classifications in a high-dimensional feature space defined by the training data set instead of the J -dimensional parameter space. For instance consider a case shown in Fig.3-3 (A). The two types of the objects represented by the input parameter vector of $J=2$ is nonlinearly separable in the 2D space. The problem can become linearly separable if one maps the two parameters of x_1 and x_2 into a higher dimensional space using a function $\phi(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$. With the function $\phi(\mathbf{x})$ a hyperplane can be determined to discriminate the given objects linearly with the largest margin [40].

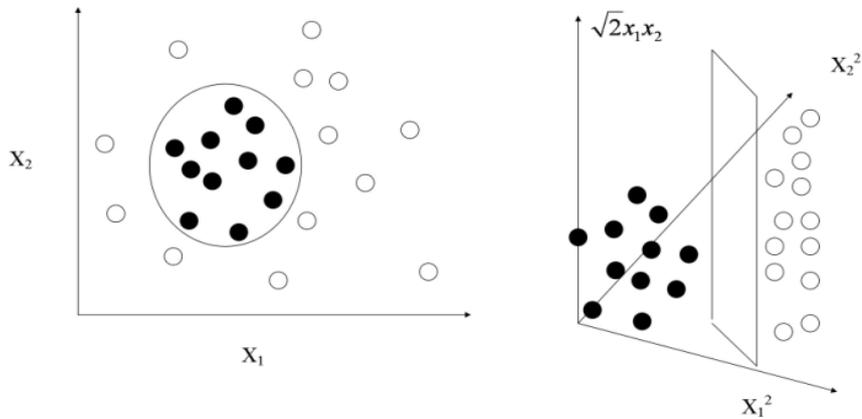


Fig.3-3 (A). A nonlinearly separable dataset. (B). New dataset after mapping

Using a mapping function ϕ , Eq. (8) can be written as following:

$$\begin{cases} \text{minimize} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \\ \text{subject to} & y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \forall (\mathbf{x}_i, y_i) \in D \text{ and } \xi_i \geq 0 \end{cases}, \quad (9)$$

where $\mathbf{w} = \sum_{i \in SV} \alpha_i y_i \phi(\mathbf{x}_i)$. In SVM, the kernel function is defined with paired parameter vectors

as

$$K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j). \quad (10)$$

Since the optimization problem shown in Eq. (9) is based on the dot product of two parameter vectors, SVM only needs the K function to map the training data into a high dimensional feature space without having to explicitly define the function $\phi(\mathbf{x})$. This technique of kernel function based mapping allows the determination of the decision function F by solving the following dual optimization problem [34]:

$$\begin{cases} \text{minimize} & \frac{1}{2} \mathbf{a}^T Q \mathbf{a} - \mathbf{e}^T \mathbf{a} \\ \text{subject to} & \mathbf{y}^T \mathbf{a} = 0, 0 \leq a_i \leq C, i = 1, \dots, N \end{cases}, \quad (11)$$

where $\mathbf{e} = [1, \dots, 1]^T$ is the vector of all ones and Q is a positive semidefinite matrix of rank N with elements given by

$$Q_{ij} \equiv y_i y_j K(\mathbf{x}_i, \mathbf{x}_j). \quad (12)$$

After solving Eq. (11), the optimized vector $\boldsymbol{\alpha}$ and the bias b can be confirmed and the decision function can be written as

$$F(\mathbf{x}) = \sum_{i=1}^N y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b. \quad (13)$$

By solving the dual problem from training dataset D , SVM outputs the values of $y_i \alpha_i$ and b , label names (-1 or 1), support vectors, and other information.

3.3.2 LIBSVM

LIBSVM is an integrated software using the SVM algorithm for classification, regression, and distribution estimation. The purpose of LIBSVM is to help users from other research fields to use SVM easily without knowing the details of the implementation. LIBSVM is available in different programming languages including C++, C#, JAVA, Python, R, and MATLAB, which can be linked with users' own programs. We select an open-source code package LIBSVM 2.86[39] and employ it in our software to examine 38 features of diffraction images. A classification task includes separating data into training and testing sets. Each training set involves a class label (cell type) and 38 features (image features in 3.2.2). LIBSVM produces a model which is based on the training data to predict the target values of the test data. Four basic Kernel functions provided by LIBSVM are:

- a. Linear: $K(x_i, x_j) = x_i^T x_j$.
- b. Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, $\gamma > 0$.
- c. Radial basis function (RBF): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$.
- d. Sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$.

γ , r , and d are kernel parameters. In LIBSVM, $\gamma=1/J$, $r=0$, and $d=3$. The effectiveness of SVM classification depends on the selection of kernel functions. The operator $x_i^T x_j$ exists in four kernel functions and is calculated by the function **dot**(x_i , x_j) in LIBSVM, which returns an inner product. For instance, if we have two cells x_i and x_j , and each has 4 parameters as following: $x_i^T = \{1, 2, 3, 4\}$, $x_j^T = \{5, 6, 7, 8\}$. We can get $K(x_i, x_j) = 1 \times 5 + 2 \times 6 + 3 \times 7 + 4 \times 8 = 70$.

In RBF kernel function, $\|\mathbf{x}_i - \mathbf{x}_j\|^2 = \mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j - 2\mathbf{x}_i^T \mathbf{x}_j$. Among 4 kernel functions, the RBF kernel function is considered as the first choice [43]. Unlike the linear kernel function, the RBF kernel function nonlinearly maps samples into a higher dimensional space. Thus it is able to classify two types of samples when class labels and attributes are nonlinear. The linear kernel function is considered as a special form of the RBF kernel function [44]. It is due to the linear kernel function with a penalty parameter C performs the same as the RBF kernel function with parameters (C, γ) . In addition, the number of hyper parameters influences the complexity of the model selection. It is obvious that the polynomial kernel function has more hyper parameters than the RBF kernel function. At last, there are fewer numerical difficulties for the RBF kernel function. For polynomial kernels function, the kernel values may go to infinity or zero when the degree is large. In contrast, the kernel value of RBF is from 0 to 1.

Accuracy (A) measures the accuracy of classification of two known samples and it is defined as following:

$$A = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

For example PC3 and PCS cells are two known data sets for cell classification, then TP (true-positive) is the number of correctly identified image pairs acquired from PC3 cells; TN (true-negative) is the number of correctly identified image pairs from PCS cells; FP (false-positive) is the number of image pairs of PCS cells but incorrectly identified as of PC3 cells; FN (false-negative) is the number of image pairs of PC3 cells but incorrectly identified as of PCS cells; N is the number of diffraction image pairs in the data set. Five-fold cross-validation is used to evaluate the individual performance of 38 image parameters with

different kernel functions. The algorithm randomly divides the training data set into five or nearly equal five parts: four parts are training sets and one part is the test set. It iterated five times to obtain the average test accuracy A_{av} . Different SVM models were then formed by a parameter vector in the training data with selected parameters M in the same sequence of ranking as components. Each SVM model was trained in the feature space with the training data and the same scheme of five-fold cross-validation is used to obtain A_{av} for the evaluation.

The data is processed by our software in the following procedure:

1. Transform the data to the format of LIBSVM.
2. Choose the texture features we want to train as the training set. The format of input file is: *[label] [index1]:[value1] [index2]:[value2] ...*, where *label* is the class of the classification (1 or -1). The index represents the name of the feature and usually is a continuous integer. Values are the actual data for training.
3. Choose the kernel function for training.
4. Use five-fold cross-validation to find the best parameters with the highest ACC for each kernel function.
5. Use the best kernel function and the best combination of parameters to train the training set and get the training model.
6. Test the model with the test set.

CHAPTER 4 EXPERIMENTAL METHODS

In order to investigate if the p-DIFC method is able to accurately classify cells which have high degree of similarity, we have acquired confocal image stacks and cross-polarized diffraction image pair data from different types of cells. Reconstruction of 3D structures and statistical analysis have been performed to calculate their morphologic parameters and to compare quantitatively the difference and similarity in 3D morphology among different types of cells. Cross-polarized diffraction image pair data has been used to obtain useful image parameters for classification. In this chapter, we present the experimental procedures of confocal imaging and diffraction imaging with an in-house developed system and related image analysis.

4.1 Confocal imaging and 3D reconstruction

4.1.1 Cell preparation

We have studied multiple human cancer cell lines and primary cells derived from non-malignant tissues of prostate and spleen. The human epithelial cells include the PC-3 prostate cancer cell line (CRL-1435, ATCC) and normal prostate cells (PCS440010, ATCC), which are denoted here as PC3 cells and PCS cells, respectively. The PC3 cells were maintained in RPMI-1640 (Gibco BRL, Life Technologies) supplemented with 10% fetal calf serum (FCS) and PCS cells were maintained in the prostate epithelial cell basal medium (PCS440030, ATCC) supplemented with the prostate epithelial cell growth kit (PCS440040, ATCC). The cancer cell lines derived from malignant T and B cells consist of the Jurkat

(TIB-152, ATCC) and Ramos (CRL-1596, ATCC) which are cultured in the same medium RPMI-1640. The cultured cells of PC3, PCS, Jurkat, and Ramos were incubated in 37°C and 5% CO₂ environment with PC3 and PCS adherent to the growing plates and detached by adding Trypsin-EDTA solution. The cells to be imaged were washed with the DPBS buffer before experiments. The protocol of detaching the cells in suspension can be found in the Appendix B.

The human primary cells used in this study include CD4⁺ and CD8⁺ T lymphocytes extracted from fresh spleen tissues. Samples of spleen tissues are from the Department of Pathology, Brody school of Medicine. The operations were within five hours. A spleen tissue was first cut and grinded with two frosted glass slides gently in the RPMI culture medium in the plate. If the cells are grinded with too much power or heavily squeezed, the cell structure may be damaged and present an irregular morphology. Therefore this grinding process requires patience and usually takes about one hour until all small pieces disappear. Acquired cell suspension is filtered with a 70 µm cell strainer to get rid of the fat particles from the spleen. After centrifuging and removing the supernatant and re-suspension, the red blood cells were removed by adding the red blood cell lysis buffer and shaking at room temperature for ten minutes. The protocol of splitting splenocytes can be found in the Appendix C.

To obtain CD4⁺ and CD8⁺T lymphocytes from the prepared suspension of splenocytes, the cells were stained with fluorochrome-conjugated CD surface markers. The suspension sample was divided into two aliquots in a ratio of 1:9. The less aliquot is a control group without staining and only used for setting up the cell sorter. Another aliquot is stained with 3 CD surface markers including CD4 PE-Cy5.5 (MHCD0418, Life Technologies), CD8 FITC

(MHCD0801, Life Technologies), and CD3 PE (CD0304, Life Technologies). Using CD3 marker is to confirm the cells selected by CD4 or CD8 marker are 100% T lymphocytes. The standard procedure of direct immunofluorescence staining of culture cells is attached in Appendix D. At Department of Microbiology and Immunology, a high-speed cell sorter (FACSVantage SE, BD) is used to separate the stained aliquot into two suspensions: CD4+ T lymphocytes and CD8+ T lymphocytes. A blue laser with a wavelength of 488 nm is used for excitation and scatter measurement. Emission filters centered at 694nm and 519 nm are used for measurement of fluorescence by PE-CY5.5 conjugated CD4 and FITC conjugated CD8 accordingly. Another emission filter centered at 575 nm measures the fluorescence by PE conjugated CD3. Signal gating was implemented by the FSC and SSC signals to prevent the presence of debris, doublets, and triplets in collected cell suspensions. After sorting, the ratio of living cells is about 95%. The T cell suspension samples were re-suspended with concentrations adjusted to about 2×10^6 cells/ml for confocal imaging and 1×10^6 cells/ml for p-DIFC measurement. The protocol of cell counting is in Appendix F. For confocal imaging, the cell suspension was further stained with Syto 61 dye for imaging nuclei and MitoTracker Orange dye (both from Life Technologies) for imaging mitochondria. The protocol of this double staining is in Appendix E.

4.1.2 Confocal imaging

Laser scanning confocal microscopes (LSCM) employ one or more laser beams for excitation of fluorescent molecules in organelles. The laser beams are focused inside a cell by the objective, which is also used to collect the fluorescent light emitted from the reagent

molecules. As the laser beams are scanned by two mirrors before they enter the objective, the laser beams' common focal spot moves over the x-y plane point by point. At each point, the fluorescent light collected by the objective is focused by another lens into a pinhole, which prevents fluorescent light coming out of the regions outside of the focal spot being detected by the photomultiplier (PMT) placed behind the pinhole. This is the principle of spatial filtering underlying LSCM that is responsible for the very short focal depth of each image consisting of the moving focal spots in the x-y plane. After the scanning within the field of view is complete, the signals acquired by PMT are stored as pixels for one image slice at a particular z position. The emitted fluorescent signals, up to three with different wavelengths bands, are stored into separate data arrays inside the computer memory. An image stack of multiple slices can be acquired by continuous translating the imaged cell over a sequence of z-positions with a step size of Δz with one image slice per z-position.

Most researchers only use LSCM to obtain high-contrast and high-resolution cell images at a particular "depth" of the imaged cell or cells. Different from that, the imaging techniques in this dissertation is to acquire the image stack for 3D reconstruction and quantification of the morphology of the imaged cell by our in-house developed software. For this reason, we emphasize on acquisition of a complete image stack of sufficient number of image slices with signal-to-noise ratios as large as possible, which is critical for the 3D reconstruction software to segment correctly the interested intracellular organelles such as nucleus and mitochondria through pixel-to-pixel calculations in each image slice of the stack. Following figures provide examples of complete image stack of 6 types of cells. The ideal top slice and bottom slice

should be both dark images without any signal to ensure the completeness of the acquisition along the z-axis.

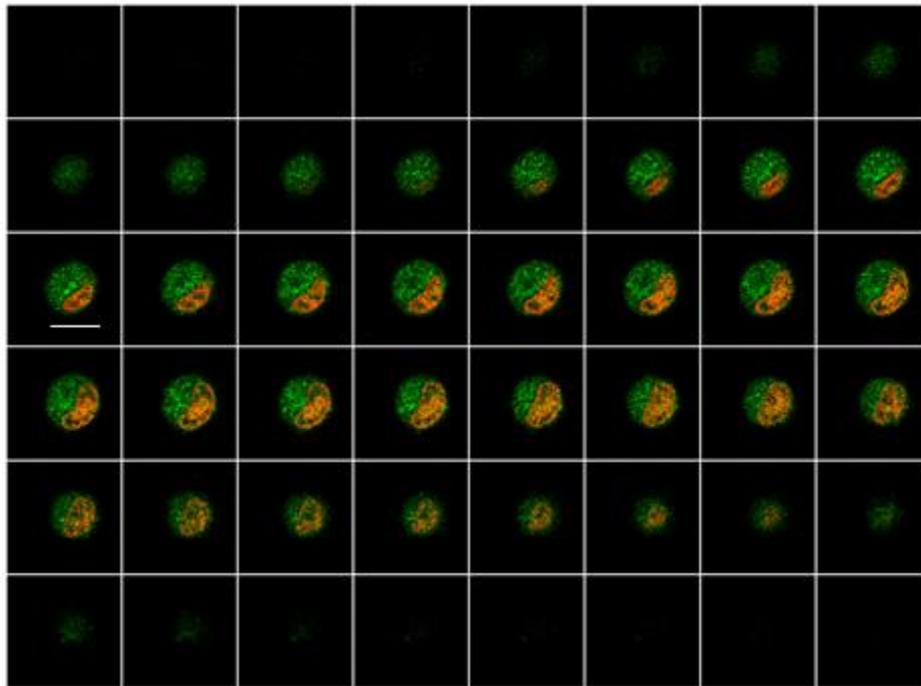


Fig.4-1 Confocal image stack acquired from a PC3 cell

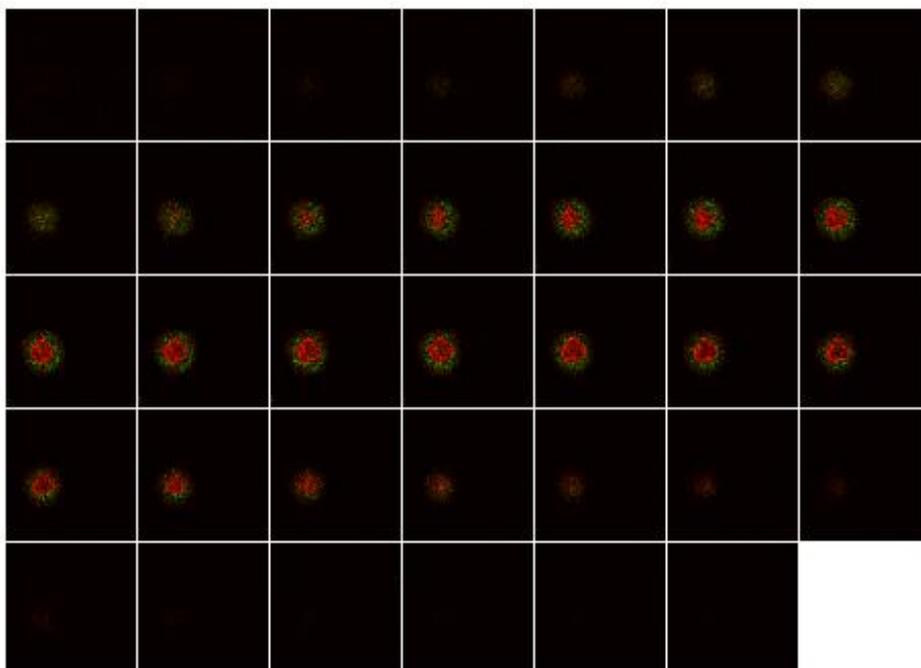


Fig.4-2 Confocal image stack acquired from a PCS cell

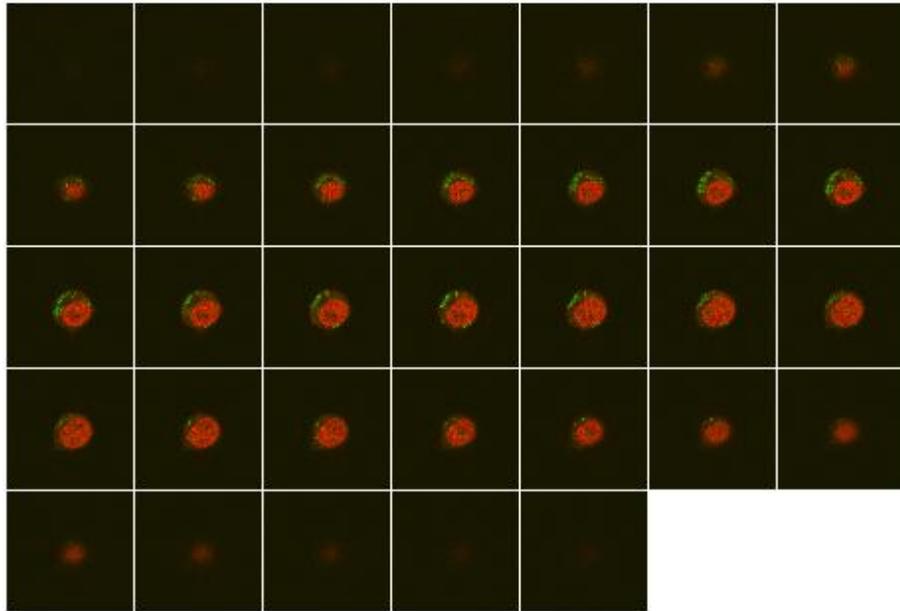


Fig.4-3 Confocal image stack acquired from a Jurkat cell

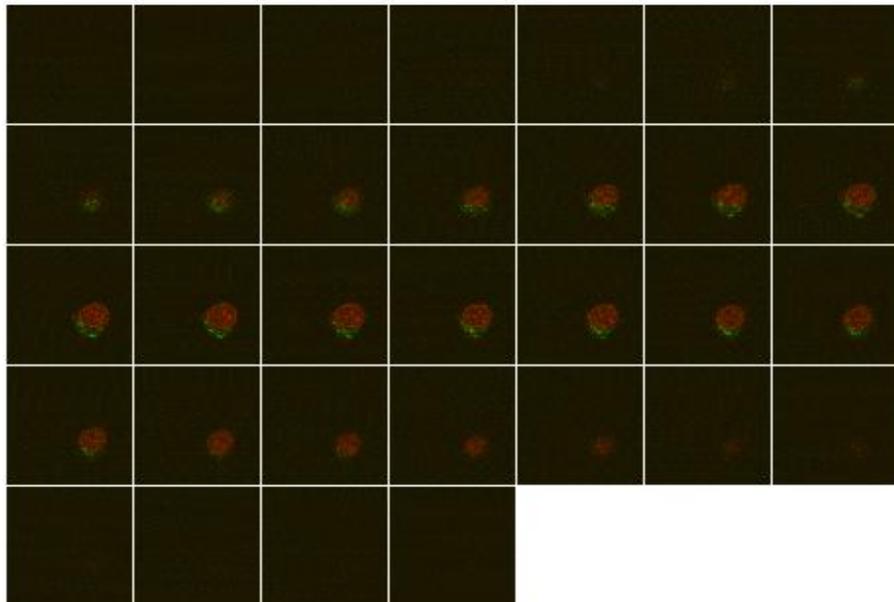


Fig.4-4 Confocal image stack acquired from a Ramos cell

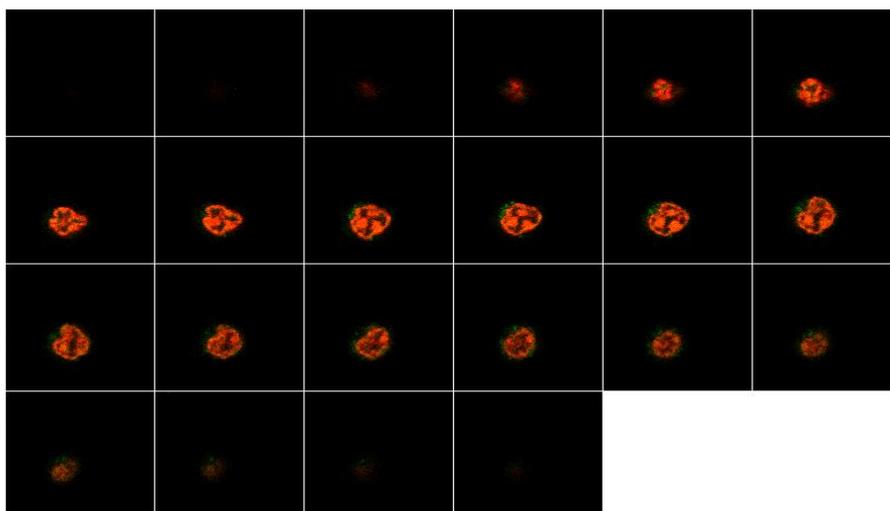


Fig.4-5 Confocal image stack acquired from a CD4+ T lymphocyte

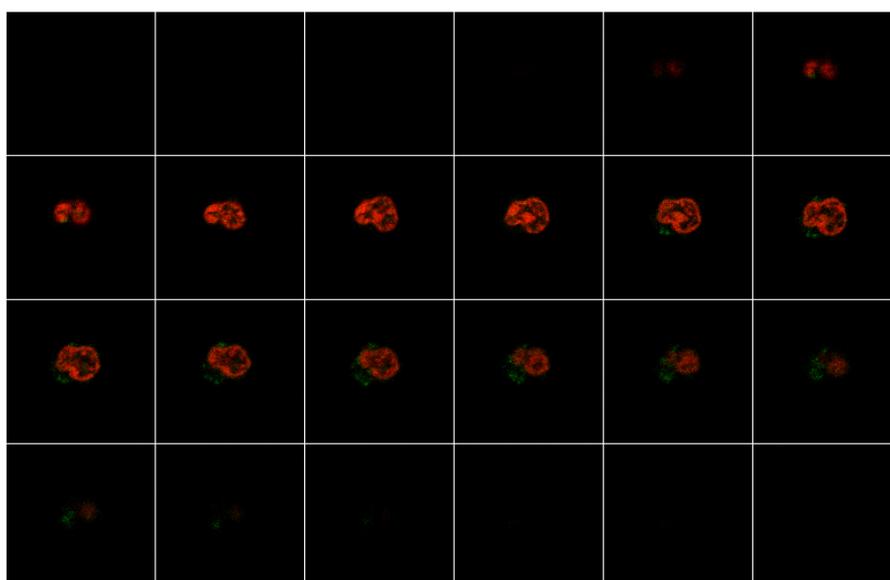


Fig.4-6 Confocal image stack acquired from a CD8+ T lymphocyte

Department of Microbiology and Immunology provides the LSCM (LSM 510, Zeiss) in this study. Since we need high-contrast and high-resolution images, the objective of highest magnification should be used. We choose 63x water-immersion objective for PC3, PCS, Jurkat, and Ramos cells, and 100x oil-immersion objective for CD4+ and CD8+ T lymphocytes. The excitation lasers of confocal imaging are the 633nm red laser and the

543nm green laser for activating Syto-61 and MitoTracker antigens accordingly. A low-pass 650 nm and band-pass 560-615 nm emission filters are selected for the red and green channels as shown in Fig.4-7 (A) and (B) accordingly. An image slice such as Fig.4-7(C) is composed of two channels: red channel for the nucleus and the cytoplasm; green channel for the mitochondria.

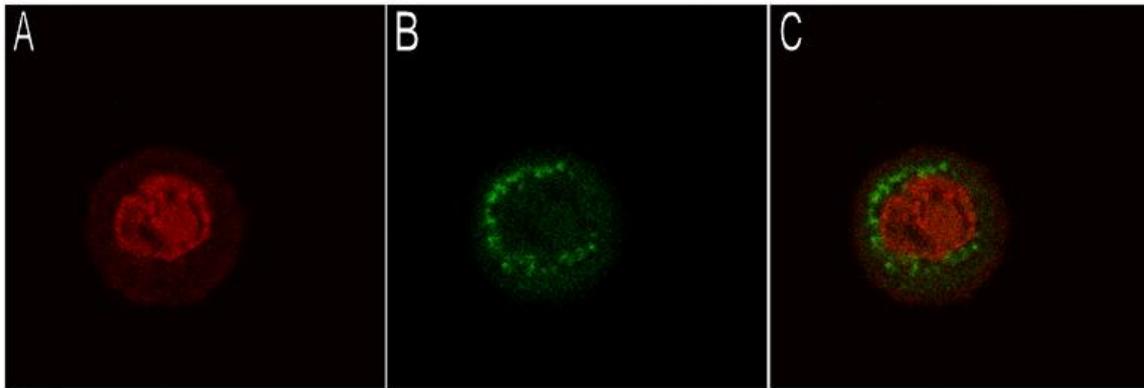


Fig. 4-7 Confocal Slice Image acquired from a normal prostate cell in different channels: (A) Image slice in the red channel; (B) Image slice in the red channel; (C) Image slice in red and green channels.

The pixel depth we choose for imaging is 12 bits and the frame size is 512x512. The z-axis translation step size Δz is typically $0.5\mu\text{m}$. The scan rate is one of the keys of the image quality. One can slow the scan speed to allow more photons to integrate on the detector, or apply image averaging to remove random noise, or a combination of both. The average numbers available in the LSM software are 1, 2, 4, and 8; the scan speed in the LSM software has a range from 1 to 12. We usually apply speed 8 to have a fast acquisition without losing too much signal. If the image quality is still not satisfactory, a scan average number of 8 can be selected to improve the signal-to-noise ratio dramatically. However the image acquisition time will be 8 times longer which increases the chance of cell motion and bleaching of the

fluorescent dyes. Another key to the image quality is the combination of the laser intensity and the detector gain which determine the total number of fluorescent photons received by the PMT and the brightness of the image. By examining the histogram, these parameters can be selected so that a few percent of pixels of the most intensive image slice are saturated. This will ensure the best use of the dynamic range of pixels for the later reconstructions and calculations. The signals from stained cells vary significantly and there is no constant number for the laser intensity or the gain. But the average intensity of the 633nm laser is 6 and the average of the 543nm laser intensity is 18. The average of detector gain for both channels is 600 and very sensitive. Therefore when one tests for the best combination of the intensity of laser and the detector gain, the detector gain should be the first thing to modify.

After initial settings, a careful observation should be made on the live stream. The stream helps users to find candidate cells for imaging in the field of view. Then cells are randomly selected for imaging by changing the digital zoom to 4x to clearly show the inner organization of the cells. The laser intensity and the gain are adjusted until one can acquire a high-contrast and high-resolution image. After finding the best gain and the best laser intensity, one returns to the live stream to find the top and the bottom of the image stacks. The experiment of capturing entire image stacks starts after selecting the scan speed and the average number. If the quality of the image stack is not as good as the one under the live stream, one should change the scan speed and the average number to increase the quality of images.

4.1.3 3D reconstruction and parameter calculation

A Matlab-based software developed previously reconstructs the 3D structures of the cells from confocal image stacks. The cytoplasm, the nuclei, and the mitochondria of the cell are assigned with different values according to the image intensity. Red channel is sorted for the nuclei and green channel is sorted for the mitochondria. The cytoplasm is sorted by both channels. The result of sorting is visible in Fig.4-8 and an alternative is provided for resorting in a fast way. The threshold of the sorted image is able to be modified by the scrolling bar on the interface.

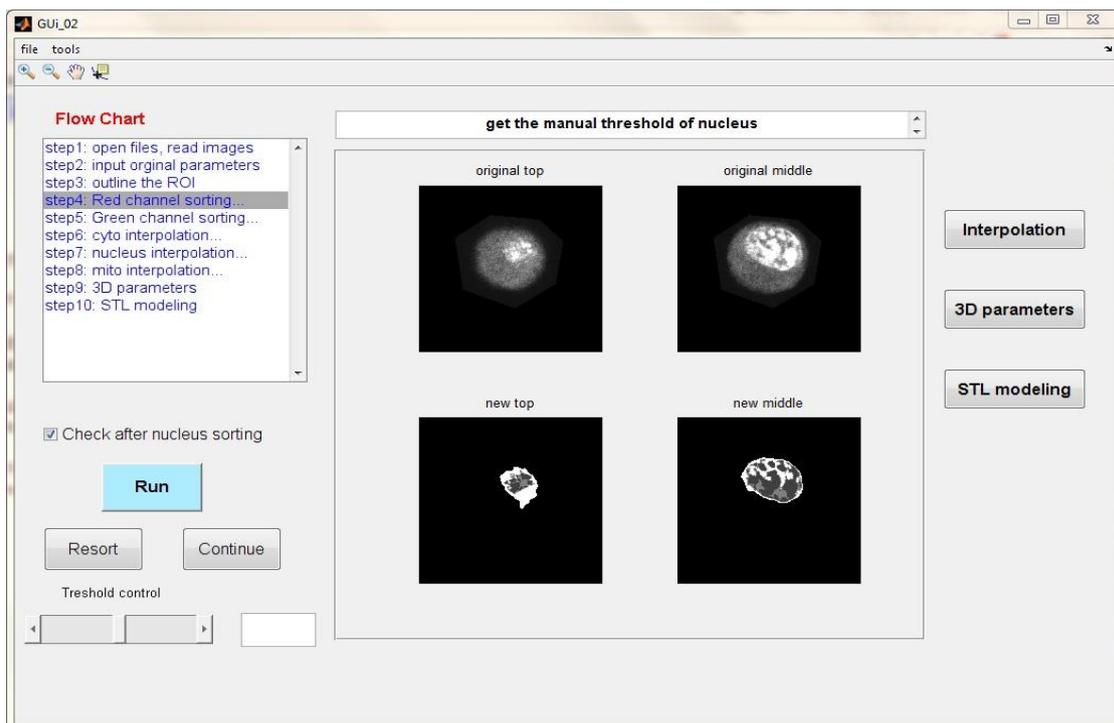


Fig. 4-8 User interface of the 3D reconstruction software

The process of resorting doesn't require any manual drawing. Also the system performs interpolation and generates about 21 3D parameters of the cell including: Cell volume, Cell surface area, Surface to volume ratio of cell, Index of surface irregularity of cell, Average

distance of cell membrane voxels to centroid, Standard deviation, Nuclear volume, Nuclear surface area, Index of surface irregularity of nucleus, Mitochondrial volume, Mitochondrial surface area, Surface to volume ratio of mitochondria, Index of surface irregularity of mitochondria, Distance between the centroids of nucleus and cell, Volume ratio of nucleus to cell, and Volume ratio of mitochondrion to cell.

To process the cell images with the software, firstly we use Zeiss LSM image browser to open the LSM image and convert the LSM file to 12-bit tiff stacks. Because our software can only process tiff image stacks. The software opens the tiff image stacks according to the files' path, the total image slice number, and the Z step. The first and the last slice numbers of the cytoplasm, the mitochondria and the nucleus are required for sorting procedure as well. We draw the interest of the area manually to select the pixels for the cell. The system will halt and wait for the threshold of the nucleus. Therefore we can sketch the area of the nucleus and select the pixels for it. Within the area of nucleus, system will calculate the average intensity as a reference. A histogram analysis will be performed, and the minimum value of the first peak will be chosen as the threshold of the cell membrane. Meanwhile the pixel intensity of the second peak will be used as the threshold of the nucleus. Sobel operation is performed based on derivative based edge detection on all slices, and a threshold of derivative is selected through the trial and the error to generate a binary stack for segmentation of the cytoplasm and the nucleus from the background. Opening and closing fill-point gap operations are implemented to remove the invalid pixels and smooth out the cell membranes' border lines. We observe the result of the nucleus after sorting. If the size of the nucleus is

too large, the threshold we used is too small. Therefore, it is necessary to resort by adjusting the threshold. After sorting the nucleus, we draw the area of mitochondria manually and pick a bulk of closed area of the mitochondria. The system will use its average intensity as the threshold for sorting the mitochondria. System uses the Watershed algorithm to produce the borderlines that separate different mitochondria clusters. Next we observe the result of the mitochondria after sorting to determine if a resorting is needed. If we are satisfied with the result, the system will do the shape based interpolation for the cell. The system interpolates the cytoplasm, the nucleus, and the mitochondria accordingly. A sliding scheme is applied to add additional slices between the neighboring slices. After reconstruction, we quantitatively analyze the 3D structures and obtain 27 morphological parameters of the cytoplasm, the nucleus and the mitochondria, which provide essential morphology data correlated to the diffraction image feature parameters obtained by the p-DIFC method. Statistical analysis of morphological parameters is performed with the SPSS software (IBM, version 17) to evaluate the significance of the differences between two selected cell types.

4.2 Diffraction imaging measurement

4.2.1 Optical design

The p-DIFC system has been improved both on software and hardware to increase the speed of imaging and the accuracy of analysis as shown in Fig. 4-9. With a concentric sheath fluid at a higher pressure entering the chamber, cells carried by the core fluid move in single file through a continuous-wave solid state green laser (MGL-III-532-100, CNI) beam and

each emitting light scatters at the same wavelength of 532nm. The laser beam is linearly polarized with its direction adjustable using a half-wave plate (WP). The power of the beam passing WP can be up to 200mW. A spherical lens of 75mm in focal length focuses the incident beam onto the core fluid with a spot diameter of about 30 μ m. The light scatter from flowing cells were collected by an infinity-corrected 50x objective (378-805-3, Mitutoyo) within an angular cone, which was centered at the scattering polar angle $\theta_s=90^\circ$ along the x-axis and of a cone angle in water. An interference filter(WF) of 532nm combined with a polarizing beam splitter(PBS) divides the side scatters collected by an objective into two components of horizontal and vertical polarizations, labeled as p- polarization and s- polarization for p-CCD cameras and s-CCD cameras (LU75M, Lumenera). Pairs of diffraction images of 640x480 pixels and 12-bit pixel depth are acquired from flowing cells with different incident beam (vertical, horizontal, and 45 $^\circ$ polarizations). In each measurement, a small portion of cell suspension sample (<200 μ l) is loaded into the core fluid syringe followed by alignment of the imaging unit to the off-focus position Δx ($\Delta x=150$ μ m for Prostate cells and $\Delta x=100$ μ m for other types of cells) and adjustment of incident beam power. About 1500 cells were imaged for each cell type from each cell sample for each polarization direction. Then we apply image preprocessing to filter out non-cell images and calculate 38 image parameters extracted from each cell to characterize image texture patterns and classify the cells. We further randomly separate the cell image pairs into two sets of training and test data and apply our classification software based on the LIBSVM package to obtain the best SVM model with the highest value of the accuracy.

In Fig. 4-9, BE is the beam expander; WP is the half-wave plate; M is the mirror; ND means neutral density filters; FL means focusing lens; FC is the flow chamber; CL is the condenser lens; Ob is the objective; WF means 532nm wavelength filter; PBS is the polarizing beam splitter; TL is the tube lenses; BS is the beam splitter; PMT means photomultiplier; CCD is the camera recording either s- or p-polarized diffraction images. The axes of x and z are labeled by red lines.

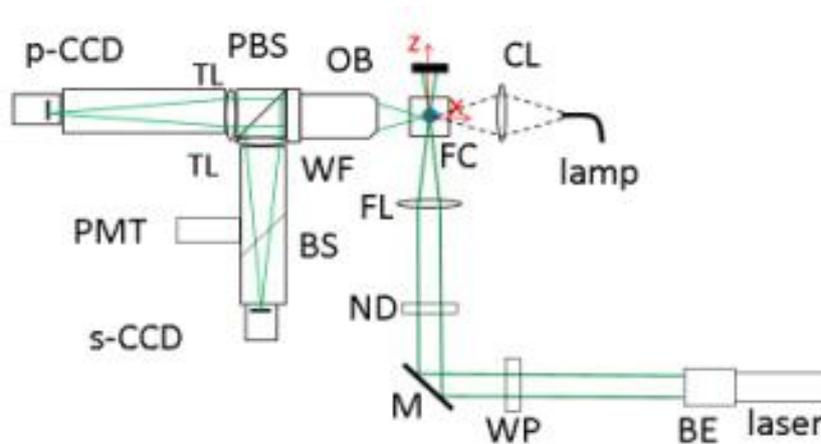


Fig. 4-9 Top view of the p-DIFC system for acquisition of s- and p-polarized diffraction images

4.2.2 Camera control and data acquisition

By calling the Application Program Interface (API) functions provided by the vendor of two CCD cameras (LU075M, Lumenera), we have developed an image data acquisition software to control the cameras and adjust camera settings such as exposure time, image signal gain, and the preview window for optic alignments to take two images simultaneously after receiving the external trigger signals produced by the PMT as shown in Fig. 4-9. The software will make a “beep” sound when it receives an external trigger to notify the users that

a pair of cell images is captured by the system. Fig. 4-10 shows the flow diagram of the software.

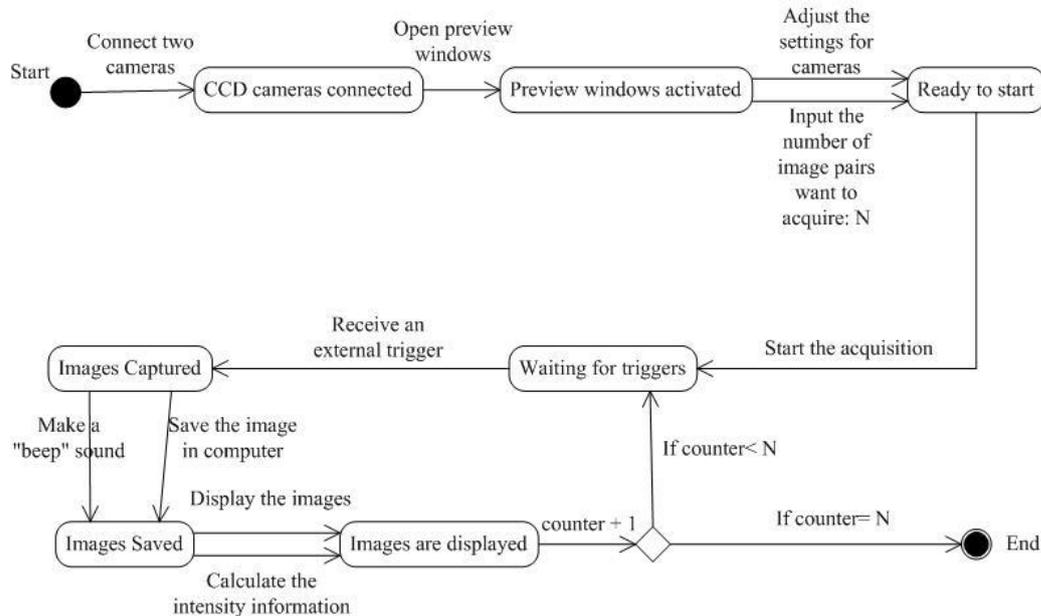


Fig. 4-10 Flow diagram of the image acquisition

After connecting the cameras to the system and setting the basic parameters for both cameras, the software will wait for the external triggering signals. When a trigger signal is received from the PMT, the system will save the images in the hard drive of the computer and make a “beep” sound. The images will be displayed with their intensity distribution. The image counter will add up one and compare with the target number of cells N . The software will terminate if the counter number equals N . In the initial version, the image acquisition software can capture one pair of cross-polarized diffraction images per second. Each paired image acquisition is implemented as a single process in the Windows Operating System (OS). A process cannot be activated until the previous one completes, which produces a delay or gap of no data acquisition in time after the initial camera triggering. If this issue cannot be

corrected, many flowing cells would not be able measured due to the waiting gap. We have improved the software by implementing a multithreading mechanism which converts the entire process of imaging data acquisition into multiple threads. In this design, a single thread of execution is the smallest sequence of programmed instructions that can be managed by the OS. Multiple threads can be executed within the same process that shares the same resource such as memory. Each paired image data acquisition is assigned as one thread and the software can process multiple threads at the same time. A test of this multithreading design has yielded the results through multiple experiments and the maximum speed of paired image acquisition can reach a rate up to ten image pairs per second. And this speed can increase significantly with an upgrade on the computer hardware.

Another significant software improvement is to process pixel intensity information and present the detail of this information in real time. This functionality is very useful for adjustment of incident laser beam power to reduce the probabilities of acquiring underexposed or overexposed image data in subsequent acquisition because the dynamic ranges of the cameras are quite limited. For this purpose, investigators using the p-DIFC system can first acquire ten to twenty pairs of diffraction images for adjustment of the incident beam power. As each image pairs are continuously acquired, the image pixel intensity information is calculated and displayed on the main interface of the software in the form of the maximum pixel intensity, the minimum pixel intensity, the average pixel intensity, and the total number of saturated pixels. The real-time feedbacks of image pixel intensities

are not only helpful to adjust the beam power before starting the data acquisition but also useful to monitor the image quality during data acquisition.

4.2.3 Data analysis

The first step to analyze the raw data is image preprocessing. The algorithm mentioned in 3.1 is implemented in a Matlab program and able to filter out the 12-bit raw diffraction images $J(z,y)$ into different categories. Only images of cells are kept for the next step. After converting the 12-bit images to 8-bit images, we calculate the GLCM parameters and the image intensities to acquire a parameter space. With the selected kernel function and certain combinations of parameters, SVM is able to create a feature space and classify cells with a higher accuracy than the parameter space. One can select the feature space according to its accuracy and save it as a training model. The basic flow of data processing in the p-DIFC method is demonstrated in Fig. 4-11.

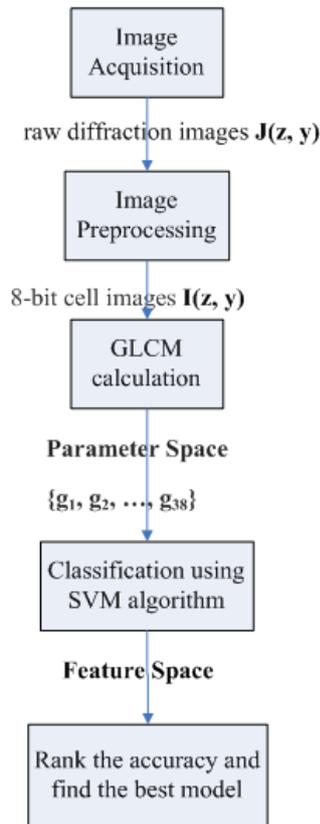


Fig. 4-11 Flow diagram of data processing in p-DIFC method

Image parameters calculated from the GLCM algorithm and the image intensity are analyzed by another software developed in C#.NET. As shown in Fig. 4-12, the user can calculate the GLCM parameters and display any two of them in a scatter plot and two histogram plots. The statistical information is also provided including the mean value, the standard deviation, the maximum value, and the minimum value. This gives us an initial estimation if we can classify two types of cells in a parameter space. A SVM ranking function has been added recently. The user can rank 38 parameters in the order of the classification accuracy and compare the accuracy with the distribution in the histogram and the scatter plots. But according to our careful tests with many types of cells, not all parameters can be clearly distinguished on a scatter plot or multiple histograms. Some parameters which overlap each other on the scatter plots or histograms have a related high

accuracy classified by SVM algorithm. This proves that the image parameters are not always linearly separable in the parameter space and not accurate enough to separate two types of cells.

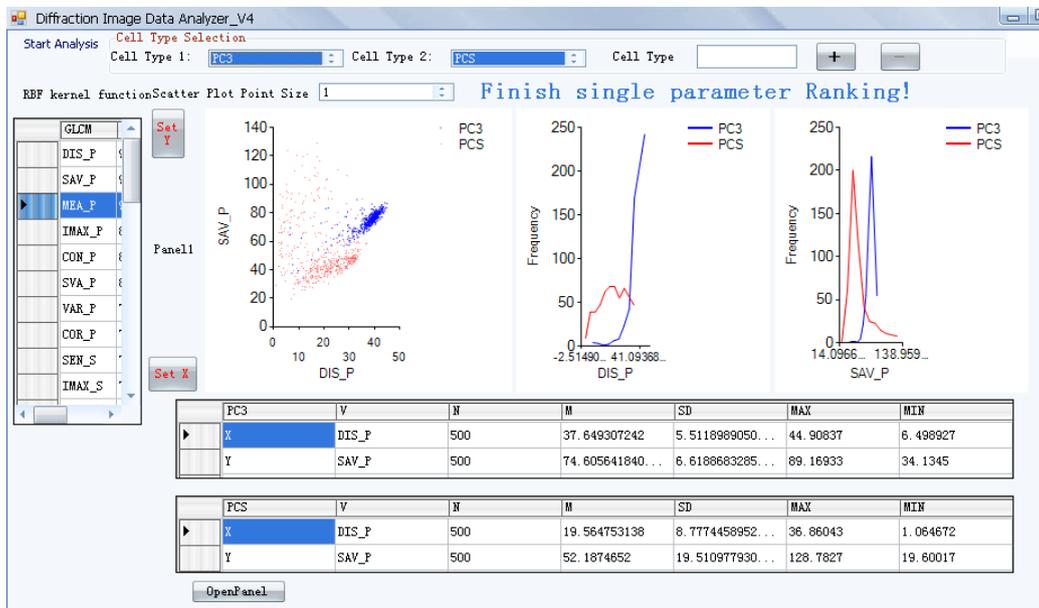


Fig. 4-12 User interface of parameter analysis software

Fig 4-13 shows the user interface of the SVM classification software which is also developed in C#.NET. After selecting the path of an Excel file containing the parameter values of diffraction images, the user is able to observe the details of experiment information and cell parameters. Four kernel functions mentioned in 3.3.2 are available in the software. One can change the fold of cross-validation n to improve the accuracy of classification. The default value of n is 5. Initially user can click a button to start the single parameter ranking with selected Kernel function. The software will call the function from LIBSVM [34] and return the accuracy (ACC) of the training with a single parameter. The ACCs for 38 parameters are ranked and displayed in the “Result” panel. The user can test different

combinations of parameters according to the ACC and find the best model for classification.

The information of training model mentioned in 3.3.1 is saved in a Text file, which can be reused to classify other data set.

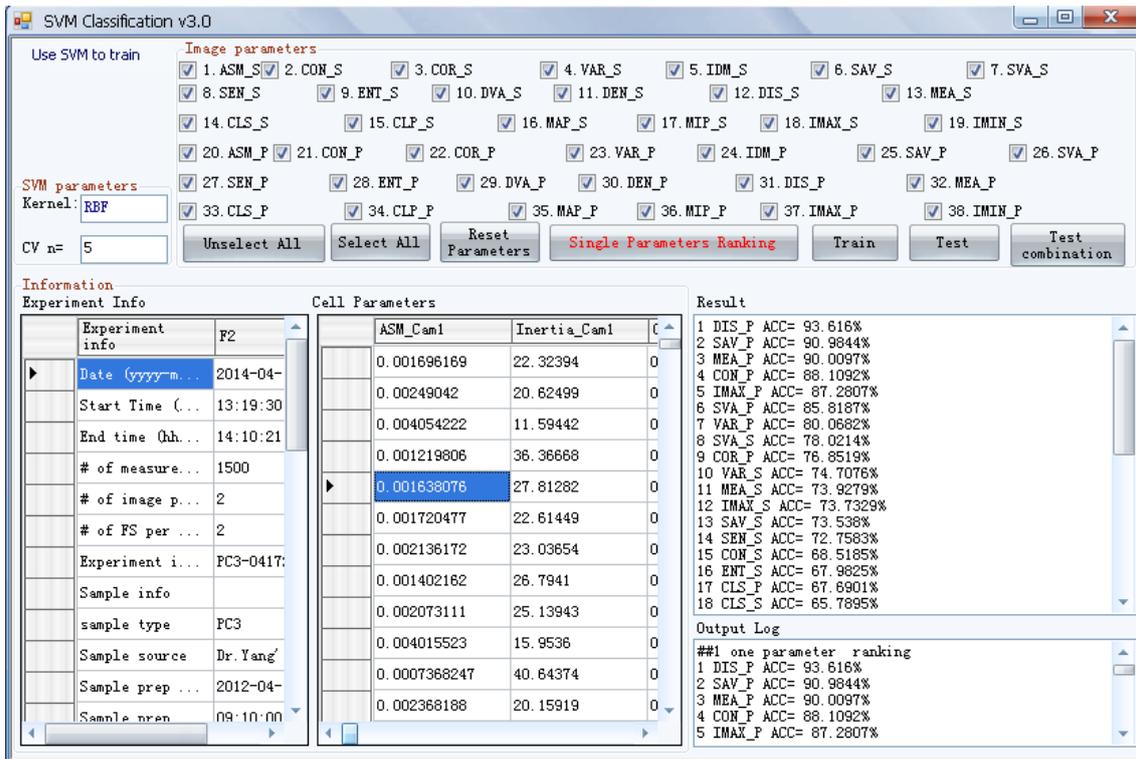


Fig. 4-13 User interface of SVM classification software

CHAPTER 5 RESULTS AND DISCUSSION

In this chapter the results of confocal imaging and 3D reconstruction will be presented to compare quantitatively the morphology parameters of different cell types for phenotyping study. Cross-polarized imaging of single cells has been carried out with the p-DIFC method and the results of measurement were used for classification study of the different cell types. With these results, we discuss the relations among the data acquired with different imaging measurements and their implications in development of a new method for cell assay and classification.

5.1 Confocal imaging and 3D reconstruction

Confocal imaging of different types of cells has been performed followed by the image segmentation and the 3D reconstruction to quantitatively determine and compare their 3D morphology parameters, which also serve as the baseline data to gain the insight into differences in light diffraction patterns exhibited by their cross-polarized diffraction images.

5.1.1 Morphology study of the Jurkat T and Ramos B cell lines

We first selected the Jurkat T and Ramos B cell lines for confocal imaging and 3D reconstruction since these two cell lines are derived from human malignant lymphocytes which have been widely recognized to have indistinguishable morphology using conventional microscopy methods. Current methods for clarifying these two cell lines require the use of

different fluorescent antibody markers on the cell membrane. A total of 60 Ramos cells and 45 Jurkat cells were imaged with the laser scanning microscope and their 3D morphology was reconstructed from the acquired image stack data with the software described in Chapter 4. Fig. 5-1 presents the perspective views of reconstructed Ramos and Jurkat cell structures using different color hues to depict the three major intracellular organelles (cytoplasm, nucleus and mitochondria) with selected morphological parameters listed below each cell. By comparing these single cell images, one can clearly see that the cell morphology varies significantly even among those within the same cell type and thus the distributions of the quantitative parameters can be quite wide.

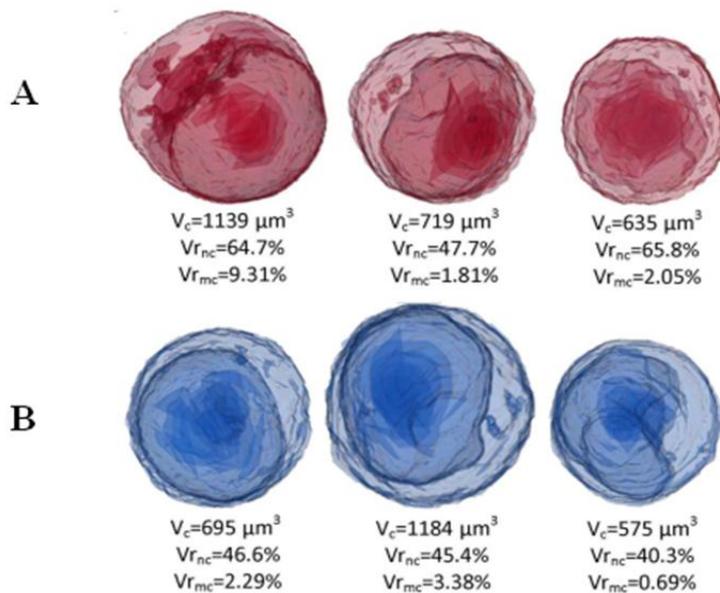


Fig. 5-1 Perspective views of reconstructed 3D structures of three Ramos (A) and three Jurkat (B) cells. Three parameters at the bottoms for each cell are cell volume V_c , volume ratio of nucleus-to-cell $V_{r_{nc}}$, and volume ratio of mitochondria-to-cell $V_{r_{mc}}$.

A total of 29 morphology parameters were calculated to characterize 3D cell morphology.

Table 5-1 provides 16 key parameters obtained with the in-house developed software to illustrate their major morphologic features and statistically significant differences. As expected, the mean values and distributions of these parameters exhibit high degree of similarity between Jurkat and Ramos cell lines in their surface area, volume, shape of the cell, and shape of the nucleus. The only statistically significant differences ($p < 0.05$) observed is the nucleus-to-cell volume ratio $V_{r_{nc}}$ which is highlighted in Table 5-1.

Table 5-1 Morphological parameters of Jurkat and Ramos cell lines

Parameter	Symbol	Unit	Mean \pm Standard Deviation		p ⁽¹⁾
			Jurkat(n=45) ⁽¹⁾	Ramos (n=60) ⁽¹⁾	
Cell surface area ⁽²⁾	S_c	μm^2	526.4 \pm 162	521.3 \pm 153	0.867
Cell volume ⁽³⁾	V_c	μm^3	663.7 \pm 308	666.5 \pm 296	0.962
Cell surface to volume ratio	SVr_c	μm^{-1}	0.8495 \pm 0.168	0.8493 \pm 0.207	0.996
Cell surface irregularity index ⁽⁴⁾	SIi_c	$\mu\text{m}^{-1/2}$	201.7 \pm 28.5	199.7 \pm 20.6	0.678
Average distance of cell membrane voxels to centroid	$\langle R_c \rangle$	μm	6.797 \pm 2.14	7.110 \pm 2.45	0.495
Standard deviation of R_c	ΔR_c	μm	1.841 \pm 0.721	2.051 \pm 0.854	0.186
Nuclear surface area	S_n	μm^2	425.5 \pm 158	388.2 \pm 130	0.187
Nuclear volume	V_n	μm^3	407.8 \pm 199	367.4 \pm 185	0.285
Nuclear surface irregularity index	SIi_n	$\mu\text{m}^{-1/2}$	206.8 \pm 42.1	204.3 \pm 28.9	0.716
Mitochondrial surface area	S_m	μm^2	578.3 \pm 803	485.5 \pm 728	0.538
Mitochondrial volume	V_m	μm^3	33.40 \pm 38.6	32.1 \pm 34.8	0.853
Mitochondrial surface to volume ratio	SVr_m	μm^{-1}	13.56 \pm 3.80	12.48 \pm 2.84	0.100
Mitochondrial surface irregularity index	SIi_m	$\mu\text{m}^{-1/2}$	731.3 \pm 695	677.0 \pm 550	0.656
Nucleus-to-cell centroid distance	D_{nc}	μm	0.1760 \pm 0.053	0.1597 \pm 0.048	0.096
Nucleus-to-cell volume ratio	$V_{r_{nc}}$	-	0.6280\pm0.126	0.5479\pm0.143	0.004
Mitochondrion-to-cell volume ratio	$V_{r_{mc}}$	-	0.0526 \pm 0.066	0.0484 \pm 0.055	0.662

⁽¹⁾ n = number of imaged cells, p-values were obtained by a two-sample t-test method.

⁽²⁾ $S = N_s \cdot s_0$ with N_s as the number of voxels on the membrane of the organelle and s_0 as the diagonal plane area of voxel.

⁽³⁾ $V = N_v \cdot v_0$ with N_v as the number of voxels inside the organelle of interest and v_0 as voxel volume.

⁽⁴⁾ $SIi = N_s \cdot a_0 / (V)^{1/2}$ with a_0 as the side length ($= 0.07 \mu\text{m}$) of voxel.

To compare the morphology data of the two cell lines in details, we provide in Fig. 5-2 the histograms of $V_{r_{nc}}$ of the Jurkat and Ramos cells. It can be seen that the values of $V_{r_{nc}}$ of most Ramos cells are centered around the point of 0.6 while two peaks exist in the histogram of the Jurkat cells located at the values of around 0.5 and 0.7. Although the difference in the histogram distributions of $V_{r_{nc}}$ appears, the difference in cell volume and nucleus volume of Jurkat and Ramos cells are not statistically significant at all according to the p-values in Table 5-1. Therefore the only statistically significant difference among the morphologic parameters of the two cell lines can only be observed in $V_{r_{nc}}$, which confirms the conventional view that microscopy study alone cannot have the capacity to distinguish the Jurkat and Ramos cells based on their morphology.

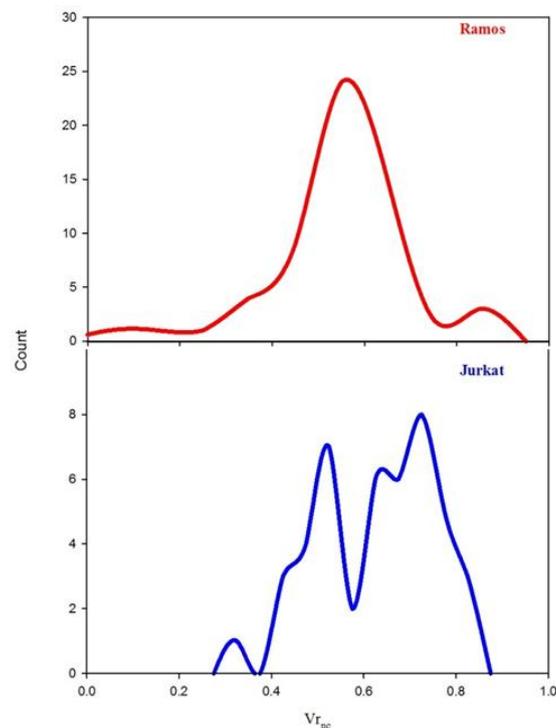


Fig. 5-2 Histograms of nucleus-to-cell volume ratio for the Jurkat and Ramos

5.1.2 Morphology study of the normal and cancerous prostate epithelial cells

To extend our morphology study to the epithelial cells, we have selected the normal and cancerous prostate cells of PCS and PC3 cells for confocal imaging and 3D reconstruction. In Fig. 5-3, we present the perspective views of the 3D structures of the prostate cells. Three parameters at the bottom of each cell are cell volume V_c , volume ratio of nucleus-to-cell $V_{r_{nc}}$, and volume ratio of mitochondria-to-cell $V_{r_{mc}}$. It can be observed directly from the image data that the major differences between PC3 and PCS cells are in the cell volume. The PC3 cells are almost two times bigger than PCS in the cell volume, while the PCS cells have a larger volume ratio of mitochondria-to-cell than PC3.

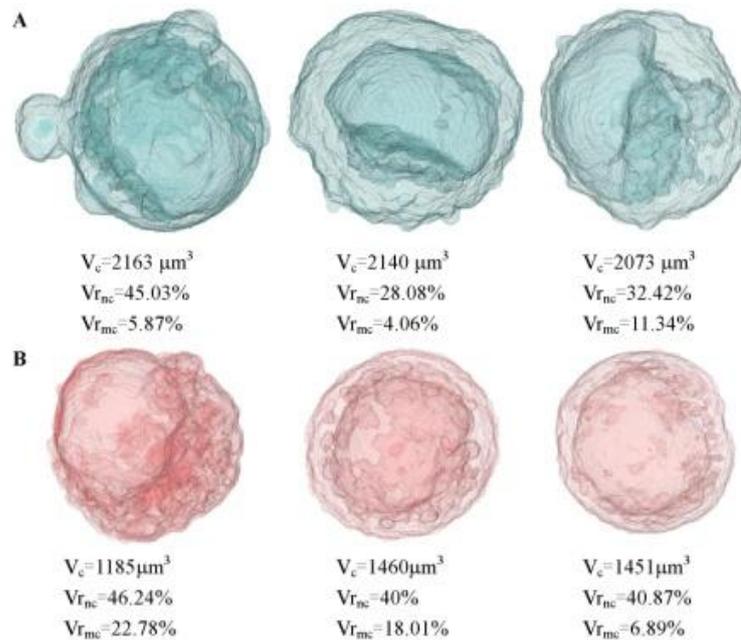


Fig. 5-3 Perspective views of reconstructed 3D structures of three PC3 (A) and PCS (B) cells. Three parameters at the bottoms for each cell are cell volume V_c , volume ratio of nucleus-to-cell $V_{r_{nc}}$, and volume ratio of mitochondria-to-cell $V_{r_{mc}}$.

Table 5-2 presents the values of the mean and the standard deviation of 17 key parameters with the p-values to test the statistical significance on the parameter difference between the two cell types. The morphology parameters show clearly that most statistically significant morphological differences are related to those of cell and nucleus.

Table 5-2 Morphological parameters of PC3 and PCS cells

Parameter	Symbol	Unit	Mean \pm Standard Deviation		p ⁽¹⁾
			PC3 (n=40) ⁽¹⁾	PCS (n=38) ⁽¹⁾	
Cell surface area ⁽²⁾	S _c	μm^2	1135 \pm 226	918.6 \pm 229	7.0 \times 10 ⁻⁴
Cell volume ⁽³⁾	V _c	μm^3	2116 \pm 623	1543 \pm 665	1.9 \times 10 ⁻⁴
Cell surface to volume ratio	SVr _c	μm^{-1}	0.5615 \pm 0.123	0.6386 \pm 0.134	0.011
Cell surface irregularity index ⁽⁴⁾	SIi _c	$\mu\text{m}^{-1/2}$	242.8 \pm 25.7	231.6 \pm 26.6	0.66
Average distance of cell membrane voxels to centroid	<R _c >	μm	8.788 \pm 1.19	7.881 \pm 1.015	1.0 \times 10 ⁻³
Standard deviation of R _c	Δ R _c	μm	2.297 \pm 0.781	2.002 \pm 0.657	0.076
Nuclear surface area	S _n	μm^2	830.5 \pm 231	665.9 \pm 344	0.015
Nuclear volume	V _n	μm^3	1022 \pm 383	679.5 \pm 379	1.7 \times 10 ⁻⁴
Nuclear surface to volume ratio	SVr _n	μm^{-1}	0.8451 \pm 0.121	1.018 \pm 0.175	2.5 \times 10 ⁻⁶
Nuclear surface irregularity index	SIi _n	$\mu\text{m}^{-1/2}$	254.2 \pm 33.4	245.3 \pm 62.4	0.43
Mitochondrial surface area	S _m	μm^2	546.9 \pm 309	629.4 \pm 347	0.27
Mitochondrial volume	V _m	μm^3	160.7 \pm 108	148.8 \pm 95.7	0.61
Mitochondrial surface to volume ratio	SVr _m	μm^{-1}	4.272 \pm 1.93	4.993 \pm 1.59	0.077
Mitochondrial surface irregularity index	SIi _m	$\mu\text{m}^{-1/2}$	441.2 \pm 138	513.3 \pm 136	0.023
Nucleus-to-cell centroid distance	D _{nc}	μm	0.1400 \pm 0.0477	0.1552 \pm 0.0475	0.12
Nucleus-to-cell volume ratio	Vr _{nc}	-	0.4933 \pm 0.135	0.4351 \pm 0.128	0.054
Mitochondrion-to-cell volume ratio	Vr _{mc}	-	0.0795 \pm 0.053	0.1056 \pm 0.0739	0.076

⁽¹⁾ n = number of imaged cells, p-values were obtained by a two-sample t-test method.

⁽²⁾ S = N_s·s₀ with N_s as the number of voxels on the membrane of the organelle and s₀ as the diagonal plane area of voxel.

⁽³⁾ V = N_v·v₀ with N_v as the number of voxels inside the organelle of interest and v₀ as voxel volume.

⁽⁴⁾ SIi = N_s·a₀/(V)^{1/2} with a₀ as the side length (=0.07 μm) of voxel.

Four scatter plots of the imaged prostate cells with the morphology parameters of p < 0.05 are provided in Fig.5-4 to compare their distributions. Although most of cells of the two types

overlap each other in the scatter plots, the PC3 cells cluster with smaller spreads in values of the nuclear and cellular parameters than those of the PCS cells, which is consistent with the standard deviations of most parameters in Table 5-2. In addition the PC3 cells have significantly larger cellular and nuclear volumes than PCS cells (also can be observed from the perspective views of the 3D structures of the prostate cells). It provides the insight on the ability to separate PC3 and PCS cells by the p-DIFC method.

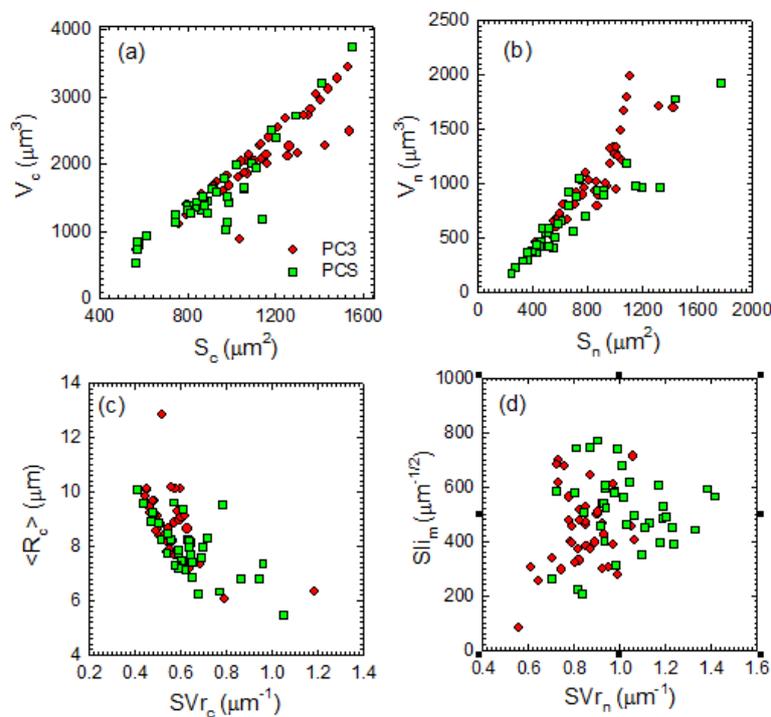


Fig. 5-4 The scatter plots of PC3 and PCS with 4 combinations of 3D parameters: (a) V_c vs S_c ; (b) V_n vs S_n ; (c) $\langle R_c \rangle$ vs SVr_c ; (d) Sll_m vs SVr_n .

Four pairs of the 3D morphologic parameters plotted in Fig. 5-4 are picked from the Table 5-2 because their p-values are less than 0.05 and thus statistically different. However in these scatter plots, one cannot find a separating line as a clear margin in the 2D parameter space to distinguish the two types of the prostate cells despite their differences are indicated

by the p-values. Most of the PC3 cells represented by the symbols of red dots and the PCS cells represented by the symbols of green dots in plots are overlapped. Even in the scatter plot of SI_{i_m} versus SV_{r_n} which has the fewest overlapped symbols of the two cell types, one could not classify these two cells. Obviously, it is very difficult to achieve classifications of high accuracy directly in the parameter space of 3D morphology even for cells of different types with relatively large differences.

5.1.3 Morphology study of the primary CD4+ and CD8+ T lymphocytes

A total of 59 CD4+ and 44 CD8+ primary T lymphocytes extracted from the human spleen tissues have been imaged with the confocal microscope and their morphology has been quantified through 3D reconstruction. Values of 17 selected 3D morphology parameters are listed in Table 5-3. Compared with the human lymphocytes derived from Jurkat and Ramos cells, the volumes of the primary T cells given in Table 5-3 are much smaller, which reveals a very interesting morphologic difference between the cultured and primary cells of the same origin. Furthermore, the marked differences illustrate clearly the effect of environment on the cell structure and phenotype development.

Table 5-3 Morphological parameters of CD4+ and CD8+ T lymphocytes

Parameter	Symbol	Unit	Mean \pm Standard Deviation		p ⁽¹⁾
			CD8+ T(n=44) ⁽¹⁾	CD4+ T(n=59) ⁽¹⁾	
Cell surface area ⁽²⁾	S _c	μm^2	231.1 \pm 66.3	186.0 \pm 62.4	0.036
Cell volume ⁽³⁾	V _c	μm^3	92.90 \pm 18.0	88.82 \pm 19.5	0.282
Cell surface to volume ratio	SVr _c	μm^{-1}	2.341 \pm 0.71	2.237 \pm 1.00	0.543
Cell surface irregularity index ⁽⁴⁾	SIi _c	$\mu\text{m}^{-1/2}$	306.6 \pm 41.6	252.5 \pm 42.4	3.85 $\times 10^{-9}$
Average distance of cell membrane voxels to centroid	$\langle R_c \rangle$	μm	3.274 \pm 1.45	3.558 \pm 1.30	0.085
Standard deviation of R _c	ΔR_c	μm	1.208 \pm 0.410	0.9290 \pm 0.217	2.20 $\times 10^{-5}$
Nuclear surface area	S _n	μm^2	196.7 \pm 69.6	166.6 \pm 58.0	0.019
Nuclear volume	V _n	μm^3	72.56 \pm 12.1	65.02 \pm 14.0	0.005
Nuclear surface to volume ratio	SVr _n	μm^{-1}	3.254 \pm 0.81	4.032 \pm 1.61	0.004
Nuclear surface irregularity index	SIi _n	$\mu\text{m}^{-1/2}$	267.3 \pm 89.9	294.7 \pm 77.1	0.099
Mitochondrial surface area	S _m	μm^2	22.79 \pm 51.6	22.93 \pm 94.1	0.933
Mitochondrial volume	V _m	μm^3	1.11 \pm 0.489	0.91 \pm 1.12	0.270
Mitochondrial surface to volume ratio	SVr _m	μm^{-1}	30.33 \pm 45.29	30.56 \pm 134.4	0.991
Mitochondrial surface irregularity index	SIi _m	$\mu\text{m}^{-1/2}$	1332 \pm 1059	494.4 \pm 1020	9.88 $\times 10^{-5}$
Nucleus-to-cell centroid distance	D _{nc}	μm	0.2235 \pm 0.118	0.2592 \pm 0.135	0.164
Nucleus-to-cell volume ratio	Vr _{nc}	-	0.7901 \pm 0.097	0.7475 \pm 0.142	0.090
Mitochondrion-to-cell volume ratio	Vr _{mc}	-	0.0111 \pm 0.008	0.0106 \pm 0.006	0.890

⁽¹⁾ n = number of imaged cells, p-values were obtained by a two-sample t-test method.

⁽²⁾ S = N_s·s₀ with N_s as the number of voxels on the membrane of the organelle and s₀ as the diagonal plane area of voxel.

⁽³⁾ V = N_v·v₀ with N_v as the number of voxels inside the organelle of interest and v₀ as voxel volume.

⁽⁴⁾ SIi = N_s·a₀/(V)^{1/2} with a₀ as the side length (=0.04 μm) of voxel.

Fig 5-5 presents two sets of CD4+ and CD8+ T lymphocytes with perspective views of the reconstructed 3D structure for each of the three cells in each set. Careful analysis of the data in Table 5-3 shows that most of the significant differences are related to the cell nucleus. The ratio of nuclear volume to cellular volume is very high (about 80%), which is consistent with the observations by hematologists leading to some differences in the shape of the cell. These results provide a useful insight on the relation between the structures and functions for the T lymphocytes. For instance, it may suggest that CD8+ T cells, which can be activated to

cytotoxic T cells, may have more activity therefore require significantly larger nuclear volumes to express molecules for performing their functions.

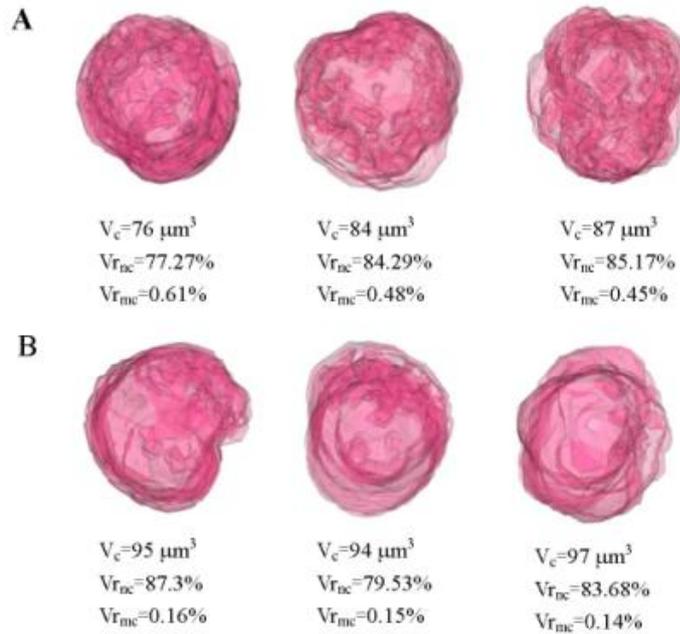


Fig. 5-5 Perspective views of reconstructed 3D structures of three CD4+ (A) and three CD8+ (B) T lymphocytes. Three parameters at the bottoms for each cell are cell volume V_c , volume ratio of nucleus-to-cell V_{rnc} , and volume ratio of mitochondria-to-cell V_{rnc} .

To clearly visualize and compare the morphological differences among six different cell types, we present in Fig. 5-6, Fig. 5-7, and Fig. 5-8 the scatter plots of selected morphology parameters of the cells investigated with the confocal imaging-based morphology quantification. In Fig. 5-6, S_c is the surface area of the cell and V_c is the volume of the cell. Although CD4+ and CD8+ T lymphocytes are overlapped with each other, they are well-separated from other cells at lower left corner, which indicates their cells' smallest surface area and volume. Jurkat and Ramos cell lines are also overlapped in S_c and V_c . But their positions are above the group of CD4+ and CD8+ T lymphocytes and under PC3 and

PCS cells. The groups of PC3 and PCS cells are widely spread in V_c and S_c and they are located on the upper right of the diagnostic line in Fig 5-6. It also shows that the cancer cell lines of human lymphocytes are much larger than primary human lymphocytes and that prostate cancer cells are larger than prostate normal cells. This allows us to conclude that although we are not able to distinguish cells within the same cell category by 3D morphology, we still have the ability to classify cells across the phenotypes and separate the cancerous lymphocytes from normal lymphocytes.

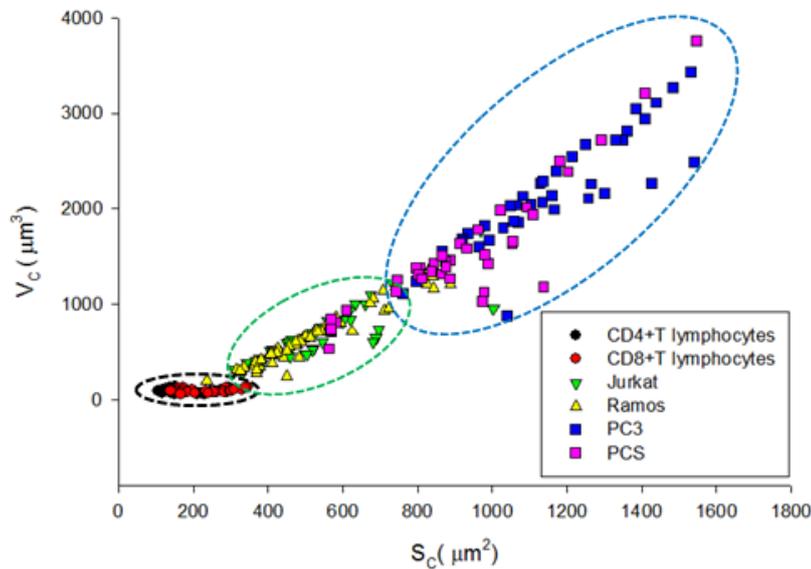


Fig. 5-6 Scatter plot of V_c versus S_c for the 6 cell lines or types

In Fig. 5-7, S_n is the surface area of the nucleus and V_n is the volume of the nucleus. CD4+ and CD8+ T lymphocytes have the smallest V_n and S_n . Meanwhile their ratios of V_c and S_c are smaller than other types of cells. We can easily distinguish CD4+ and CD8+ T lymphocytes from other cells because they are grouped at the lower left corner in the plot. PC3 and PCS make up the broad band that stretches from the upper right corner down

through the lower left corner above the group of CD4+ and CD8+ T lymphocytes. Jurkat and Ramos cells appear in the lower left parts of the scatter plot.

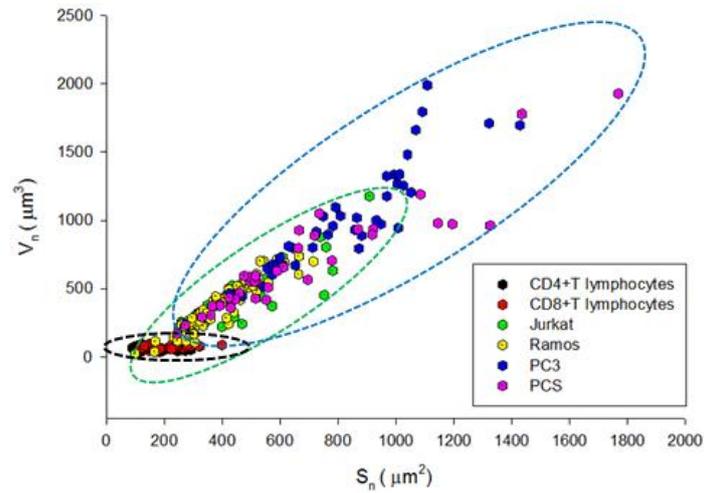


Fig. 5-7 Scatter plot of V_n versus S_n for the 6 cell lines or types as noted in the legends.

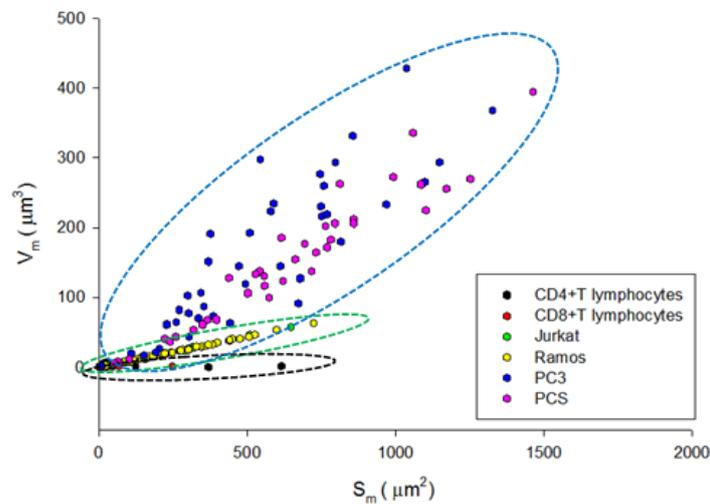


Fig. 5-8 Scatter plot of V_m versus S_m for the 6 cell lines or types as noted in the legends.

Scatter plot of volume of the mitochondria V_m versus surface area of the mitochondria S_m in Fig. 5-8 shows an interesting finding that most CD4+ and CD8+ T lymphocytes have the same ratio of V_m and S_m as Jurkat and Ramos cancer cell lines. Their ratios are significantly

different from the ratios of prostate cells. This provides strong evidence that it is possible to distinguish white blood cells from prostate epithelial cells by morphology assay.

5.2 Results of p-DIFC measurement

We have performed multiple diffraction imaging measurements on the pairs of the Jurkat versus Ramos cells and the PC3 versus PCS cells to investigate cell classification by the p-DIFC method between cells of high and moderate similarity in their 3D morphology as discussed in the previous section. Cross-polarized diffraction images of CD4+ and CD8+ T lymphocytes have also been acquired recently which are still under processing and the results will be represented elsewhere. Fig. 5-9 shows selected pairs of the cross-polarized diffraction images acquired from the Jurkat versus the Ramos cells and the PC3 versus the PCS cells. The size of speckles in PC3 and PCS images can be seen to be slightly larger than those in Jurkat and Ramos images. But the total amounts of speckles in Jurkat and Ramos images are less than those in PC3 and PCS images. With the naked eyes, one is unable to quantitatively tell the difference among these diffraction images between Jurkat and Ramos cells as well as between PC3 and PCS cells. To quantitatively extract image texture information, we have applied the GLCM based image processing software as described in Chapter 4 to obtain a total of 38 parameters from each diffraction image pair for each imaged cell, and these parameters were assembled in a feature vector in different combinations to represent this cell in a multidimensional feature space. The SVM algorithm was applied to statistically evaluate

the feature vectors with the training data and obtain an optimized model for accurate cell classification.

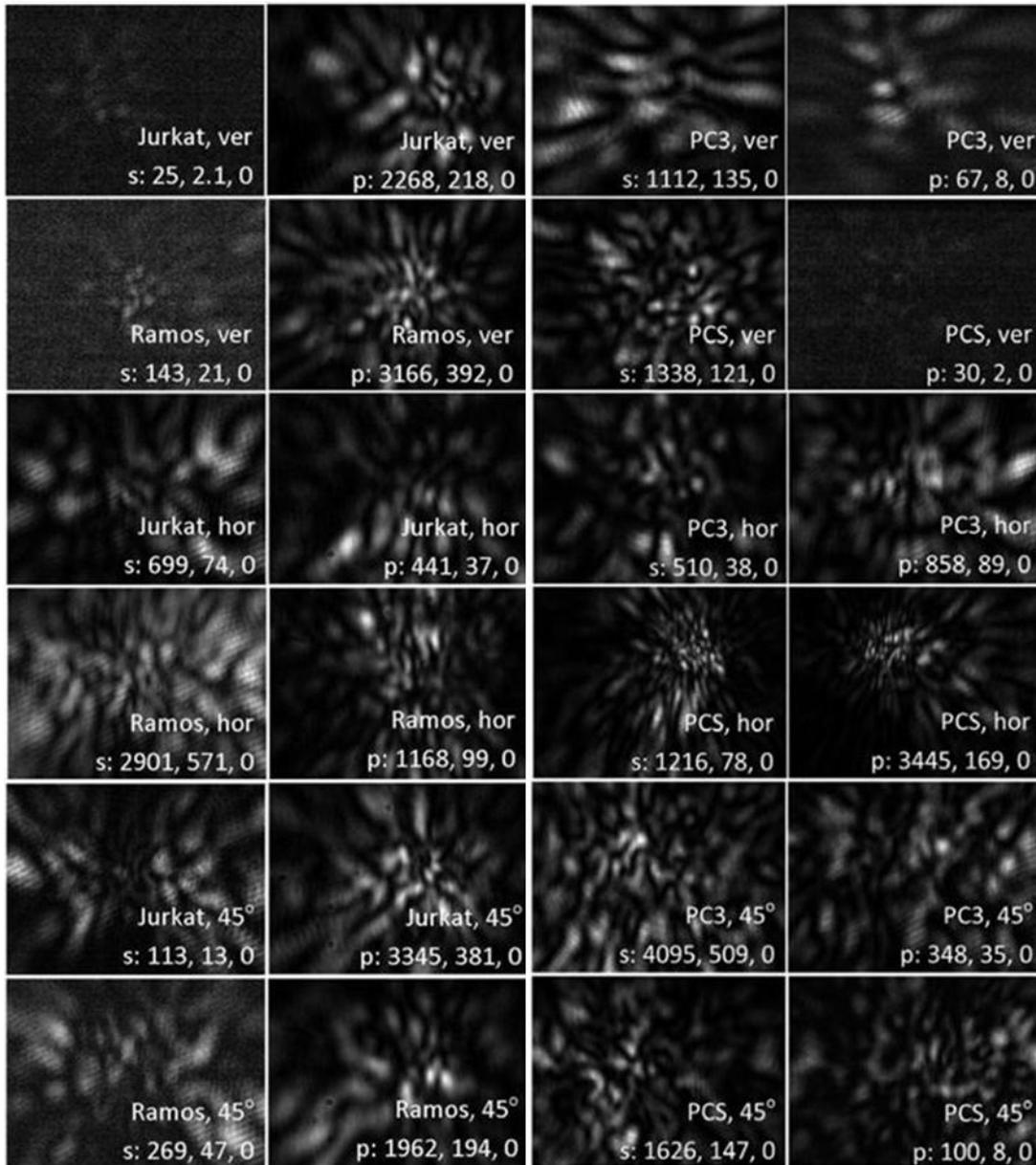


Fig. 5-9 Selected pairs of polarization-resolved diffraction images acquired from Jurkat and Ramos cells, and PC3 and PCS cells. Each image is labeled with the cell type, polarization direction of incident beam, polarization direction of scattered light, maximum, average, and minimum pixel intensities of the 12-bit diffraction image.

5.2.1 p-DIFC measurement of Jurkat T and Ramos B cell lines

After the acquisition and pre-processing of the cross-polarized diffraction image pair data, the image data deemed as those of intact cells were further processed by the GLCM based software to extract 38 image parameters for classification study. Two p-DIFC measurements of the Jurkat and Ramos cells have been carried out to examine the repeatability of the data acquisition and the result. After the acquisition and image processing, the diffraction image pair data were divided into a training data set and a test data set for each measurement. Table 5-4 lists the numbers of imaged cells or diffraction image pairs in the training and test data sets and the best accuracy of classification for Jurkat and Ramos cells using the SVM algorithm and optimized SVM models.

From the results presented in Table 5-4, one can see that polynomial kernel function has the best performance with only three to seven image parameters to form an optimized SVM model. Among the three polarization directions of the incident laser beam, horizontal and vertical polarizations provide the best accuracies with both above 97% in the data acquired in both measurements. In the data of measurement #1, the classification accuracy of horizontal polarization can reach 100%. With these results, a conclusion can be made that the accurate classification of the Jurkat and Ramos cells can be achieved robustly under the conditions of using the linear kernel function for the optimized SVM model and the vertical or the horizontal for the incident beam polarization.

Table 5-4 Experimental parameters and classification results of Jurkat and Ramos cell lines

Measurement sequence	Incident polarization	Cell type	$N_{tot}^{(1)}$	$N_{tra}^{(1)}$	$N_{tes}^{(1)}$	$A_{av}(\%)$		Kernel function and $M^{(2)}$ of best SVM model
						training	test	
#1	vertical	Jurkat	328	200	128	99.8	100	Polynomial 3
		Ramos	253	200	53			
	horizontal	Jurkat	1374	400	974	100	99.4	Polynomial 4
		Ramos	1046	400	646			
	45°	Jurkat	606	400	206	99.3	99.3	Polynomial 7
		Ramos	899	400	499			
#2	vertical	Jurkat	1630	1000	630	98.3	97.8	Polynomial 3
		Ramos	1277	1000	277			
	horizontal	Jurkat	1577	1000	577	97.9	98.1	Polynomial 4
		Ramos	1885	1000	885			
	45°	Jurkat	899	700	199	81.4	83.8	Polynomial 7
		Ramos	1530	700	830			

(1) N_{tot} = number of diffraction image pairs of viable cells for extraction of 38 image parameters; N_{tra} = number of diffraction image pairs in the training data set; N_{tes} = number of diffraction image pairs in the test data set; $N_{tes} = N_{tot} - N_{tra}$.

(2) M is the number of image parameters used in the SVM model for classification.

Two parameters p-COR, and p-CON from the best training model are selected to verify the classification of data from the measurement #1 with the incident polarization of 45 degree.

Fig. 5-10 shows the 2D scatter plot and histograms of them and one can find that Jurkat and Ramos cells can be separated in two groups in the scatter plot with these two parameters.

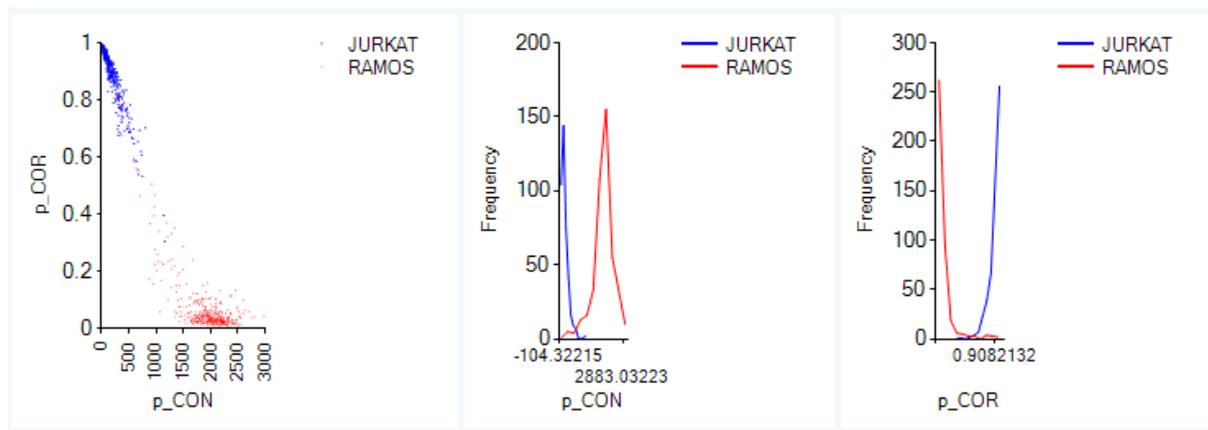


Fig.5-10 2 D scatter plot and histograms of selected parameters for Ramos and Jurkat cell lines from sequence #1. The polarization of incident beam is 45 degree. p-COR: correlation of p-polarized images; p-CON: contrast of p-polarized images.

Another parameter p-DIS is also added from the best training model to form a 3D space to verify the classification. As shown in Fig. 5-11 the distribution of red dot symbols (Ramos) and blue square symbols (Jurkat) in the 3D parameter space are separated well only in the 2D domains of p-COR and p-CON. About eight Ramos cells are overlapped with Jurkat cells in the axis of p-DIS. But the accuracy is still quite high, which can be significantly improved to the values shown in Table 5-4 by mapping the distribution of the cells in a high-dimensional feature space defined by the training data and the kernel function used in the SVM model. Therefore, classification between the Jurkat and Ramos cells can be achieved relatively well in the parameter space which can be further improved in the feature space defined by the SVM algorithm using an optimized model as shown in Fig. 5-12.

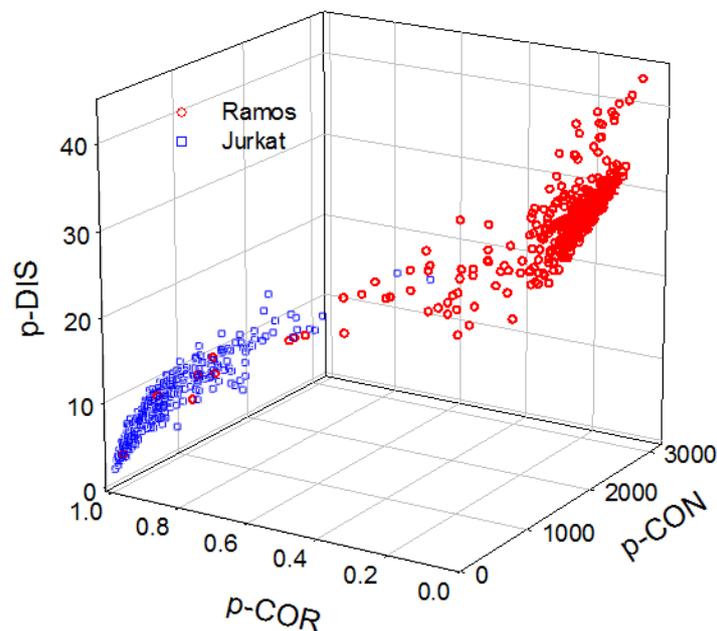


Fig.5-11 3D scatter plot of selected parameters for Ramos and Jurkat cell lines from sequence #1. The polarization of incident beam is 45 degree. p-DIS: dissimilarity of p-polarized images; p-COR: correlation of p-polarized images; p-CON: contrast of p-polarized images.

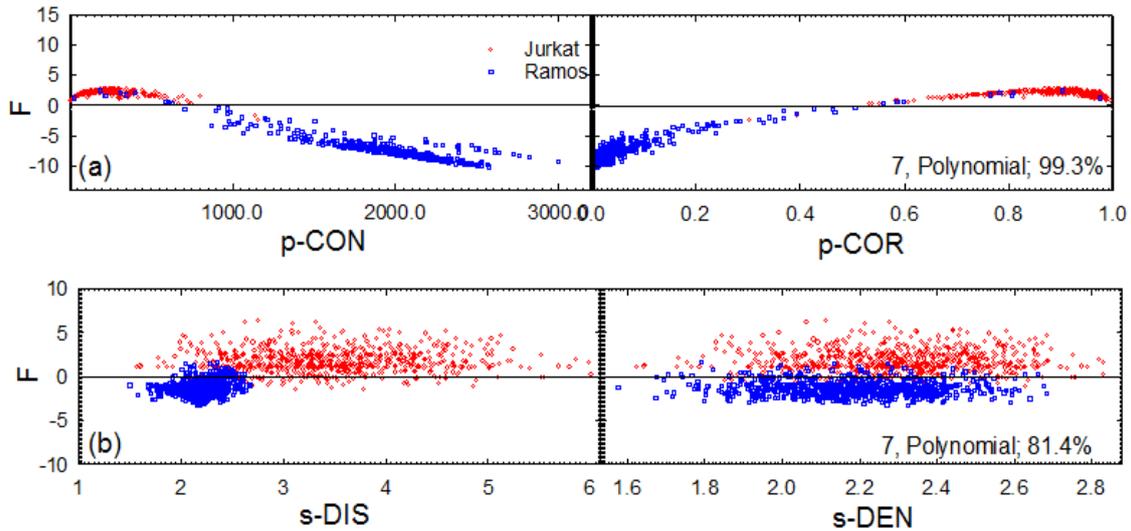


Fig.5-12 The scatter plots of training data with values of classifier F versus the top two ranked GLCM parameters used by the best SVM model established for: (a) data acquired in sequence #1 with 45 degree of incident beam; p-COR: correlation of p-polarized images; p-CON: contrast of p-polarized images; (b) data acquired in sequence #2 with 45 degree polarization; s-DEN: difference entropy of s-polarized images; s-DIS: dissimilarity of s-polarized images. The cells with $F > 0$ are classified by the SVM model as Jurkat cells and those with $F < 0$ as Ramos cells. The values of M, kernel function and Aav of the best SVM model are labeled.

5.2.2 p-DIFC measurement of the normal and cancerous prostate epithelial cells

In the previous study, we have performed the p-DIFC based classification study of the Jurkat T and Ramos B cells which have been shown to have highly similar morphology through the confocal imaging. Both of the Jurkat and Ramos cells are of white blood cell origin and are suspension cells. To fully explore the potentials of the p-DIFC method in cell assay and classification, we chose two epithelial cells derived from human prostate tissues with the cancer cell line of PC3 and normal cell type of PCS. The study of these two cell types not only provide an interesting contrast among the phenotypes of different tissue

origins but also have the practical importance in future applications of detecting circulating tumor cells in peripheral blood samples.

Three measurements have been carried out on the two prostate cell types in different weeks.

Table 5-5 provides the number of image pairs acquired and processed in the training and test data sets for PC3 and PCS cells in different measurements and corresponding results of classification with the best SVM models. From this table, one can observe that both Polynomial and Linear kernel functions perform well for classification of the two prostate cell types in the cases of horizontal and vertical polarization for the incident laser beam, which is similar as the result obtained from Jurkat and Ramos cells in comparison the case of 45° polarization. In data of measurement #2, the average accuracy for both training and test data of the incident beam as horizontal polarization are 100%.

Table 5-5 Experimental parameters and classification results of PC3 and PCS cells

Measurement Sequence	Incident polarization	Cell type	$N_{tot}^{(1)}$	$N_{tra}^{(1)}$	$N_{tes}^{(1)}$	$A_{av}(\%)$		Kernel function and $M^{(2)}$ of best SVM model
						training	test	
#1	vertical	PC3	716	500	216	99.1	97.1	Linear 10
		PCS	668	500	168			
	horizontal	PC3	681	500	181	93.7	84.5	Polynomial 10
		PCS	623	500	123			
	45°	PC3	770	300	470	80.7	64.8	Polynomial 10
		PCS	378	300	78			
#2	vertical	PC3	998	800	198	76.9	74.8	Polynomial 13
		PCS	1393	800	593			
	horizontal	PC3	890	400	490	100	100	Linear 6
		PCS	578	400	178			
	45°	PC3	897	600	297	76.3	78.2	RBF 5
		PCS	758	600	158			
#3	vertical	PC3	1130	800	330	93.5	93.0	Linear 9
		PCS	1006	800	206			
	horizontal	PC3	1104	800	304	99.5	99.5	Polynomial 14
		PCS	1337	800	537			
	45°	PC3	1137	800	337	86.0	89.0	Linear 1
		PCS	1092	800	292			

⁽¹⁾ N_{tot} = number of diffraction image pairs of viable cells for extraction of 38 image parameters; N_{tra} = number of diffraction image pairs in the training data set; N_{tes} = number of diffraction image pairs in the test data set; $N_{tes} = N_{tot} - N_{tra}$.

⁽²⁾ M is the number of image parameters used in the SVM model for classification.

We analyzed the classification result using two parameters in 2D parameter space for data of measurement # 1 as shown in Fig. 5-13. The image parameters are selected according to their single-parameter rankings determined by the SVM based classification. One can find from this figure that certain amount of PC3 and PCS cells are overlapped using only these two parameters.

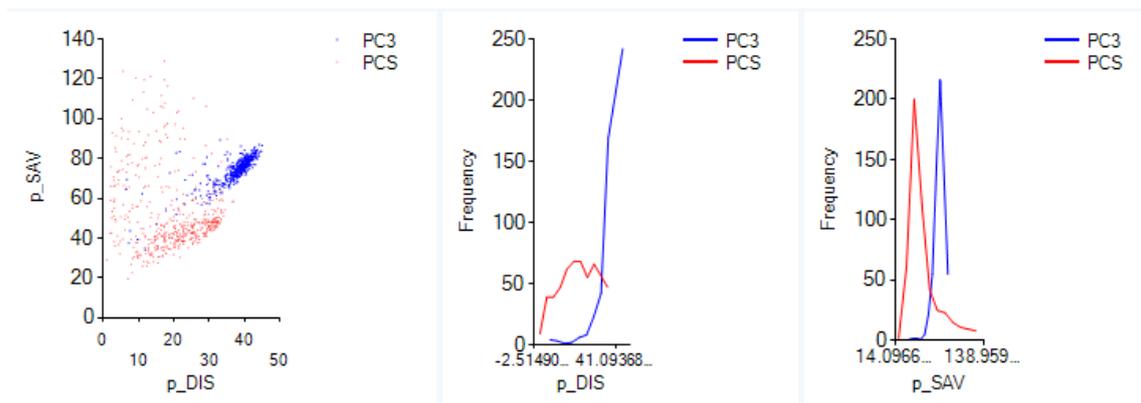


Fig.5-13 2D scatter plot and histograms of selected parameters for PC3 and PCS cells from sequence #1. The polarization of incident beam is vertical. p-DIS: dissimilarity of p-polarized images; p-SAV: sum of the average of p-polarized images.

A 3D scatter plot using top 3 parameters ranked by the SVM is in Fig. 5-14, which demonstrates clearly that the result of direct cell classification with these parameters is not satisfactory in the parameter space and most PC3 and PCS cells are mixed even using the three parameters ranked as the three best single parameters for classification. These results provide clear evidences that the SVM algorithm is more effective to classify cells. SVM improves classification by using a kernel function to map the image parameters into a higher dimensional feature space. Since the kernel functions can be nonlinear, the mapping provides a robust approach for classification with the GLCM based image parameters.

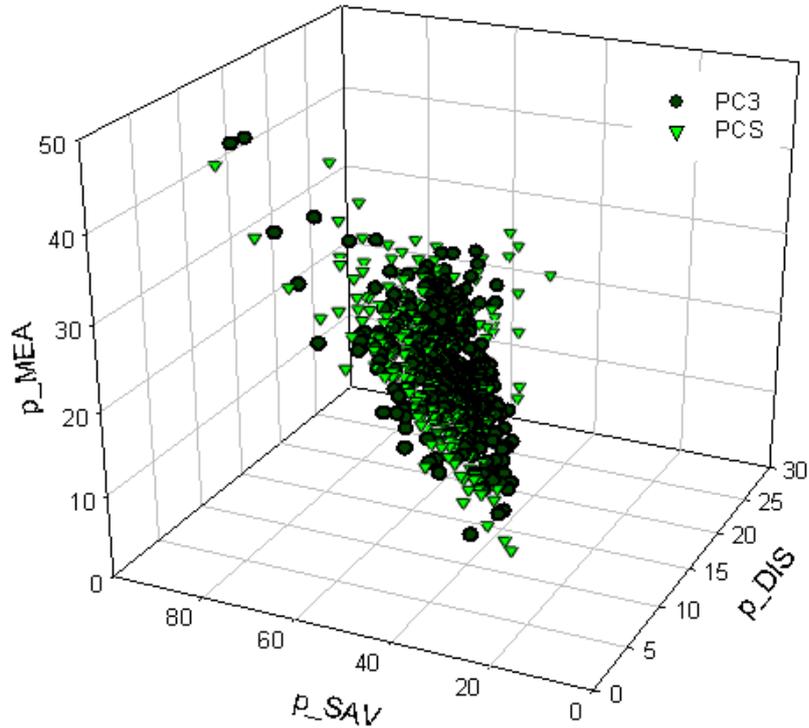


Fig.5-14 3D scatter plot of selected parameters for PC3 and PCS cells from sequence #1. The polarization of incident beam is vertical. p-DIS: dissimilarity of p-polarized images; p-SAV: sum of the average of p-polarized images; p-MEA: mean of p-polarized images .

To demonstrate the effectiveness of the SVM algorithm, scatter plots of the training result is presented in Fig.5-15 for the cell classification of sequence #1-vertical polarization and sequence #3 with the best SVM model. In Fig. 5-15 (a), the best two GLCM parameters of the dissimilarity and the sum average extracted from p-polarized images yield respectively the classification accuracy A_{av} of 91.0% and 87.7% for the training data, which are significantly smaller than the accuracy of 99.1% from the best SVM model of $M=10$ parameters and the linear kernel function. Similar improvement in classification can be observed in Fig. 5-15 (b). The plots clearly show that the SVM algorithm provides a powerful tool to improve cell classification with extracted image parameters by mapping them from the parameter space into the feature space using the kernel function. With these results, a

conclusion can be made that the accurate classification of PC3 and PCS cells can be achieved robustly using either polynomial or linear kernel function for the SVM model and the vertical or horizontal for the incident beam polarization. The best overall results of A_{av} for classifying PC3 and PCS cells can be obtained with the linear kernel function and a feature vector of six parameters from the diffraction image pairs acquired with the horizontal beam polarization.

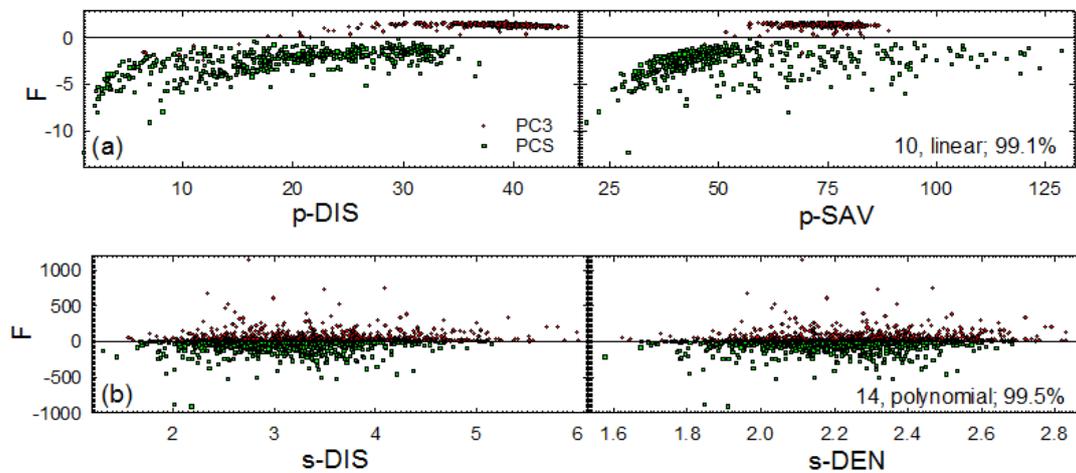


Fig.5-15 The scatter plots of training data with values of classifier F versus the top two ranked GLCM parameters used by the best SVM model established for: (a) data acquired in sequence #1 with vertical incident beam polarization; p-DIS: dissimilarity of p-polarized images, p-SAV: sum average of p-polarized images; (b) data acquired in sequence #3 with horizontal polarization; s-DEN: difference entropy of s-polarized images; s-DIS: dissimilarity of s-polarized images. The cells with $F > 0$ are classified by the SVM model as PC3 cells and those with $F < 0$ as PCS cells. The values of M, kernel function and A_{av} of the best SVM model are labeled.

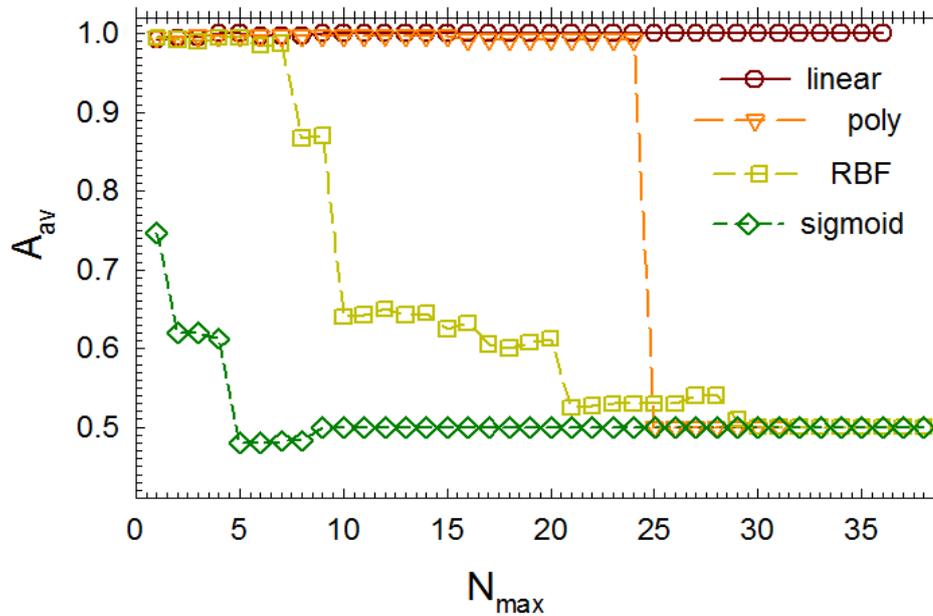


Fig.5-16 The averaged accuracy A_{av} versus the maximum number of image parameters in a feature vector N_{max} for Jurkat and Ramos classification.

The results in Fig.5-16 were obtained by performing SVM classification with four different kernel functions in the training data sets of Jurkat and Ramos cells with feature vectors constructed by the image parameters sequenced according to their rankings. The data were acquired from the first sample set with a vertically polarized incident laser beam. Linear and polynomial kernel functions have the best performance and they both have a very high accuracy with only a few parameters. But the accuracy of polynomial kernel function drops rapidly when $N_{max} > 25$. Sigmoid kernel function has the worst performance. Its highest accuracy is around 75%.

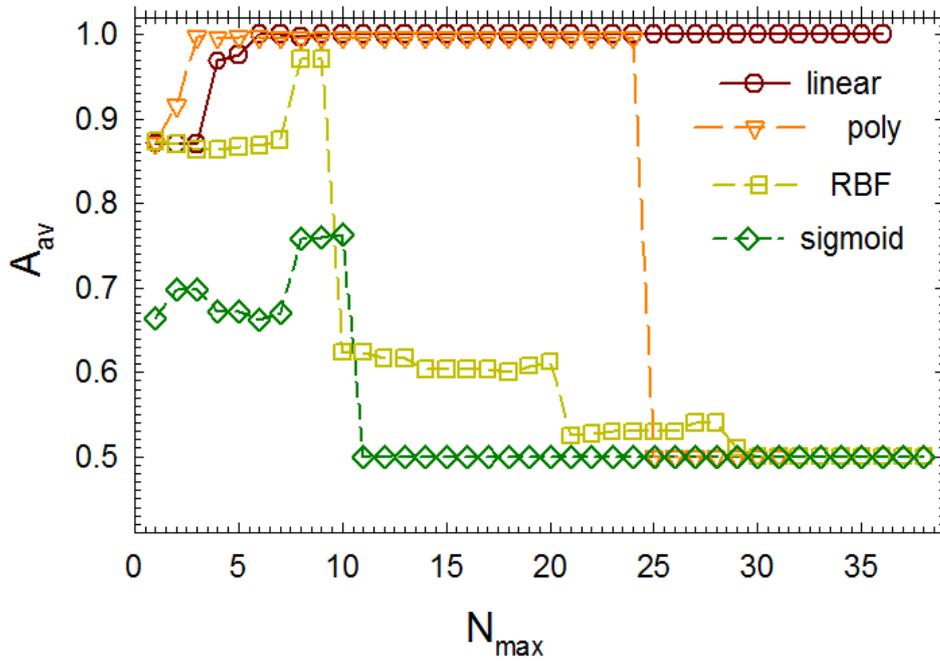


Fig.5-17 The averaged accuracy A_{av} versus the maximum number of image parameters in a feature vector N_{max} for PC3 and PCS classification.

The results in Fig.5-17 were obtained by performing SVM classification for PC3 and PCS cells with four different kernel functions in the training data set with feature vectors constructed by the image parameters sequenced according to their rankings. The data were acquired from the sequence #2 with a horizontal polarized incident laser beam. Polynomial and linear kernel functions have the best performance again. Different from Fig. 5-16, there are some build-up areas for four kernel functions. Both linear and polynomial kernel functions have the best accuracy when $N_{max} > 6$. And the accuracy of polynomial kernel functions drops when $N_{max} > 25$. Sigmoid kernel function still has the worst performance.

5.3 Discussion

T and B subpopulations of lymphocytes have been widely deemed as morphologically indistinguishable and can only be separated by the fluorescent surface marker [16], which agrees with the quantitative results of 3D morphological measurement for Jurkat and Ramos cells. Using the p-DIFC method, an optimized set of diffraction image parameters performs as the “fingerprints” encoded by the 3D morphological traits of the cells and provides high accuracy of classification. Therefore the p-DIFC method is capable to distinguish T and B subpopulations of lymphocytes. Although PC3 and PCS cells are significantly different in the size of cell and nuclear according to Table 5-2, accurate classification via statistical distribution for them cannot be achieved. As shown in Fig. 5-4, considerable overlaps exist in the distribution of the cellular and nuclear volume and surface area. Hence the differences in morphologic parameters of single cells cannot serve as effective markers for the detection of prostatic carcinoma cells. But PC3 cells can be accurately distinguished from PCS cells with the cross-polarized diffraction image pairs acquired rapidly with the p-DIFC method.

The mechanism underlying the ability of p-DIFC to distinguish the two different types of cells may be traced to the morphology of a cell in terms of its refractive index distribution. Through imaging, the difference in the intracellular refractive index distribution can be reflected by dissimilarity in the diffraction patterns quantified by the image parameters of texture and intensity. Therefore diffraction image parameters can serve as the morphologic fingerprints of the cell impressed by the coherent EM wavefields of the incident laser beam. Although these parameters from diffraction images correlate strongly with the cell

morphology, they are formed through very complex interactions of the incident EM wavefields with the molecules inside the cell. The actual intracellular distribution refractive index is still unknown, thus the detailed relations between the texture parameters extracted from a pair of 2D cross-polarized diffraction images and 3D cell morphology remains to be discovered. The results presented in this dissertation still provide strong evidences that the p-DIFC method is capable to establish an empirical approach for accurate classification of T lymphocytes and B lymphocytes as well as cancer epithelial cells and cancer cells. With the powerful machine learning application such as SVM, the diffraction image data is converted into a multidimensional feature space defined by the training data and optimized kernel functions for significantly improving the accuracy of classification as shown in Fig. 5-12.

The examination of the data for PC3 and PCS cells also proves that the scattering efficiency and the distribution of scattered light intensity between the two cross-polarized diffraction images vary dramatically among the three polarization directions of the incident beam, which can also be observed in the values of A_{av} in Table 5-5. This data clearly indicates that the diffraction image parameters carry rich information on intracellular biomolecules in terms of their ability to polarize in the wavefields of the incident beam. As a consequence the p-DIFC method's capability to accurately separate two cell types without the need of extraneous labeling relies not only on the 3D morphology but also on the molecular response to the incident wavefields. Among three incident polarization direction, 45° tends to produce smaller values of A_{av} for each of the three measurements. Similar results have been observed in Table 5-4 for Jurkat and Ramos cells. The reduction of A_{av} with polarization at

45° can be explained as following considerations. If incident beam is propagating along the z-axis with polarization at 45°, the intracellular molecules can induce dipoles to oscillate along any direction within the x-y plane (The coordinate is defined in Fig. 4-9). The probability of inducing molecular dipoles along the horizontal direction of x-axis is the same as the one along the vertical direction of the y-axis, which reduces the selectivity of the p-DIFC method to contrast the differences among cells with different types of molecular response to the incident wavefields. The cross-polarized diffraction images in Fig.5-9 also verify this explanation although the image number is limited.

CHAPTER 6 SUMMARY

The research projects described in this dissertation provide quantitative results of investigations on 3D morphology among six types of human cells of different origins of tissues. More importantly, the dissertation research focuses on the exploration of a new approach within the platform of flow cytometry technology through imaging of highly coherent scattered light for label-free cell assay and classification.

The morphology investigations have been carried out with confocal microscopy to acquire image stack data of single cells and perform 3D reconstruction by a Matlab-based software for segmentation and interpolation. A total of 29 morphology parameters related to the cell, the nucleus, and the mitochondria were calculated for quantitative analysis and comparison among different phenotypes on the basis of the two-samples t-test.

Multiple software components have been developed through this dissertation research to improve an experimental system to study the potentials of the p-DIFC method for rapid acquisition of cross-polarized diffraction image pairs from single cells excited by a linearly polarized laser beam. The speed of image acquisition has been improved from less than 1 image pair per second to about 10 image pairs per second by applying a multi-threading mechanism in the camera control software module. Image preprocessing has also been automated with an in-house developed software component to increase the speed of image data analysis after data acquisitions. An existing GLCM algorithm has been implemented into

the image processing software to obtain a total of 38 image texture and intensity parameters for each imaged cell to quantify the diffraction patterns by the cell. To perform an automated cell classification, a software has been developed using the SVM algorithm to map the image parameters into the feature space formed by the training data of the cross-polarized diffraction image pairs.

The cells that have been investigated include Jurkat T cell lines, Ramos B cell lines, CD4+ T lymphocytes, CD8+ T lymphocytes, normal prostate epithelial cells, and cancerous prostate epithelial cells. A careful morphology study on the Jurkat T and Ramos B cell lines showed that their 3D parameters are highly similar except the nucleus-to-cell volume ratio and therefore they are indistinguishable in morphology by the traditional microscopy. But using p-DIFC method, accurate classification can be achieved with the accuracy ranging from 97% to 100% with optimized SVM models consisting of the polynomial kernel function and 4 GLCM image parameters. These results provide strong evidences that by detecting the intracellular refractive index distribution the p-DIFC approach is capable of detecting the subtle morphologic differences that are difficult to quantify by the traditional microscopy measurement.

The morphology study on the normal and cancerous prostate epithelial cells shows that the cancerous prostate epithelial cells tend to have larger sizes in the cell and nucleus than the normal prostate epithelial cells. These differences are quite obvious in the statistical analysis of the mean value and the standard deviation. Still the scatter plots of those parameters with p-values less than 0.05 demonstrate that morphology parameters alone can hardly be used for

effective classification of the two cell types. The p-DIFC measurement has been applied on the prostate epithelial cells, and it has been shown that accurate classification of the two cell types can also be achieved with the accuracies ranging from 97% to 100% using optimized SVM models of either linear or polynomial kernel functions and up to 10 GLCM image parameters. These results prove again that the p-DIFC method has the ability to distinguish between the two types of cells, which have larger morphologic differences in comparison to the Jurkat and Ramos cells.

We have also carried out the morphology study of primary human CD4+ and CD8+ T lymphocytes extracted from spleen tissues and the preliminary measurement of the cross-polarized diffraction images (not included in this dissertation). Based on the p-values, the major difference between these two T cell subtypes lies in the volume of the nucleus in which the CD8+ T lymphocytes have a larger volume on average. The diffraction images acquired with the CD4+ and CD8+ T lymphocytes sorted from the extracted splenocytes shows more noises and larger variations in diffraction patterns as compared with the Jurkat and Ramos cells. A study is underway to develop improved image analysis algorithms that would allow for the separation of different pattern types from the acquired data and results will be reported elsewhere.

With the 3D parameters, we were able to evaluate the morphologic differences among the six phenotypes of prostate epithelial cells, cell lines derived from lymphocytes, and primary lymphocytes for the first time. By pooling the morphologic parameters of all of the cells imaged by the confocal microscopy method, the cell volume and nuclear volume of the T

lymphocytes cell lines can be observed to be much larger than those of primary T lymphocytes, which can be used to understand the difficulty in analyzing the diffraction image data acquired with primary T cells. Similar differences are found between the normal and cancerous prostate epithelial cells in which the formers tend to have smaller cell and nuclear sizes than the cancer cell line. These findings provide insights on the structural differences among the cell types and a basis to understand the ability of p-DIFC method for morphology based cell classification. Another significant finding of this dissertation research is the statistically significant difference in the size of nucleus between CD4+ and CD8+ T lymphocytes, which could be attributed to their distinct functions after activation.

With the cross-polarized diffraction image data, the p-DIFC method has been shown to have the ability to accurately classify the Jurkat T cell line from the Ramos B cell line. Their morphology was quantified to be highly similar and cannot be distinguished without staining with different CD markers. The similar measurements and classification studies on the two types of cancerous and normal prostate epithelial cells have also demonstrated the ability of the p-DIFC method to accurately differentiate them from the cross-polarized diffraction image data. These results lead to the conclusion that the p-DIFC method has significant potential to be developed into a rapid and label-free method for cell assay and morphology based classification to discriminate white blood cells and epithelial cells of both high similarity and moderate similarity in their morphology.

We should note that the p-DIFC method requires significant improvement before it can be established as an effective tool for cell assay and phenotyping. Based on the results of this

dissertation research, it becomes clear that one of the research tasks of high priority for future improvement is to prove and establish quantitative relations between the 3D cell morphologic parameters and the diffraction pattern features of the diffraction image texture parameters.

Previously, numerical analysis has been performed by our research group on the effect of the nuclear and mitochondrial morphology on the patterns of diffraction images with finite-difference-time-domain (FDTD) and discrete-dipole-approximation (DDA) models of light scattering [23,45]. The simulation results indicate that the difference in the nuclear volume or the volume ratio of nucleus-to-cell can lead to obvious changes in the GLCM parameters extracted from the diffraction images. However, the correlations of the changes between the nuclear morphology and GLCM parameters are convoluted among the volume, shape, and values of refractive index heterogeneity of the organelles inside the cell. A detailed and large-scale numerical study needs to be pursued in the near future to extract quantitative relations which can be used to develop a comprehensive method for extracting 3D morphology information from the diffraction image data and understand the morphology foundations of the p-DIFC method. Another issue is related to the application of the p-DIFC method to classify different types of cells in multiple measurements. For example, its capability of distinguishing three or more than three types of cells remains unknown and needs to be carefully examined. Moreover, current texture analysis algorithm GLCM requires significant computational cost for extraction of image parameters which are very sensitive to the minor changes in the optic settings due to ambient light noise and variation of the cell sample conditions. Therefore better algorithms for processing and characterization of image

textures needs to be developed for increasing the speed of calculation and reducing the sensitivity to noise. Hardware-wise, the current p-DIFC system needs to be significantly improved as a semi-automatic system without too many manual control and alignment requirements and the system must increase the repeatability of data acquisition and subsequent analysis.

REFERENCES

- [1] Shapiro, Howard M. "Multistation multiparameter flow cytometry: a critical review and rationale." *Cytometry* 3.4 (1983): 227-243.
- [2] Lievens, L. A., et al. "Tumor-associated macrophages in thoracic malignancies." *Lung Cancer* 80.3 (2013): 256-262.
- [3] Allard, W. Jeffrey, et al. "Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases." *Clinical Cancer Research* 10.20 (2004): 6897-6904.
- [4] Hacker, Jörg, and James B. Kaper. "Pathogenicity islands and the evolution of microbes." *Annual Reviews in Microbiology* 54.1 (2000): 641-679.
- [5] Suratt, Benjamin T., et al. "Neutrophil maturation and activation determine anatomic site of clearance from circulation." *American Journal of Physiology-Lung Cellular and Molecular Physiology* 281.4 (2001): L913-L921.
- [6] Gest, Howard. "The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society." *Notes and Records of the Royal Society* 58.2 (2004): 187-201.
- [7] Johannsen, Wilhelm. "The genotype conception of heredity." *International journal of epidemiology* 43.4 (2014): 989-1000.

- [8] Johnsen, Sönke, and Edith A. Widder. "The physical basis of transparency in biological tissue: ultrastructure and the minimization of light scattering." *Journal of theoretical biology* 199.2 (1999): 181-198.
- [9] Drezek, Rebekah, Andrew Dunn, and Rebekah Richards-Kortum. "Light scattering from cells: finite-difference time-domain simulations and goniometric measurements." *Applied Optics* 38.16 (1999): 3651-3661.
- [10] Ruban, Gennady I., et al. "Light scattering and morphology of the lymphocyte as applied to flow cytometry for distinguishing healthy and infected individuals." *Journal of biomedical optics* 15.5 (2010): 057008-057008.
- [11] Melamed, Myron R., et al. "Flow cytometry and sorting." *American Journal of Clinical Oncology* 14.1 (1991): 90.
- [12] Forrest, J. W., H. W. Straat, and R. K. Dakin. "Autocollimating Spectrometer." *JOSA* 46.2 (1956): 143-143.
- [13] Steiger, James H., Alexander Shapiro, and Michael W. Browne. "On the multivariate asymptotic distribution of sequential chi-square statistics." *Psychometrika* 50.3 (1985): 253-263.
- [14] Armstrong, D. T., and William Hansel. "Alteration of the bovine estrous cycle with oxytocin." *Journal of Dairy Science* 42.3 (1959): 533-542.
- [15] Herzenberg, Leonard A., Richard G. Sweet, and Leonore A. Herzenberg. "Fluorescence-activated cell sorting." *Sci Am* 234.3 (1976): 108-117.

- [16] Wintrobe, Maxwell Myer. *Wintrobe's clinical hematology*. Ed. John P. Greer. Vol. 1. Lippincott Williams & Wilkins, 2009.
- [17] Ong, S. H., et al. "Development of an imaging flow cytometer." *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology* 9.5 (1987): 375-382.
- [18] George, Thaddeus C., et al. "Distinguishing modes of cell death using the ImageStream® multispectral imaging flow cytometer." *Cytometry Part A* 59.2 (2004): 237-245.
- [19] Holler, Stephen, et al. "Two-dimensional angular optical scattering for the characterization of airborne microparticles." *Optics letters* 23.18 (1998): 1489-1491.
- [20] Neukammer, Jörg, et al. "Angular distribution of light scattered by single biological cells and oriented particle agglomerates." *Applied optics* 42.31 (2003): 6388-6397.
- [21] Jacobs, Kenneth M., Jun Q. Lu, and Xin-Hua Hu. "Development of a diffraction imaging flow cytometer." *Optics letters* 34.19 (2009): 2985-2987.
- [22] Jacobs, Kenneth M., et al. "Diffraction imaging of spheres and melanoma cells with a microscope objective." *Journal of biophotonics* 2.8-9 (2009): 521-527.
- [23] Dong, Ke, et al. "Label-free classification of cultured cells through diffraction imaging." *Biomedical optics express* 2.6 (2011): 1717-1726.
- [24] Yu, Sa, et al. "A novel method of diffraction imaging flow cytometry for sizing microspheres." *Optics express* 20.20 (2012): 22245-22251.

- [25] Kaighn, M. E., et al. "Establishment and characterization of a human prostatic carcinoma cell line (PC-3)." *Investigative urology* 17.1 (1979): 16-23.
- [26] Maton, Anthea. *Human biology and health*. Prentice Hall, 1997.
- [27] Hollowell, J. G., et al. "Hematological and iron-related analytes--reference data for persons aged 1 year and over: United States, 1988-94." *Vital and health statistics. Series 11, Data from the national health survey* 247 (2005): 1-156.
- [28] Schneider, Ulrich, Hans-Ulrich Schwenk, and Georg Bornkamm. "Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma." *International journal of cancer* 19.5 (1977): 621-626.
- [29] Sato, Eiichi, et al. "Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer." *Proceedings of the National Academy of Sciences of the United States of America* 102.51 (2005): 18538-18543.
- [30] Aerts, Joachim G., and Joost P. Hegmans. "Tumor-specific cytotoxic T cells are crucial for efficacy of immunomodulatory antibodies in patients with lung cancer." *Cancer research* 73.8 (2013): 2381-2388.
- [31] Rosenberg, Steven A. "Cell transfer immunotherapy for metastatic solid cancer—what clinicians need to know." *Nature reviews Clinical oncology* 8.10 (2011): 577-585.
- [32] Haralick, Robert M. "Statistical and structural approaches to texture." *Proceedings of the IEEE* 67.5 (1979): 786-804.

- [33] Haralick, Robert M., Karthikeyan Shanmugam, and Its' Hak Dinstein. "Textural features for image classification." *Systems, Man and Cybernetics, IEEE Transactions on* 6 (1973): 610-621.
- [34] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: A library for support vector machines." *ACM Transactions on Intelligent Systems and Technology (TIST)*2.3 (2011): 27.
- [35] Liu, XiaoHang, Keith Clarke, and Martin Herold. "Population density and image texture." *Photogrammetric Engineering & Remote Sensing* 72.2 (2006): 187-196.
- [36] Gong, Peng, Danielle J. Marceau, and Philip J. Howarth. "A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data." *Remote sensing of environment* 40.2 (1992): 137-151.
- [37] Baraldi, Andrea, and Flavio Parmiggiani. "An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters." *Geoscience and Remote Sensing, IEEE Transactions on* 33.2 (1995): 293-304.
- [38] Unser, Michael. "Sum and difference histograms for texture classification." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 1 (1986): 118-125.
- [39] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: A Library for Support Vector Machines. Department of Computer Science and Information Engineering, National Taiwan University." (2001).
- [40] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20.3 (1995): 273-297.

- [41] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers." *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992.
- [42] Suykens, Johan AK, and Joos Vandewalle. "Least squares support vector machine classifiers." *Neural processing letters* 9.3 (1999): 293-300.
- [43] Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003).
- [44] Keerthi, S. Sathiya, and Chih-Jen Lin. "Asymptotic behaviors of support vector machines with Gaussian kernel." *Neural computation* 15.7 (2003): 1667-1689.
- [45] Zhang, Jun, et al. "Analysis of cellular objects through diffraction images acquired by flow cytometry." *Optics express* 21.21 (2013): 24819-24828.
- [46] Feng, Yuanming, et al. "Polarization imaging and classification of Jurkat T and Ramos B cells using a flow cytometer." *Cytometry Part A* 85.9 (2014): 817-826.
- [47] Kanopoulos, Nick, Nagesh Vasanthavada, and Robert L. Baker. "Design of an image edge detection filter using the Sobel operator." *Solid-State Circuits, IEEE Journal of* 23.2 (1988): 358-367.
- [48] Kanungo, Tapas, et al. "An efficient k-means clustering algorithm: Analysis and implementation." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7 (2002): 881-892.

- [49] Smith, A. M. S., et al. "Texture based feature extraction: application to burn scar detection in Earth observation satellite sensor imagery." *International Journal of Remote Sensing* 23.8 (2002): 1733-1739.
- [50] Mohanty, Aswini Kumar, Swapnasikta Bebertha, and Saroj Kumar Lenka. "Classifying benign and malignant mass using GLCM and GLRLM based texture features from mammogram." *International Journal of Engineering Research and Applications* 1.3 (2011): 687-693.

APPENDIX A DIFFRACTION IMAGE PARAMETERS

DEFINITION

There are 19 parameters for each image including 17 GLCM parameters and 2 intensity parameters. The parameter names, symbol, and their definition equation is in the following table.

Parameter	Symbol	Definition Equation
Correlation	COR	$\frac{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu_x)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y}$ $= \frac{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Dissimilarity	DIS	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} i - j p(i, j)$
Contrast (or inertia)	CON	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - j)^2 p(i, j)$
Inverse Difference Moment (or local homogeneity)	IDM	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i - j)^2} p(i, j)$
Entropy	ENT	$-\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i, j) \cdot \log(p(i, j))$
Sum entropy	SEN	$-\sum_{k=0}^{2G-2} p_{x+y}(k) \cdot \log(p_{x+y}(k))$
Difference entropy	DEN	$-\sum_{k=0}^{G-1} p_{x-y}(k) \cdot \log(p_{x-y}(k))$
ASM (Angular Second Moment or energy or homogeneity)	ASM	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{p(i, j)\}^2$
Variance	VAR	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 p(i, j)$
Sum variance	SVA	$\sum_{k=0}^{2G-2} (k - SEN)^2 p_{x+y}(k)$
Difference variance	DVA	$\frac{1}{G-1} \sum_{k=0}^{G-1} (p_{x-y}(k) - \bar{p}_{x-y})^2$

Mean	MEA	$\sum_{i=0}^{G-1} i \sum_{j=0}^{G-1} p(i, j) = \mu_x$
Sum average	SAV	$\sum_{k=0}^{2G-2} k p_{x+y}(k)$
Cluster shade	CLS	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i + j - 2\mu)^3 p(i, j)$
Cluster prominence	CLP	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i + j - 2\mu)^4 p(i, j)$
Maximum probability	MAP	$\max(p(i, j))$
Minimum probability	MIP	$\min(p(i, j))$
Relative minimum pixel intensity	IMIN	$\min(J(x, y)) / \text{mean}(J(x, y))$
Relative maximum pixel intensity	IMAX	$\max(J(x, y)) / \text{mean}(J(x, y))$

APPENDIX B PROTOCOL OF DETACHING CELLS

This protocol is used for detaching cells from the culture plates. PC3 and PCS cells are attached to the plate.

1. Aliquot out enough media and warm up the media in the 37 degree water bath.
2. When the media is warmed up, get ready of the DPBS and Trypsin-EDTA.
3. Take out the plate to be detached from the incubator, and suck out the media.
4. Add 8-10 ml of DPBS to the plate to wash once (removing the FBS in the media), and then suck out the DPBS.
5. Add 2ml of Trypsin-EDTA to each plate, and put into the incubator; let it incubate for 3-5 minutes.
6. During that 3-5 minutes, take the plate out, tap the plate from the side to break up the cell clusters, and watch under the microscope to make sure cells are detached.
7. In the hood, add 5ml of the media to the plate and pipette up and down several times to break the cell clusters and collect the cells to the 15 ml tube.
8. Spine down for 5 minutes. Take out the plate and other tubes needed and label them.

9. After the 5 minutes centrifugation, suck out the media, pipette up and down several times to break the cell clusters.
10. Count the cell number under the microscope. And calculate how much volume of the cells is required to take out.
11. Put some media in the new plate first and add the required amount of cells into the plate. Mix them well, tap the plate, and put into the incubator.
12. For the 24- or 96-well plate, usually prepare the fresh media and cells in another tube first, and aliquot into the wells.

APPENDIX C SPLENOCYTES ISOLATION

PROCEDURE

This protocol is for acquiring the splenocytes from the spleen tissue. The key is to keep the tissue on ice during transportation and use caution to avoid split during operation.

1. Keep the tissue sample on ice for transportation.
2. Cut the tissue into small pieces with medical scissors in RPMI 1640 medium.
3. Generate a single cell suspension of splenocytes by grinding all small tissue pieces gently with two pieces of frosted glass slides in RPMI medium.
4. Filter cells using 70 μm cell strainer to get rid particles such as fat in the suspension.
5. Centrifuge the cell suspension in 1500 RPM for 5 minutes and discard the supernatant.
6. Add 10 ml red blood cell lysis buffer and shake at room temperature for 10 minutes to remove the red blood cells.
7. Wash cells PBS/BSA buffer and count them.

APPENDIX D IMMUNOFLUORESCENCE STAINING

PROCEDURE

This protocol is for staining cells for cell sorter to get the specific cells. Make sure read the manual of the CD marker before the experiment. The proper amount of CD marker (from Life Technology) to use in Step 3 is 5 μl / million cells.

1. Pellet cells and spin them at 1500 rpm for 5 minutes. Adjust the cell suspension to a concentration of 5×10^6 cells/ml with PBS/BSA buffer (phosphate buffered saline pH 7.4 and 1% BSA).
2. Aliquot 100 μl of cell suspension into as many test tubes as required.
3. Add conjugated antibody at the recommended dilution from the manufacturer. Mix well and incubate at on ice for 30 minutes. Include an unstained control for cytometer set-up.
4. Wash cells with 2 ml of PBS/BSA, centrifuge at 1500 rpm for 5 minutes and discard the supernatant.
5. Resuspend cells in 0.3 ml of PBS/BSA or with 0.2 ml of 0.5% paraformaldehyde in PBS/BSA if required.

APPENDIX E PROTOCOL OF STAINING WITH SYTO 61 AND MITOTRACKER ORANGE

The double staining protocol is for acquiring confocal image stack. Primary cells are easily to bleach or to lose the signals after long-time staining. Therefore the incubation time in step 8 can be shorter than 30 minutes for primary cells.

1. Collect cells in 15mL conical tube from the cell culture.
2. Spin cells at 1500 rpm for 5 min.
3. Aspirate the media on top of cells to obtain a cell pellet.
4. Break up cell pellet by tapping the tube and resuspend cells by adding 5mL of culture media.
5. Pipette the cell suspension several times to make sure a single cell suspension.
6. Add 1 μ L of Syto-61 and MitoTracker Orange stock solution to the 5mL cell suspension so the final concentration is 1 μ M for Syto-61 and 0.2 μ M for MitoTracker Orange - start of staining.
7. Invert tube several times to mix media well.

8. Incubate at 37deg C and 5% CO2 for 30 min.
9. Spin cells at 1500 rpm for 3 min.
10. Break up cell pellet and resuspend cells in 2mL of media as previous procedure – start of 1st wash.
11. Incubate at 37°C and 5% CO2 for at least 5 min.
12. Spin cells at 1500 rpm for 3 min.
13. Aspirate the media on top of cells to obtain a cell pellet, and this is the end of 1st wash.
14. If too much fluorescence background appears in the confocal images, a 2nd wash may be taken by repeating above steps.
15. Resuspend pellet in 0.5 to 1mL media to have high cell concentration for confocal observation –cell suspension for imaging.
16. Add 150 µL to depression slide, put a glass cover slide on top and invert the assembly for the inverted microscope viewing. If cells are too close to each other, dilute cell suspension with more media.

APPENDIX F CELL COUNTING PROTOCOL

Cell counting should be applied after cell isolation or before staining. Also it is necessary to count the cell after sorting to estimate the viability of cells.

1. Ensure the hemocytometer is clean using 70% ethanol.
2. Make sure the cell suspension to be counted is well mixed by either gentle agitation of the tube containing the cells (or other appropriate container). A serological pipette may be used if required.
3. Using a 100 μ L pipette, take out the cells and drop them in a small tube (gently to avoid lysing them). Take out 100 μ L trypan blue and mix gently.
4. Using a Gilson pipette, draw up some cell suspension containing trypan blue.
Carefully fill the hemocytometer by gently resting the end of the Gilson tip at the edge of the chambers. Take care not to overfill the chamber. Allow the sample to be drawn out of the pipette by capillary action, the fluid should run to the edges of the grooves only. Re-load the pipette and fill the second chamber if required.
5. Focus on the grid lines of the hemocytometer using the 10x objective of the microscope. Focus on one set of 16 corner squares.
6. Using a hand tally counter, count the number of cells in this area of 16 squares. When counting, always count only live cells that look healthy. Count cells that are within the

square and any positioned on the right hand or bottom boundary line. Dead cells stained blue with trypan blue can be counted separately for a viability count.

7. Move the hemocytometer to another set of 16 corner squares and carry on counting until all 4 sets of 16 corner squares are counted.
8. The hemocytometer is designed so that the number of cells in one set of 16 corner squares is equivalent to the number of cells $\times 10^4/\text{mL}$. Therefore, to obtain the count, and the total count from 4 sets of 16 corner = $(\text{cells}/\text{mL} \times 10^4) \times 4$ squares from one hemocytometer grid.
9. So divide the count by 4. Then multiply by 2 to adjust for the 1:2 dilution in trypan blue. These two steps are equivalent to dividing the cell count by 2. As an example: if the cell count is 145, the cell density is: $145/2 = 72.5 \times 10^4/\text{mL}$.