UNIVERSITY OF CALIFORNIA SAN DIEGO

Image-Guided Cell Classification and Sorting

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Photonics)

by

Yi Gu

Committee in charge:

Professor Yuhwa Lo, Chair
Professor Zhaowei Liu, Co-Chair
Professor Joseph Ford
Professor Tina Ng
Professor Kun Zhang

2019

The Dissertation of Yi Gu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California San Diego

2019

DEDICATION

To my family and loved ones…

EPIGRAPH

I have not failed. I've just found 10,000 ways that won't work.

Thomas Edison

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First of all, I would like to acknowledge my advisor Professor Yu-Hwa Lo for his guidance, support and encouragement as the chair of committee. This work could not possibly be accomplished without his tremendous input. I would like to thank my committee members, Professor Zhaowei Liu, Professor Joseph Ford, Professor Tina Ng, and Professor Kun Zhang for their time to give me advice on my project and read my dissertation.

I also wish to thank all the other Lo group members, Yuanyuan Han, Alex Ce Zhang, Xinyu Chen, Rui Tang, Zunming Zhang, Jiajie Chen, Violet Castor, Jill Tsai, Edward Wang, Roger Chiu, Tony Yen, Wei Cai, Pingwei-Chen, Brian Lewis, Chi-Yang Tseng, Ramkumar Subramanian, Lujiang Yan, Yugang Yu, Jiayun Zhou, Iftikhar Ahmad Niaz, Zihan Xu, YuHsin Liu, David Hall, for their valuable assistance.

Last but not least, I wish to thank my wife Siyan, my son Aaron, my parents and all my fiends for their support and encouragement.

Chapter 2 is based on and mostly a reprint of: Alex Ce Zhang, **Yi Gu(co-first author)**, Yuanyuan Han, Zhe Mei, Yu-Jui Chiu, Lina Geng, Sung Hwan Cho, and Yu-

Hwa Lo. "Computational cell analysis for label-free detection of cell properties in a microfluidic laminar flow." **Analyst** 141, no. 13 (2016): 4142-4150.

Chapter 3 is based on and mostly a reprint of: **Yi Gu**, Alex Ce Zhang, Yuanyuan Han, Jie Li, Clark Chen, and Yu-Hwa Lo. "Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification." **Cytometry Part A** (2019); **Yi Gu**, Rui Tang, Alex Ce Zhang, Yuanyuan Han and Yu-Hwa Lo, "Image-guided microfluidic cell sorter with machine learning", **CLEO** (2019)， Control Number: 3080140; Yuanyuan Han, **Yi Gu**, Alex Ce Zhang, and Yu-Hwa Lo. "Imaging technologies for flow cytometry." **Lab on a Chip** 16, no. 24 (2016): 4639-4647.

Chapter 4 is based on and mostly a reprint of: **Yi Gu**, Xinyu Chen, Sung Hwan Cho, Chang-Hung Lee, Ivan Gagne, Alex Ce Zhang, Rui Tang, Yuanyuan Han, and Yu-Hwa Lo. "Image-Guided Cell Sorting Using Laser Point Scanning" in preparation

The dissertation author was the primary author of the work in these chapters, and coauthors have approved the use of the material for this dissertation.

Yi Gu

La Jolla, CA

March 2019

VITA

EDUCATION

2008 – 2012	Bachelor of Engineering in Electrical Engineering, Shanghai Jiao Tong University, China
2012 – 2014	Master of Science in Applied Physics, Cornell University, Ithaca, USA
2014 – 2019	Ph. D. in Electrical Engineering (Photonics), University of California, San Diego, USA

PUBLICATIONS

Yi Gu, Guangqiang He, and Xufei Wu. "Generation of six-partite continuous-variable entanglement using a nonlinear photonic crystal by frequency conversions." **Physical Review A** 85, no. 5 (2012): 052328.

Alex Ce Zhang, **Yi Gu(co-first author)**, Yuanyuan Han, Zhe Mei, Yu-Jui Chiu, Lina Geng, Sung Hwan Cho, and Yu-Hwa Lo. "Computational cell analysis for label-free detection of cell properties in a microfluidic laminar flow." **Analyst** 141, no. 13 (2016): 4142-4150.

Yuanyuan Han, **Yi Gu**, Alex Ce Zhang, and Yu-Hwa Lo. "Imaging technologies for flow cytometry." **Lab on a Chip** 16, no. 24 (2016): 4639-4647.

Yuanyuan Han, Rui Tang, **Yi Gu**, Alex Ce Zhang, Wei Cai, Violet Castor, Sung Hwan Cho, William Alaynick, and Yu-Hwa Lo. "Cameraless High-throughput 3D Imaging Flow Cytometry." **arXiv preprint** arXiv:1902.01393 (2019).

Yi Gu, Alex Ce Zhang, Yuanyuan Han, Jie Li, Clark Chen, and Yu-Hwa Lo. "Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification." **Cytometry Part A** (2019)

Yi Gu, Xinyu Chen, Sung Hwan Cho, Chang-Hung Lee, Ivan Gagne, Alex Ce Zhang, Rui Tang, Yuanyuan Han, and Yu-Hwa Lo. "Image-Guided Cell Sorting Using Laser Point Scanning" **in preparation**

PATENTS

Yu-Hwa Lo, Yuanyuan Han, **Yi Gu**, Alex Ce Zhang. "IMAGE-BASED CELL SORTING SYSTEMS AND METHODS." WO/2017/214572

Yu-Hwa Lo, **Yi Gu**, "2D Image-Guided Cell Sorter" in application

CONFERENCES

Yi Gu, Alex Zhang, Yuanyuan Han, Sunghwan Cho, Yuhwa Lo, "Cell Sorter based on Fluorescence and Bright Field Images", **Congress of the International Society for Advancement of Cytometry 2017** (named finalist for The 2017 Exceptional Student Award)

Yi Gu, Alex Zhang, Yuanyuan Han, Sunghwan Cho, Yuhwa Lo, "Image Guided High-Throughput Cell Sorter", **Congress of the International Society for Advancement of Cytometry 2018** (named finalist for The 2018 Exceptional Student Award)

Yi Gu, Rui Tang, Alex Ce Zhang, Yuanyuan Han and Yu-Hwa Lo, "Image-guided microfluidic cell sorter with machine learning", **CLEO** (2019)， Control Number: 3080140

ABSTRACT OF THE DISSERTATION

Image-Guided Cell Classification and Sorting

by

Yi Gu

Doctor of Philosophy in Electrical Engineering (Photonics)

University of California San Diego, 2019

Professor Yuhwa Lo, Chair
Professor Zhaowei Liu, Co-Chair

The ability to classify and map numerous cell types as well as healthy and diseased cells can bring significant insight to biology and medicine. While single-cell sequencing becomes cornerstone for cell classification and mapping, isolation of interested cells for genomic analyses rely on fluorescence activated cell sorting (FACS), which can only isolate cells based on integrated intensities. The availability

of flow cytometers with the capability to classify and isolate cells guided by high-content cell images is enabling and transformative. It provides a new paradigm to allow researchers and clinicians to isolate cells using multiple user-defined characteristics encoded by both fluorescent signals and morphological and spatial features. In this thesis, we demonstrated the “Image-Guided Cell Classification and Sorting” technology. This technology possesses high throughput isolation capability of FACS and high information content of microscopy.

To achieve “Image-Guided Cell Classification and Sorting”, we combined the techniques of machine learning, photonics, real-time signal processing and microfluidics.

Chapter 1 Introduction

1.1 Overview

1.1.1 Motivation

There are a far greater number of cell types than people realized in the past, and classifying cells from healthy and diseased tissues in much finer detail than before can bring significant insight in biology and medicine. While sequencing of single cells becomes the technology cornerstone for cell classification, selection of these single cells for genomic analyses rely on fluorescence activated cell sorting (FACS) systems.[1,2] A small biological sample can contain millions of cells, hence analyzing even as many as 100,000 single cells represent only a very small percentage of cells in the sample. Thus intelligent selection of this small percentage of cells for downstream analysis is critical to efficient and accurate cell classification. However, today's cell selection techniques are purely based on fluorescent biomarkers and/or light scattering intensity, without resorting to high content image information that has the most distinctive power to support smart and logical selection of cells, especially those rare cells and cells without known or unique biomarkers.

Cell sorting based on high-content images has not been achieved partly due to the incompatibility between cell imaging and cell sorting systems and more importantly, due to the inability of real-time image acquisition, feature extraction, and construction of image-based gating as criteria for cell sorting. We demonstrate an

image-guided flow cytometer cell sorter. The availability of flow cytometers with the capability to classify and isolate cells guided by high-content cell images is enabling and transformative. It provides a new paradigm to allow researchers and clinicians to isolate cells using multiple user-defined characteristics encoded by both fluorescent signals and morphological and spatial features.

Examples of applications include isolation of cells based on organelle translocation, cell cycle, detection and counting of phagocytosed particles, and protein co-localization, to name a few.[3-6] Some specific applications include translocation of glucocorticoid receptor (GR) from cytosol to nucleus under dexamethasone treatment[7], glucocorticoid receptor and sequential p53 activation by drug mediated apoptosis[8], and translocation of protein kinase C (PKC) from cytosol to membrane in the context of oncogenesis[9]. β-arrestin-GFP is often used to measure the internalization (inactivation) of g-protein coupled receptors (gpcrs) as β-arrestin-GFP moves from cytosol to membrane. The ~800 Gpcrs include the opioid receptors (heroin, morphine, pain pills), the dopamine receptors (cocaine, methamphetamine, addiction/reward), and hundreds of others, many awaiting discovery or "adoption" of ligands. Other specific application examples include immunology studies of B-cell or T-cell responses to various drug treatments, asymmetric B-cell division in the germinal center reaction[10,11], the erythroblast enucleation process, signaling and cytoskeletal requirements in erythroblast enucleation[12,13], uptake and

internalization of exosomes by various cancer cells, response of infected cells to drugs, use of antibody-drug conjugates for tracking drugs in/outside sub-cellular compartments, and locating antigens, enzymes or other molecules.[14-18]

1.1.2 Potential Solutions

The image-guided cell sorting and classification technology possesses the high throughput of flow cytometer and high information content of microscopy, being able to isolate cells according to their imaging features at 1000X faster rate than laser microdissection and single-cell aspiration. Two methods are employed to capture weak fluorescence signals at high throughput. The first is the spatial-temporal transform, and the second is the laser point scanning. We also developed a methodology of user-interface (UI) to generate sorting criteria by supervised machine learning, as described next.

After hundreds of cells pass through the imaging flow cytometer, the software generates a distribution of cell parameters, as well as several categories of cell images based on the built-in image processing and statistical classification algorithms. Users then apply point-and-click selection of desired cell images for the basis of gating the cells for sorting from the sample. After collecting an additional number of cells based on user's instructions, the software displays both the conventional flow cytometer parameters (i.e. fluorescence intensity) and a new set of image/morphology related

parameters (e.g. nucleus size, cell area, circularity, fluorescence patterns, etc.), as well as the representative cell images of the cells. This iterative feedback process gives users the chance to confirm their original choice criteria and to modify the "gating". Based on the displayed image and conventional data feedback, users may adjust the gating criteria. These criteria can be "ratio of fluorescence area over the total cell area", "variations of fluoresce intensity profile over the cell", "size of nucleus", or numerous other choices utilizing the spatial features of the cells. As a result, the image-guided cell selection process becomes a user-interface (UI) and user-experience (UX) interactive process with machine learning occurring in the background to present users with representative images of cell classes that most closely match the user needs and even suggest features possibly overlooked by users. As a result, users are given unprecedented intuitive visual assistance and insight to enhance their studies.

1.2 Scope of Thesis

In this thesis, chapter 2 is Computational Cell Analysis for Label-free Detection of Cell Properties in a Microfluidic Laminar Flow Using Spatial-Temporal Transformation. Chapter 3 is Image-Guided Cell Sorting Using Spatial-Temporal Transform. Chapter 4 is Image-Guided Cell Sorting Using Laser Point Scanning.

Chapter 2 Computational Cell Analysis for Label-free Detection of Cell Properties in a Microfluidic Laminar Flow Using Spatial-Temporal Transformation

2.1 Computational cell analysis technique

2.1.1 Measurement of cell position within a microfluidic channel

In a microfluidic channel, cells of different physical properties (size, shape, stiffness, morphology, etc.) experience different magnitudes of lift and drag force, thus yielding different equilibrium positions in the laminar flow[19-21].

To determine the equilibrium position of a particular cell in the microfluidic channel, a spatial coding method was used to obtain the horizontal position and the velocity of the cell. The design and configuration of the system is illustrated in figure 2.1. The spatial mask has two oppositely oriented trapezoidal slits with the base lengths being $100um$ and $50um$. An LED source was used to illuminate from the bottom of the microfluidic channel. The transmitted signal was detected by a variable gain photoreceiver made of a Si photodiode and a transimpedance amplifier (Thorlab). All light blocking areas on the spatial mask was coated with a layer of Ti/Au on a glass slide. When cells flew through the spatial mask area, their forward scattering signal gave rise to a characteristic waveform encoded by the mask. The microfluidic channel

is 5 cm long and has its inlet and outlet at the ends. The rectangular cross section of the channel is $100um$ wide and 50 um high. The mask is located at 4.5 cm from the inlet. In the following discussion, we will represent the channel width direction as x-axis and channel height direction as y-axis. The intensity modulated FS signal by the trapezoidal slits displays 2 peaks, as shown in figure 2.2(b). The ratio between the width of the first peak and the second peak provides information of the cell position in the X-axis; and the absolute value of the signal width gives information about cell velocity. Knowing the position along the x-axis and the cell velocity, the cell position along the y-axis can be obtained using the property of laminar flow that gives rise to a parabolic velocity profile represented by the following relation[22-25]:

$$y = h \times \sqrt{1 - \frac{V(x,y)}{V_{max} - (\frac{x - L/2}{L/2})^2}}$$

In the above equation, y is the cell position in the channel height direction (y-axis), h is the half channel height, $25\mu m$ in this work. x is the horizontal position. $V(x,y)$ is the velocity at a specific position. L is the channel length in horizontal direction, in our case 100 μm. V_{max} is the velocity at $x = \frac{L}{2}, y = 0$, in our case, $x = 50\ \mu m, y = 0\ \mu m$.

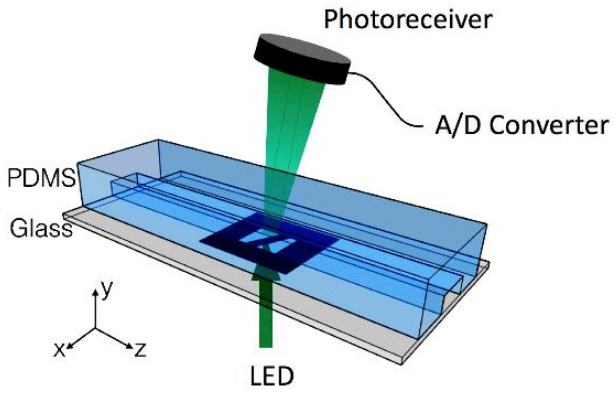


Figure 2.1 Experiment setup

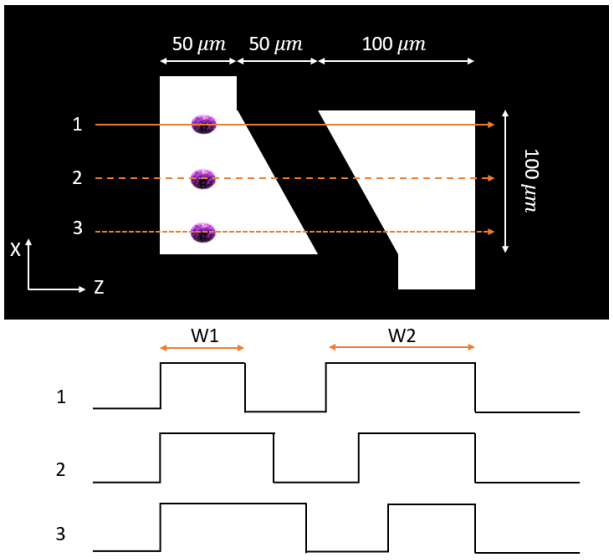


Figure 2.2 (a) Spatial mask design with two oppositely oriented trapezoidal slits. W1 and W2 represent the widths of the slits experienced by a cell traversing the mask from different positions. Also shown are the anticipated waveforms for cells crossing the mask area via different paths.

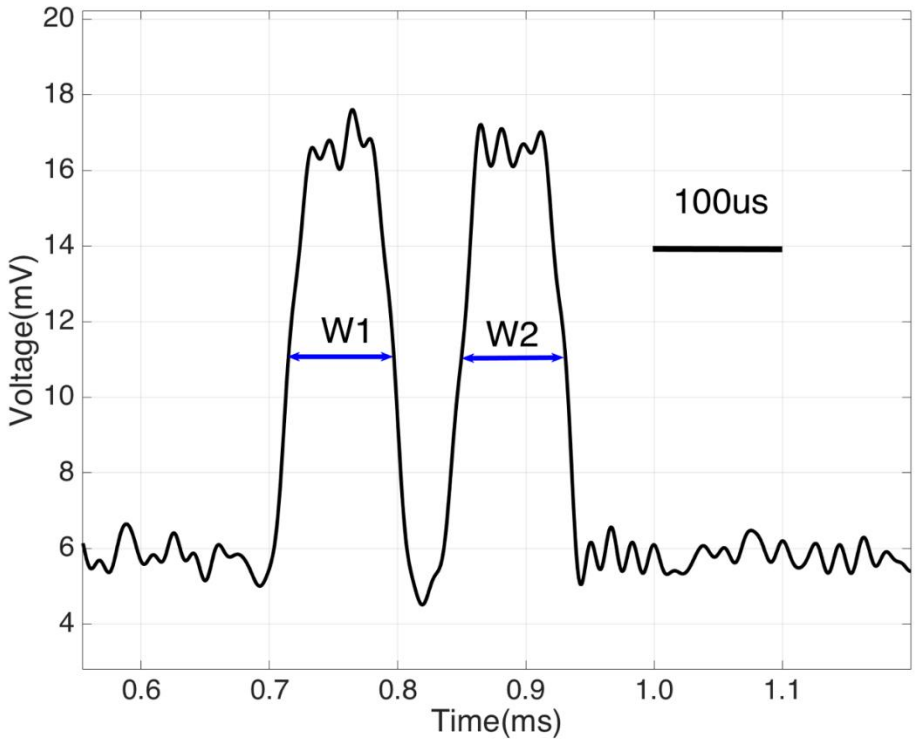


Figure 2.2 (b) Intensity modulated forward scattering signal by the trapezoidal slits

2.1.2 Computational cell analysis methods

For a given type of cells in a channel, their equilibrium positions can be represented by a characteristic distribution $R(x, y)$. The characteristic distribution can be obtained in a diluted sample to avoid any effects caused by cell-cell interactions in the flow. Figure 2.3 shows the characteristic distribution of Live MDA-MB-231 cells. Without cell-cell interactions in the fluid, the spatial distribution of a sample containing multiple cell groups is the linear combination of the characteristic functions of each cell type.

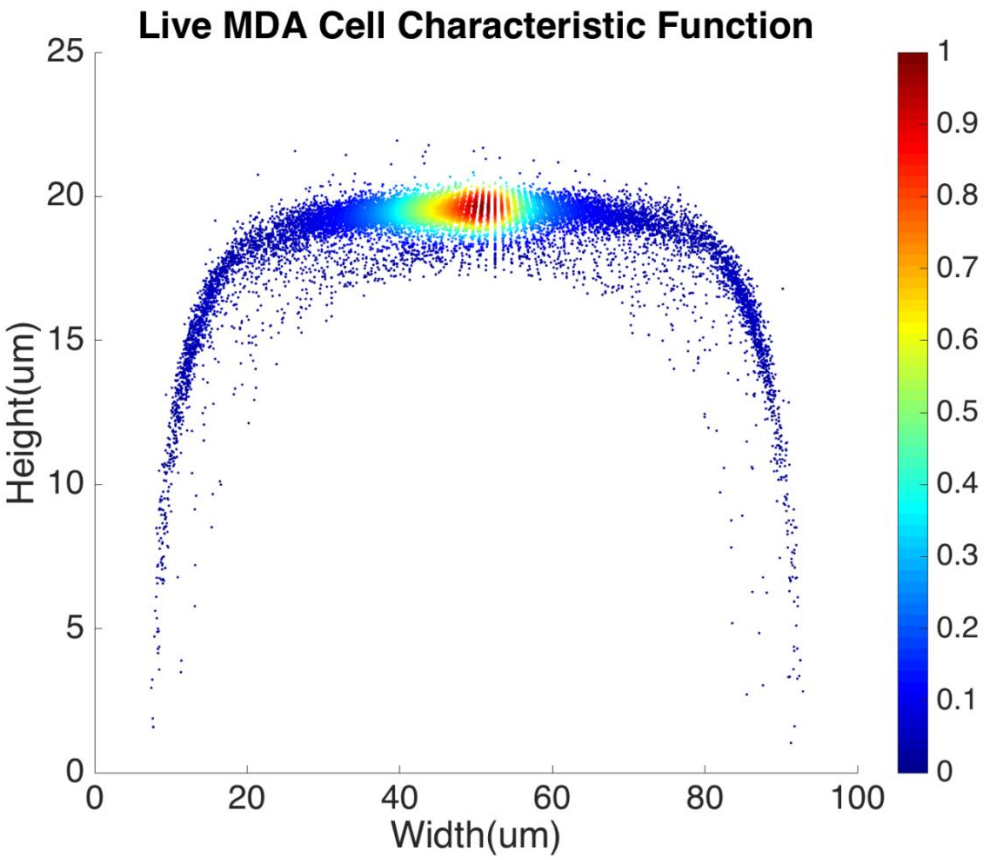


Figure 2.3 Spatial characteristic function of live MDA cell

If $R_A(x,y)$ and $R_B(x,y)$ are the characteristic functions of cell type A and cell type B, the spatial distribution, $S(x,y)$, for a mixture of cell A and cell B can be described by Eq. (1) where the coefficient, C, is the fraction of cell A in the sample.

$$S(x,y) = C \times R_A(x,y) + (1-C) \times R_B(x,y) \qquad (1)$$

Since both $R_A(x,y)$ and $R_B(x,y)$ are normalized over the entire cross section of the fluidic channel,

$$\iint_{x,y} R_A(x,y)dxdy = 1 \text{ and } \iint_{x,y} R_B(x,y)dxdy = 1$$

The distribution function $S(x,y)$ for the sample is also normalized as in (2).

$$\iint_{x,y} S(x,y)dxdy = C \times \iint_{x,y} R_A(x,y)dxdy + (1-C) \times \iint_{x,y} R_B(x,y)dxdy = 1 \quad (2)$$

If we already know $R_A(x,y), R_B(x,y)$ from the training data and measure $S(x,y)$ from the sample, we can find from Eq. 1 the only unknown, C, being the population of cell A, which is the information of interest.

In practice, the values of $R_A(x,y), R_B(x,y), and\ S(x,y)$ at each specific position (x,y) are random variables. We divide the whole area of the channel cross section into meshes so the sets of random variables $R_A(x,y), R_B(x,y), and\ S(x,y)$ follow the relations:

$$\sum_{x,y} R_A(x,y) = 1\ and\ \sum_{x,y} R_B(x,y) = 1$$

$$\sum_{x,y} S(x,y) = C \times \sum_{x,y} R_A(x,y) + (1-C) \times \sum_{x,y} R_B(x,y) = 1 \quad (3)$$

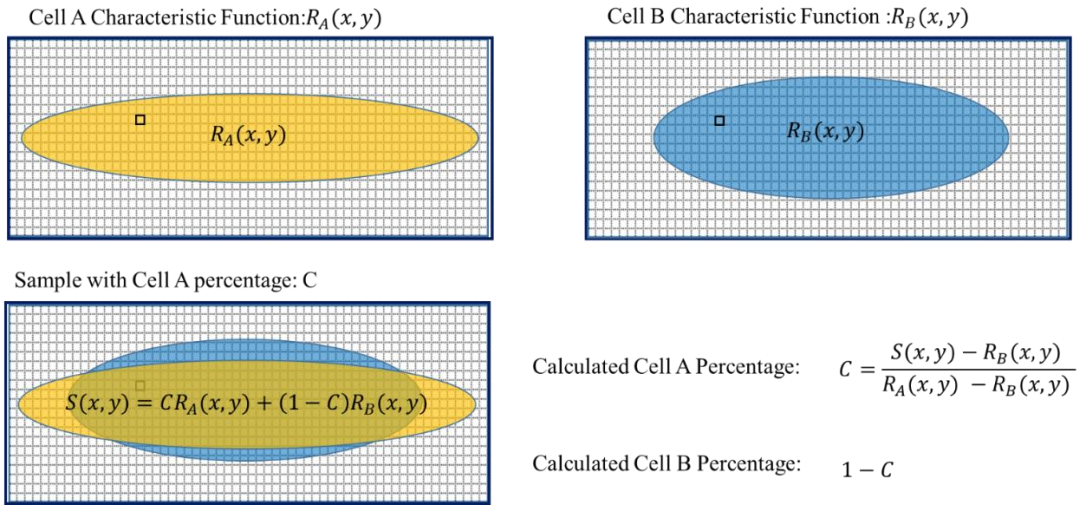


Figure 2.4. Illustration of the steps to calculate the cell ratio in a sample of cell mixture. The yellow and blue patterns represent the characteristic distributions for cell A and cell B.

Due to the statistical nature of the problem, the resulting value of C, calculated at each position (x, y), is also a random variable. Thus we can plot the distribution of C, the percentage of cell A in the sample, in a histogram. The mean value of the histogram produces the most likely percentage for cell A in the sample and the spread of the histogram provides a good indication of the quality of the measurement. Figure 2.5 shows an example of the histogram for C.

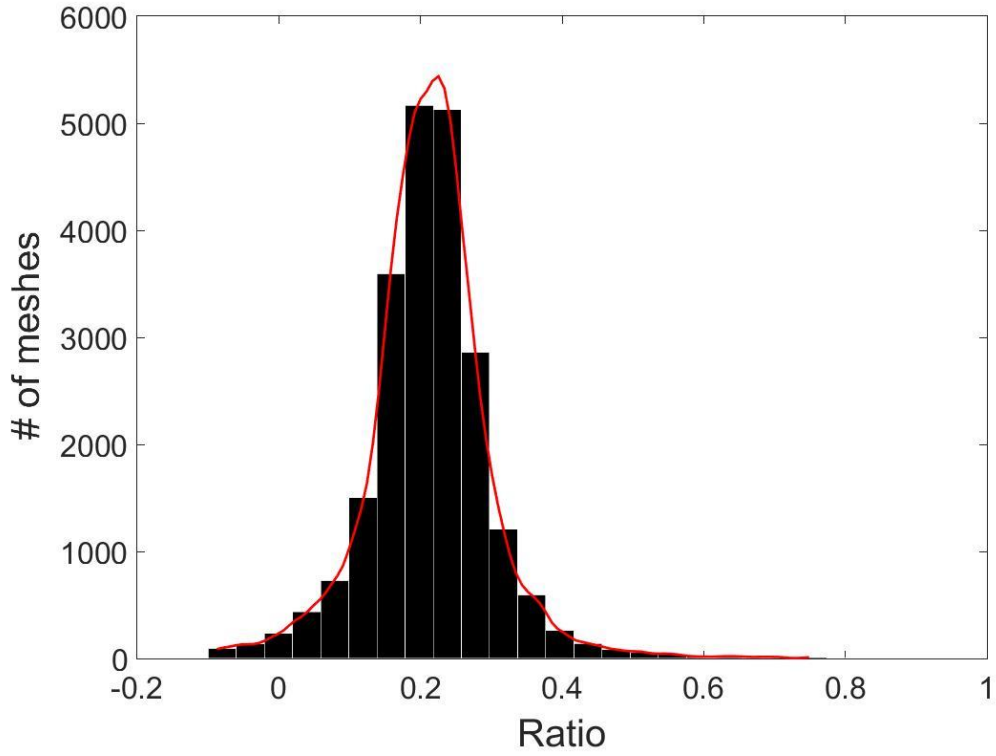


Figure 2.5. Histogram of the cell ratio, C.

Therefore, for any cell type possessing even a slightly different physical property than the rest of the cells in the sample, manifested by its largely overlapped spatial distribution with the rest of the samples, our method is still able to find the relative population of such specific cell type in a cell mixture. This is a unique capability of the proposed method.

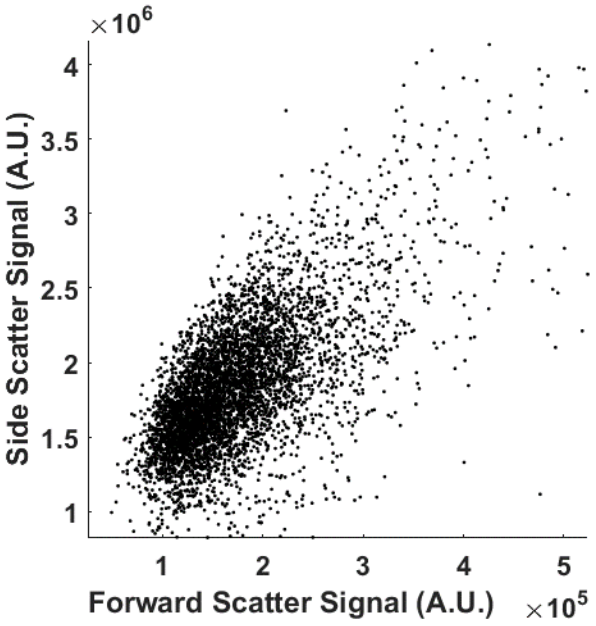
2.2 Demonstrate the technology using cell samples

2.2.1 Detection of the population of live and fixed cells

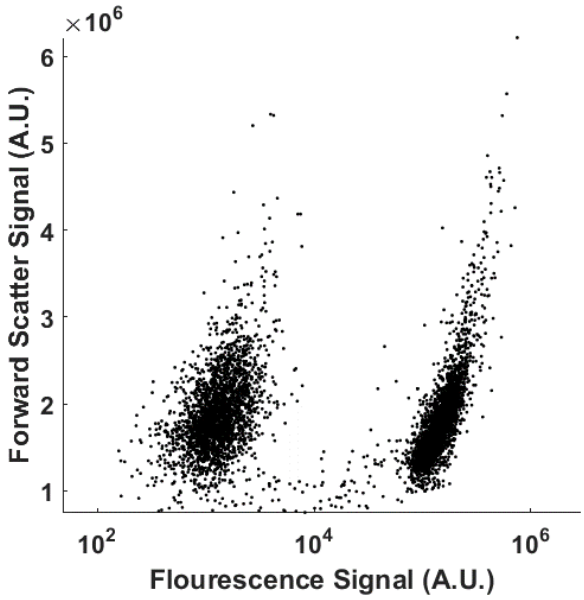
Without labelling, today's flow cytometers cannot find the percentage of live and dead cells in a culture because the scattering signals (forward, side, and back scattering) between live and dead cells overlap significantly in the distribution plot. We demonstrate that in spite of the high similarity in their physical properties of live and dead cells, the computation cell analysis technique enables us to give an unequivocal answer to the subpopulation of live cells from dead cells.

We used live and fixed MDA cells to demonstrate the concept. MDA-MB-231 is a cell line for human breast cancer cells. In the experiment, we fixed one group of cultured MDA cells and labeled them fluorescently (Propidium lodide), and then mixed these fixed cells with live cells in different ratios. Each sample with a specific ratio of live and fixed cells was divided into two parts, one running through our device and the other running through a commercial flow cytometer. Figure 2.6(a) shows the distribution plot for forward and side scattering signals from a flow cytometer. It was impossible to determine the ratio between live and fixed MDA cells from the scattering signals by any gating or machine learning algorithms due to the significant overlap of the signals from live and dead cells. The only reliable method for a flow cytometer to detect the relative population of live cells from dead cells is by fluorescent labeling as shown in Figure 2.6(b). Next we demonstrate how the computation cell analysis method can solve this problem.

To produce the characteristic function for live and dead MDA cells, we ran through our system with two samples with 100% live MDA and 100% fixed MDA. Then we ran multiple test samples with different ratios of live and fixed MDA cells. For each test sample, we ran the experiment 10 times, taking 3 minutes for each run, to test the reproducibility and reliability of the results. Figure 2.7 shows comparisons between our method and the flow cytometer results by detecting the fluorescent signals of fluorescently labelled fixed cells. The excellent agreement and similar repeatability (10 repeats) of the results from both methods demonstrate the accuracy and reliability of our label-free computation cell analysis method.



(A)



(B)

Figure 2.6. (a) Forward and side scattering of live and fixed MDA cells; (b) fluorescent signal of live and fixed MDA cells. The cluster on the left was auto fluorescence from live cells and the cluster on the right was Propidium Iodide labelled fluorescent signal from fixed cells.

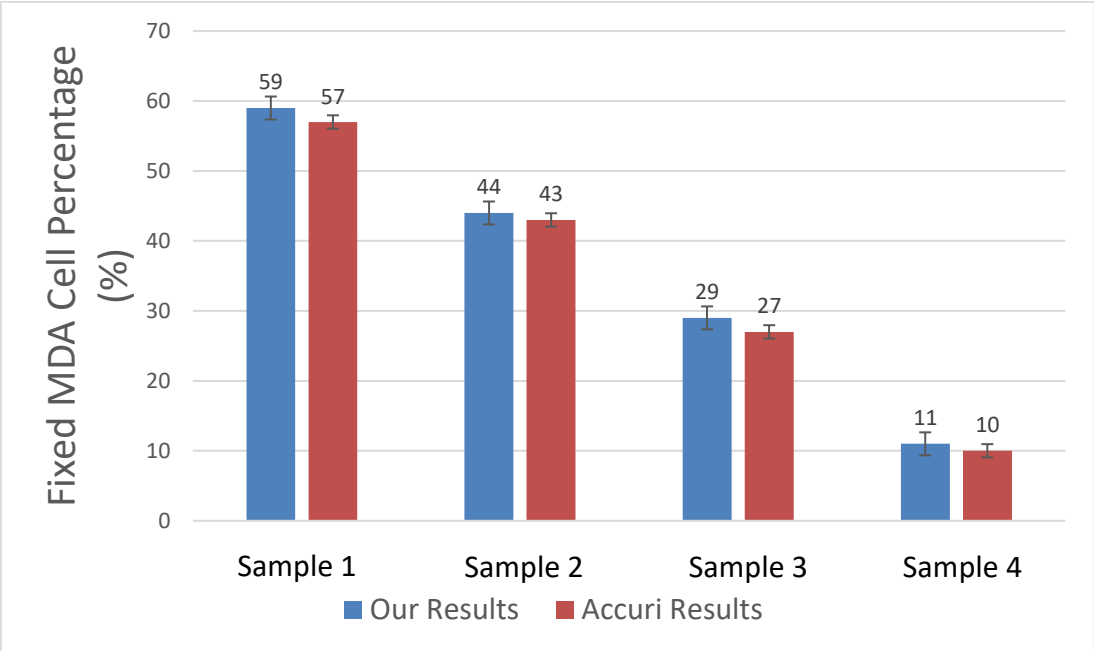


Figure 2.7. Measured mean value of live cell percentage in 4 samples. The error bars show the variations for 10 repeats for each sample using our method and a commercial flow cytometer (Accuri C6).

2.2.2 Neutrophil counter for point-of-care applications

There are three major types of white blood cells, neutrophil, lymphocyte, and monocyte. Neutrophil count is an indicator of patient's immunity to infections and is particularly important for cancer patients undergoing chemotherapy since the treatment can lower neutrophil count. Neutropenia develops when the neutrophil count falls below certain levels, substantially increasing the risk of infections.[26,27] Therefore, the neutrophil count of chemotherapy patients has to be closely monitored, presenting the need for point-of-care neutrophil counter without fluorescent labelling. In the following we demonstrate how the computation cell analysis technique can count neutrophil in a point-of-care setting.

We performed the experiment using purchased blood from San Diego blood bank. After red blood cell lysing, the blood was diluted with 1X PBS solution. Since we were interested in neutrophil count, we treated neutrophil as cell A and all non-neutrophil WBCs as cell B. Then we represent the characteristic function for neutrophil as $R_A(x, y)$ and all non-neutrophil WBCs as $R_B(x, y)$. To obtain $R_A(x, y)$ and $R_B(x, y)$, we did not use blood samples with 100% pure neutrophils since complete removal of neutrophil from the samples can be practically difficult. Instead, we chose two blood samples with different neutrophil to non-neutrophil ratios. We used superparamagnetic beads (Dynabeads from ThermoFisher) to remove some neutrophils from the blood to produce samples with lower than normal amounts of neutrophil, which also simulated neutropenia patients. Using the aforementioned protocol, the superparamagnetic beads--Dynabeads CD15--that were covalently coupled with an anti-human CD15 antibody were used to deplete human CD15+ myeloid cells, predominantly neutrophils, directly from whole blood. Different concentrations of Dynabeads CD15 were used to create blood samples having various percentages of neutrophil.

Using the above mentioned algorithm (see Fig. 2.4), we obtained the characteristic functions of Neutrophil and non-neutrophil as shown in Figure 2.8. From these characteristic functions, one could apply Eq. (4) to find the neutrophil ratio from samples of an unknown neutrophil population. Again, we have divided each test sample into two parts, one going through our device and another going through a

commercial flow cytometer (Accuri C6). For each sample the test was repeated 10 times. Figure 2.9 shows the comparison of results from both methods. The experiment was designed to cover the whole range of neutrophil ratio to simulate healthy samples and samples with different degrees of neutropenia.

Table 2.1 summarizes results from another set of experiment out of 8 purchased blood samples from San Diego blood bank. The excellent agreement between the proposed method and the commercial flow cytometer shows that the computational cell analysis device, being a highly flexible and versatile technique, can operate as a cost effective, point-of-care neutrophil counter.

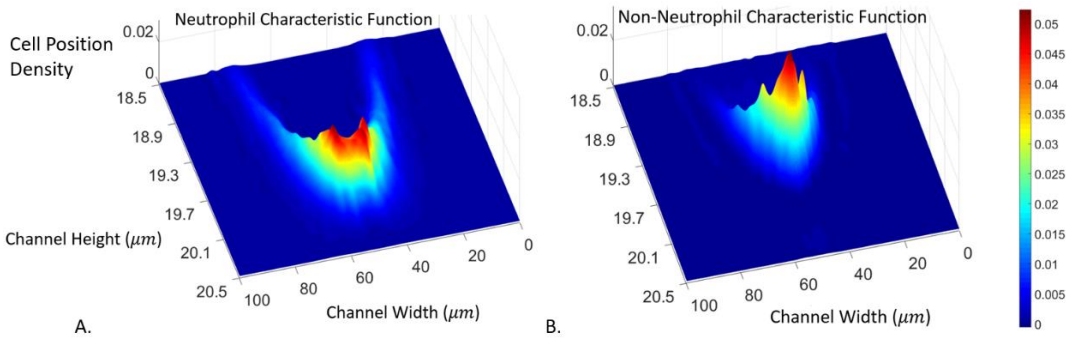


Figure 2.8. Neutrophil and Non-neutrophil characteristic functions.

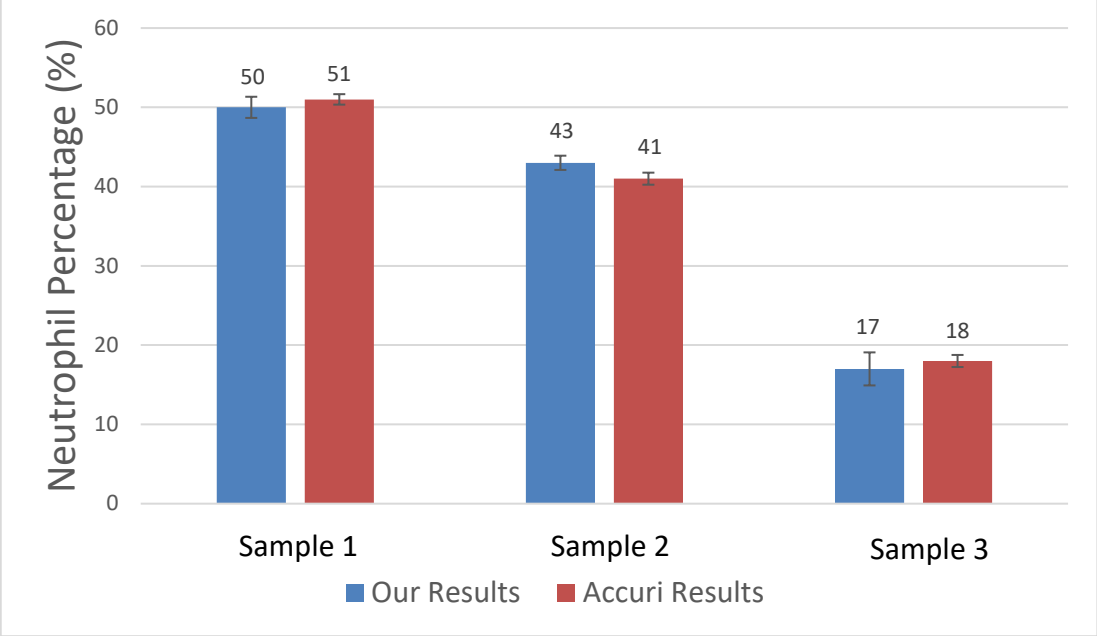


Figure 2.9. Measured mean value of neutrophil percentage over WBCs in 3 samples. The error bars show the variations for 10 repeats for each sample using our method and a commercial flow cytometer (Accuri C6).

Table 2.1 Measured neutrophil percentage over WBCs from 8 blood samples using our method and a commercial flow cytometer. Samples 9 and 10 were used to obtain the characteristic functions for neutrophil and non-neutrophil

Test			
Sample	Flow cytometer (%)	Our Method (%)	Deviation (%)
1	50	52	4.0
2	50	51	2.0
3	56	58	3.6
4	59	59	0.0
5	64	62	3.1
6	50	50	0.0
7	66	64	3.0
8	52	52	0.0

Training	
9	60
10	49

Chapter 2 is based on and mostly a reprint of: Alex Ce Zhang, **Yi Gu(co-first author)**, Yuanyuan Han, Zhe Mei, Yu-Jui Chiu, Lina Geng, Sung Hwan Cho, and Yu-Hwa Lo. "Computational cell analysis for label-free detection of cell properties in a microfluidic laminar flow." **Analyst** 141, no. 13 (2016): 4142-4150. The dissertation author was the primary author of the work.

Chapter 3 Image-Guided Cell Sorting Using Spatial-Temporal Transform

3.1 System Design of Image-Guided Cell Sorter

As shown in Figure 3.1 (a), the image-guided cell sorting and classification system consists of: (1) an imaging optical system with a spatially coded optical filter to perform spatial-temporal transformation, (2) a real-time image processing and feature extraction module, (3) an off-line post processing module for construction of cell images for human vision and generation of cell classification criteria, and (4) a microfluidic chip integrated with an on-chip piezoelectric (PZT) cell sorting actuator.

The optical filter encodes the fluorescent or light scattering signal of a cell into a temporal photocurrent waveform in the output of a photomultiplier tube (PMT) detector. Through a mathematical transformation described in[28,29], the 1D time domain signal is transformed into a 2D cell image. Due to the simplicity of the transformation algorithm, real-time signal processing can be implemented to extract image features of each cell to allow cell sorting based on these image features.

Machine learning is needed to generate and adjust image features to guide cell sorting.[30-32] To start, training samples are flowed through the system to produce a set of training data. Off-line processing is employed to construct high resolution, cell images to interface with users whose inputs will aid the selection and adaptation of cell classification criteria for real-time sorting, a method of supervised machine learning. During the real-time cell sorting experiments, real-time processing module

reconstructs cell images, extracts image features and makes sorting decisions based on the off-line trained sorting criteria. When a decision is made to sort a cell, a voltage pulse is applied to the on-chip PZT actuator, which instantaneously bends the bimorph PZT disk to deflect the cell away from the central flow into the sorting channel. For a proof-of-concept prototype, we have used a field-programmable-gate-array (FPGA) platform to implement real-time image processing which produces a latency of a few milliseconds. Higher performance GPU processors can reduce the processing time by 100 folds or more to microseconds.

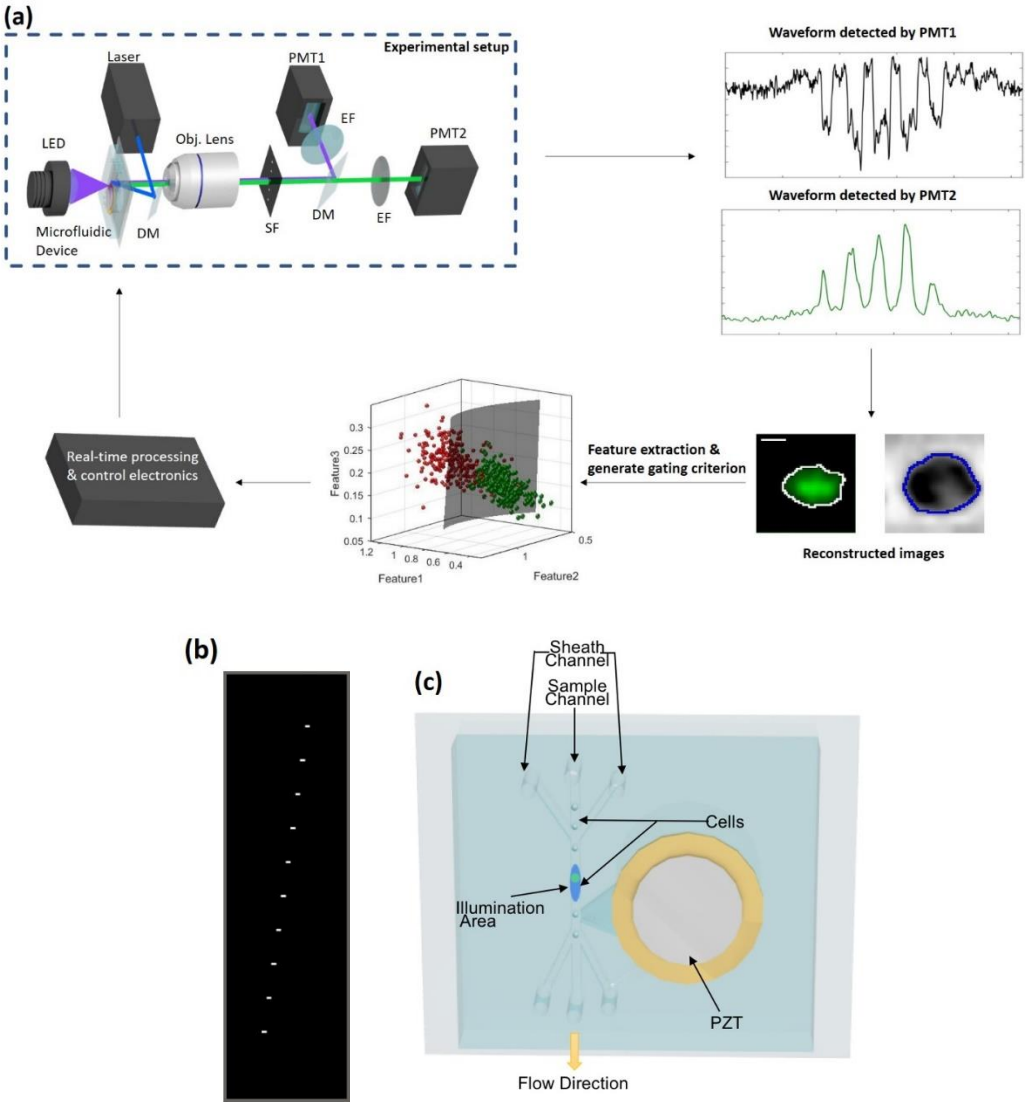


Figure 3.1 The Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification system. (a) Schematic diagram of the image-guided cell sorting system. (Scale bar is 5μm). Bright field and fluorescence cell images are at first encoded into time domain waveforms and detected by PMTs. Cell images are then reconstructed from time domain waveforms. Next, image features are extracted and image-based gating criteria are generated. Finally, cell images are processed and sorting decision is made in real-time based on the image-guided gating criteria with supervised machine learning. DM, dichroic mirror; SF, spatial filter; EF, emission filter; PMT, photomultiplier tube. (b) Design of optical spatial filter having ten 100 μm by 50 μm slits positioned apart. (c) Microfluidic device with an on-chip piezoelectric PZT actuator to deflect selected cells in the microfluidic channel for image-guided cell sorting.

3.2 Optical imaging setup

In the optical imaging system, suspended single cells flow in a microfluidic channel made of soft-molded polydimethylsiloxane (PDMS) bonded to a glass substrate. Sheath flow is used to hydro-dynamically focus the travelling cells to the center of the microfluidic channel. At the interrogation zone, each flowing cell is illuminated simultaneously by a 500mW 455nm LED (Thorlabs) and a 100mW 488nm laser (iBeam-SMART, Toptica) to generate bright field and fluorescent images. The output beam of 488nm laser is collimated, focused and expanded to illuminate a 100μm (x-direction) by 250μm (y-direction) area. The LED light is collimated and focused at the laser illumination area. Both the fluorescent emissions and the transmitted bright field signal are detected by PMTs (H9307-02, Hamamatsu). To accommodate the geometry of the microfluidic device, the laser beam is introduced to the optical interrogation area by a 52-degree miniature dichroic mirror positioned in front of a 50X objective lens (NA=0.55, working distance=13 mm, Mituyoyo). The LED is placed at the opposite side of the channel and the light is focused to the same position as the laser beam. The spatially coded optical filter is inserted at the image plane in the detection path. The pattern of the filter is shown in Figure 3.1 (b). With the spatial filter, fluorescence/scattering from different parts of the cell will pass different slits at different times. As a result, the waveform of the fluorescent/scattering signal from the PMT consists of a sequence of patterns separated in time domain, and each section of the signal in the time domain corresponds to the fluorescent/scattering signal generated by each particular segment of the cell. After the light intensity profile

over each slit is received, the cell image of the entire cell can be constructed by splicing all the profile together. The image resolution in x- (transverse to the flow) direction is primarily determined by the number of slits on the spatial filter, and the resolution in y- (flow) direction is mainly determined by the sampling rate and cell flow speed. In our system, the raw image resolution is 2 μm in x-direction and 0.4 μm in y-direction.

Dichroic mirrors are used to route the desired emission bands to their respective PMTs.

3.3 Real-Time image processing and cell classification

The real-time image processing module is implemented in a field-programmable-gate-array (FPGA) platform (National Instrument cRIO-9039). The processing module performs the functions of cell detection and image processing, to be discussed next.

Both the bright field and fluorescent images of a cell are reconstructed from respective PMT readouts. Each image covers a field-of-view of 20X20μm^2 with 50X50 pixels, matched to the finest spatial resolution achievable by the system. All relevant image features are calculated and compared against the cell selection (gating) criteria to make sorting decision.

3.3.1 Cell detection algorithm

The cell detection function determines whether a cell exists within a certain time interval, and subsequently instructs the system whether to store and transmit the signal over this time interval. The system records the outputs from PMTs in the First-in First-out (FIFO) data structure over a chosen length of time. Each time a new set of

PMT readout enters the FIFO, cell detection algorithm is activated to determine whether there is a cell within the optical system's field of view. As soon as the system detects the presence of cell within the data set, the system processes the data immediately to construct images and extract image features. Otherwise, the system continues to examine the next set of PMT readout.

Since fluorescent and bright-field images are generated simultaneously, only one fluorescent signal is used in cell detection algorithm. To shorten the processing time, we calculate the *Brightness* of signal by integrating the fluorescent intensity stored in the FIFO. As shown in Figure. 3.2, if the *Brightness* is greater than the first threshold defined as *Threshold1*, the system assumes a cell is entering the field of view. Next, the time derivative of *Brightness* is compared to another preset threshold, *Threshold2*. If the time derivative of *Brightness* is smaller than *Threshold2*, the algorithm determines that the cell is within the field of view, and the function of real-time image processing is activated. After image processing is completed, *Brightness* calculated at later time is compared to *Threshold1* again. If the *Brightness* falls below *Threshold1*, the algorithm determines the cell has left the field of view.

Brightness is calculated from the PMT readout stored in the FIFO data structure. The FIFO for fluorescent signal is referred to as $FIFO_{PMT}$. Each time when a new PMT readout enters $FIFO_{PMT}$, *Brightness* value is updated based on the incoming PMT readout and the value of the dequeued element in $FIFO_{PMT}$. *Brightness* is calculated by equation 1.

$$Brightness[t_n] = Brightness[t_{n-1}] + Voltage_{PMT} - FIFO_{PMT}[dequeue] \quad (1)$$

where $Brightness[t_n]$ is the updated *Brightness*, $Brightness[t_{n-1}]$ is *Brightness* of last time step, $Voltage_{PMT}$ is the incoming PMT readout, which is the enqueued element of $FIFO_{PMT}$, and $FIFO_{PMT}[dequeue]$ is the dequeued element in $FIFO_{PMT}$.

Calculated *Brightness* enters another FIFO named $FIFO_{Brightness}$. The length of $FIFO_{Brightness}$ is 100. When a cell is entering the field of view, the time derivative of *Brightness* is evaluated to check if the cell is within the field of view. To evaluate if the time derivative of *Brightness* is small enough, the maximum element in $FIFO_{Brightness}$ ($max_{Brightness}$) is compared to the minimum element in $FIFO_{Brightness}$ ($min_{Brightness}$). If the difference is smaller than a predefined value (e.g. 5%), the system considers the cell is within the field of view. Both $max_{Brightness}$ and $min_{Brightness}$ are updated each time a new element enters $FIFO_{Brightness}$. The algorithm to calculate $max_{Brightness}$ and $min_{Brightness}$ is shown as Figure. 3.3.

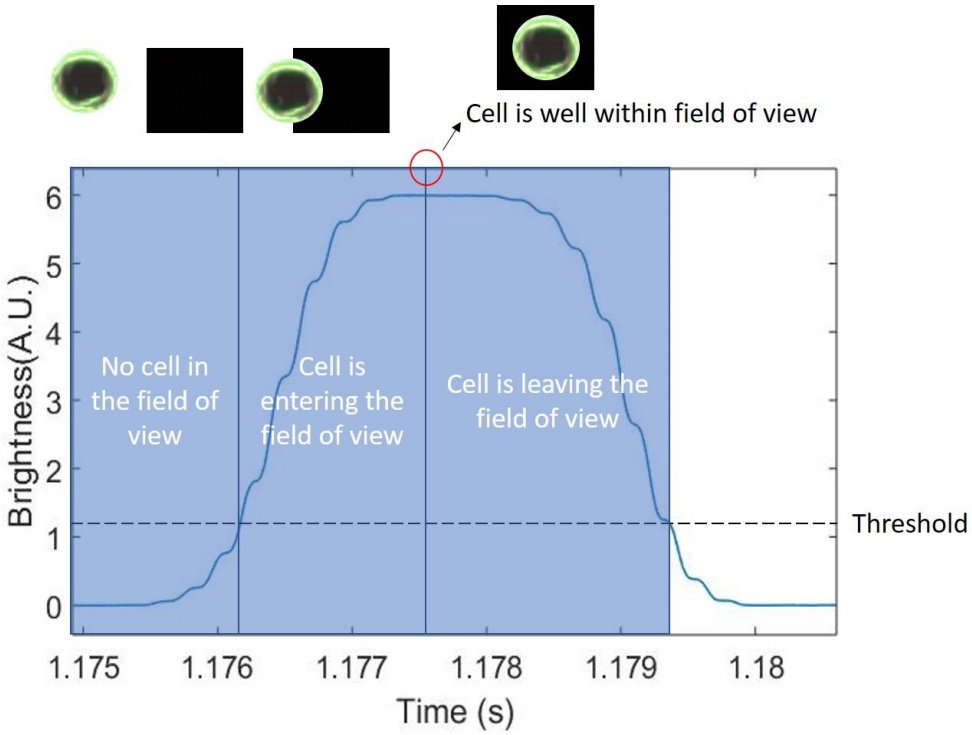


Figure 3.2 Illustration of cell detection algorithm.

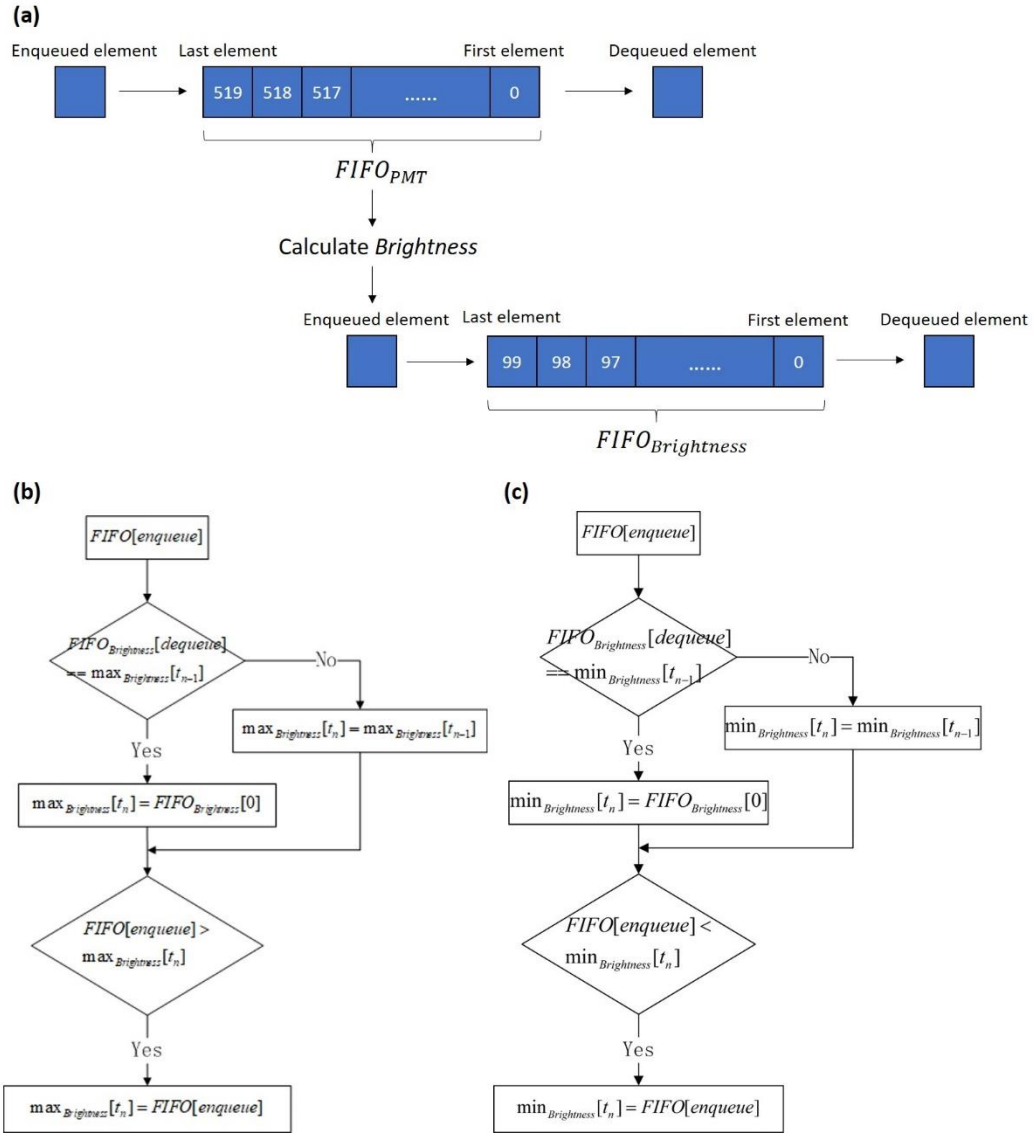


Figure 3.3 (a) Schematic diagram of $FIFO_{PMT}$ and $FIFO_{Brightness}$. (b) Flow chart for $\max_{Brightness}$ calculation. (c) Flow chart for $\min_{Brightness}$ calculation.

3.3.2 Image processing algorithm

For all applications, we essentially follow the same flow for real time image processing, involving denoising, image resizing, contour definition, area calculation, feature enumeration, etc. Some of the processes can run in parallel to simultaneously

extract multiple image derived features pertinent to image-guided sorting. In the following, we depict the specific processes applicable to each specific experiment.

Real time image processing algorithm for protein translocation experiment.

The image processing algorithm is illustrated in Figure 3.4. The image processing algorithm includes the following steps: (1) Denoise PMT signals with a 10^{th}-order Hamming low-pass filter. (2) Reconstruct both bright-field and fluorescent images from PMT signals. Since both bright-field and fluorescent signals are generated by the same slit although from different light sources, they are well synchronized. Thus the image reconstruction algorithm only needs to be launched once for both bright-field and fluorescent images. (3) Resize the images from 10x50 pixels (due to asymmetric resolution in raw images) to 50x50 pixels. (4) Detect contours of cell images by first converting the grayscale images to binary images, then eliminating spurious noise with open filter, and finally applying the contour detection algorithm to the binary images. (5) Extract all image-derived parameters.

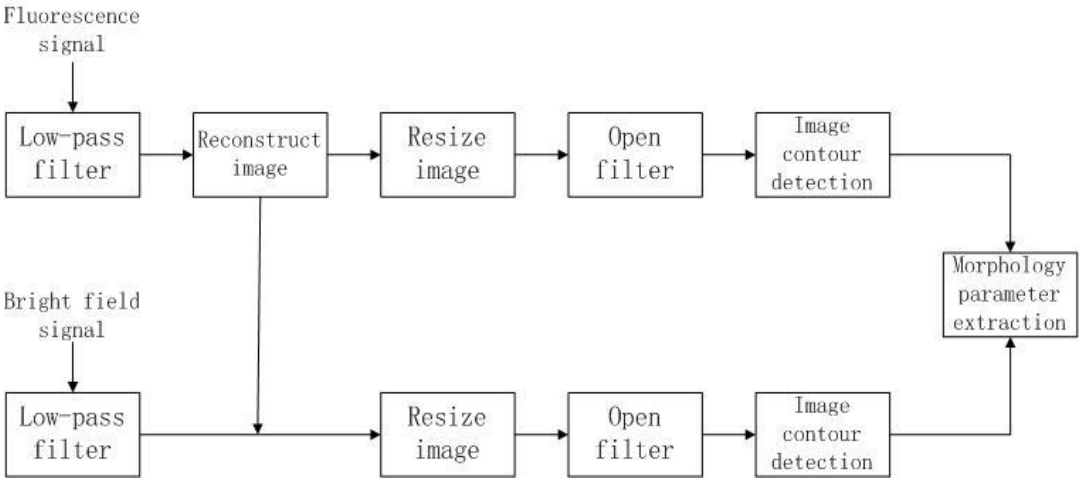


Figure 3.4 (a) Flow chart of real-time image processing algorithm for protein translocation experiment.

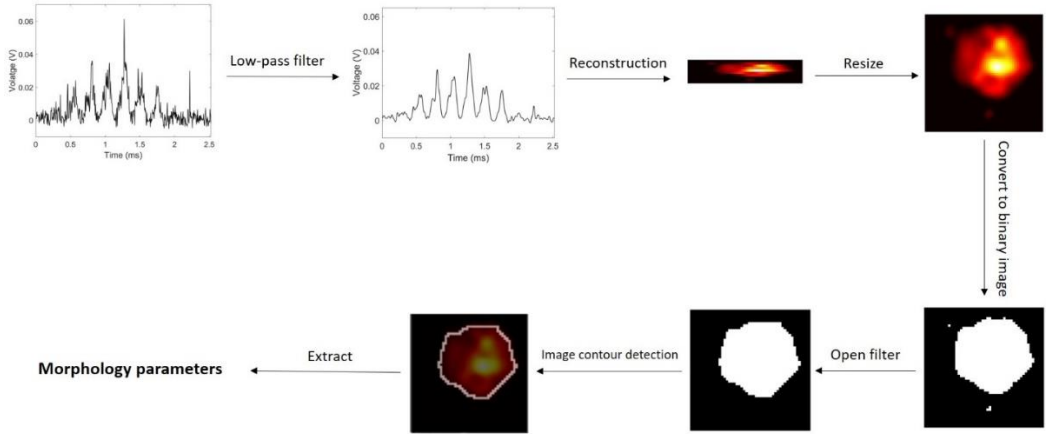


Figure 3.4 (b) An example of image processing algorithm for protein translocation experiment.

Examples of image-derived parameters that can be extracted from each cell in real time are shown in Table 3.1. As discussed later in this paper, these image-derived parameters are evaluated with Receiver Operating Characteristics (ROC). The 3 parameters receiving the highest ROC score are used as default parameters for real-time image-guided sorting. The latency of real-time processing algorithm is

5.8ms/cell with the current FPGA system, and the latency can be reduced by over 100 times with high performance GPU(e.g. NVIDIA QUADRO P6000).

Table 3.1. Extracted image-derived parameters.

Fluorescent image	**Bright-field image**	**Fluorescent image + Bright-field image**
Area	Area	Fluorescent area/bright-field area (Area ratio)
Perimeter	Perimeter	Fluorescent perimeter/bright-field perimeter (Perimeter ratio)
Shape factor (Area/Perimeter)	Shape factor (Area/Perimeter)	
Diameter (in x direction)	Diameter (in x direction)	
Diameter (in y direction)	Diameter (in y direction)	

Real time image reconstruction and speed detection. An optical spatial filter consisting of 10 slits is placed at the image plane of the signal. With a 50X (or 20X if desired for extended focal depth) objective lens, the image projected onto the optical

spatial filter is 50X (20X) times greater than the object in the microfluidic channel. For each cell travelling in the microfluidic channel, its PMT readout produces10 peaks, each corresponding to the cell's fluorescent or bright-field transmitted signal passing one of 10 slits on the spatial filter. At a cell travel speed of 8 cm/s and at 200 kSamples/s, each peak consists of 50 sampling points. To reconstruct cell images, two factors need to be considered. First, since cells do not travel at a perfectly uniform speed, the actual number of sampling points for each of 10 peaks may vary slightly. Second, since both cell travelling speed and cell position within the 20 μm by 20 μm image area can vary, the starting time point of the PMT readout also varies. In the image reconstruction algorithm, we refer "m" to be the starting point of PMT readout, "n" to be the number of sampled points in each peak. Based on cell speed variations, n ranges from 46 to 51. This leads to a range of m from 0 to 519-10n. The algorithm sweeps m and n to assure the best combination of (m,n) to reconstruct the cell image. Summation of intensities at the starting point of each peak is calculated for every (m,n) combination. The combination yielding the smallest sum is the right (m,n) which we use to reconstruct the image. Travelling speed of the cell can be detected based on the calculated value "n". The algorithm is shown in equation 2 and Figure. 3.5. In Figure. 3.5, The black "*" are starting points for each peak found by the reconstruction algorithm.

$$(m,n)=\left\{(m,n)\left|\min\sum_{i=0}^{10}\sum_{m=0}^{519-10n}FIFO_{PMT1}[m+i\times n]\right.\right\} \quad (2)$$

After image reconstruction, grayscale cell images are resized to 50x50 pixel images by linear interpolation. Then resized grayscale images are converted to binary images based on the preset intensity threshold. The conversion is described as follows:

$$binary\_image(i,j)=\begin{cases}1, & if \quad grayscale\_image(i,j) > threshold \\ 0, & if \quad grayscale\_image(i,j) < threshold\end{cases} \quad (3)$$

where (i, j) refers to the pixel located at row i and column j.

In the step of open filter, we use a 3x3 pixel neighborhood in our image processing algorithm.

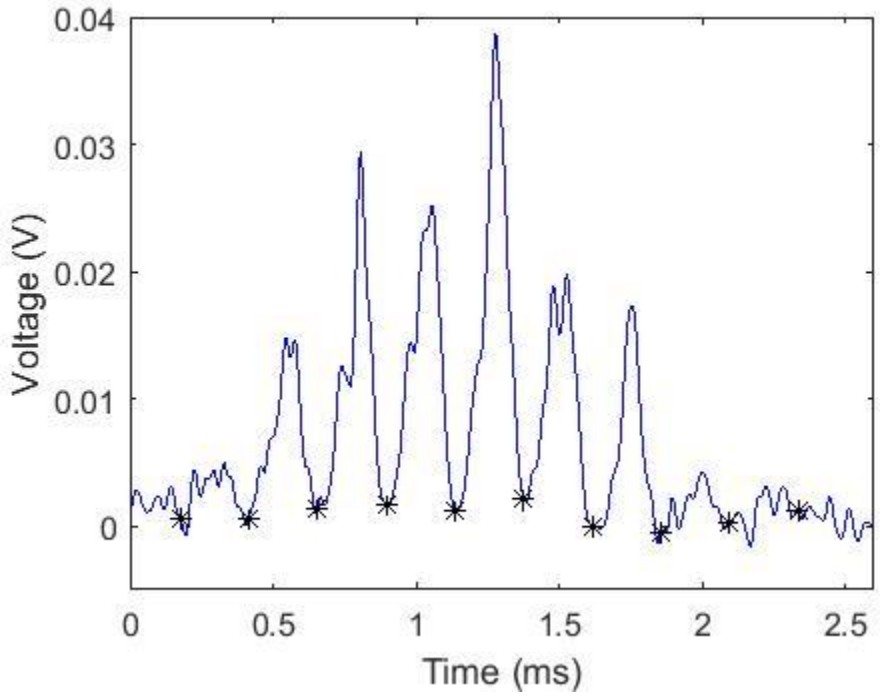


Figure 3.5 The result of searching for the best combination of starting time point (represented by "*") and number of sampling points for each peak.

Real time image contour detection. As shown in Figure. 3.6, in the contour finding algorithm all the pixels in a binary cell image are scanned. For those pixels of

non-zero value, the algorithm checks all eight pixels surrounding the center pixel in a 3x3 matrix. If the number of non-zero neighboring pixels is between 1 and 7, then this pixel is determined to be on the cell contour. Otherwise, the pixel is either inside or outside the contour. The criteria can be described in (4):

$$image\_contour(i, j) = \begin{cases} 1, & if \quad 0 < n < 8 \quad \& \quad binary\_image(i, j) = 1 \\ 0, & Otherwise \end{cases} \tag{4}$$

where n is the number of non-zeros pixels surrounding the center pixel of the 3x3 matrix.

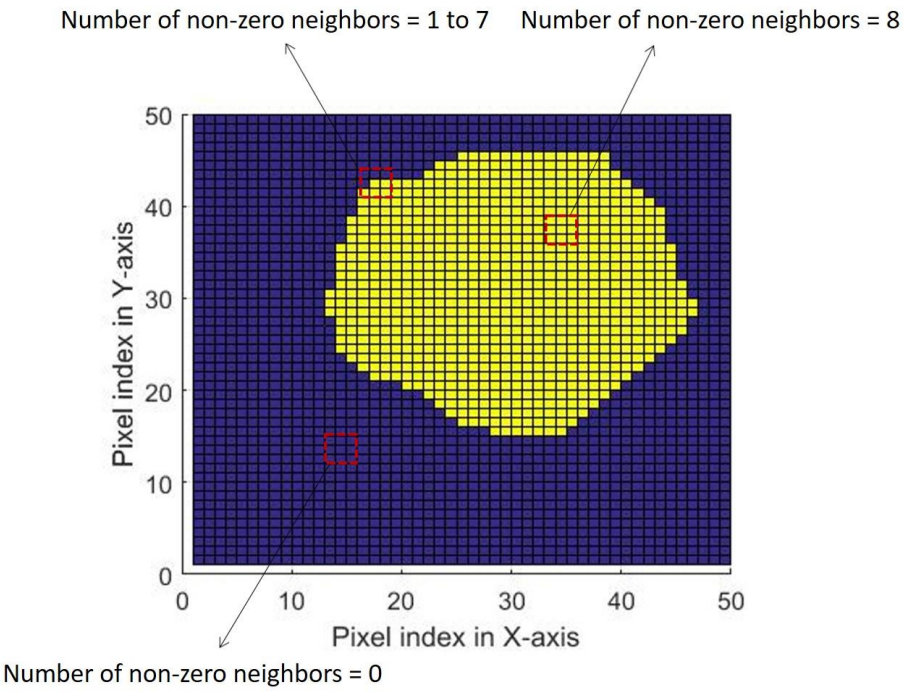


Figure 3.6 Criteria used to determine the contour of cell image.

Real time image processing algorithm for sorting MDCK cells according to the number of particles on cell surfaces.

The image processing algorithm for sorting MDCK Madin-Darby Canine Kidney Epithelial Cells based on the number of beads bonded to the cells is essentially the same as the previous cases except a top-hat filter with 7X7 pixels is used to extract features of small particles. The entire algorithm takes about 6 ms with current hardware. The flow chart of real-time image processing algorithm is shown in Figure. 3.7.

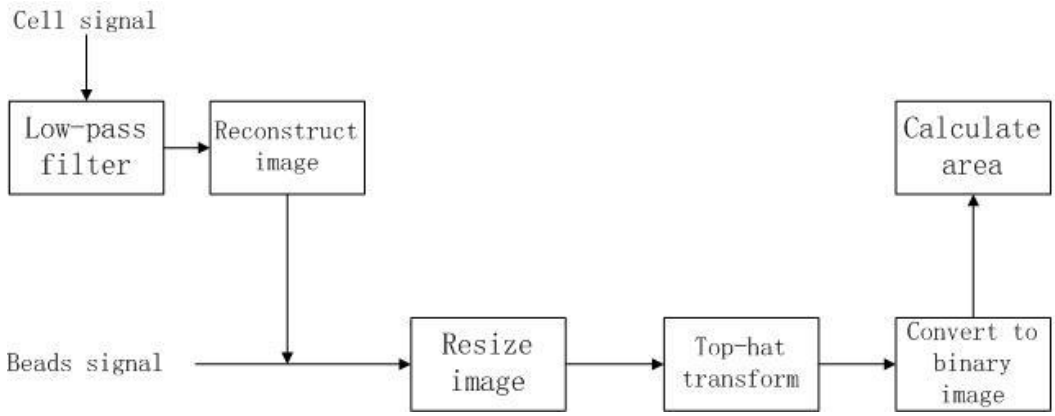


Figure 3.7 Flow chart of real time image processing algorithm for sorting MDCK cells by the number of particles bonded to the cells.

Real time image processing algorithm for sorting human glioblastoma cells by the extent of radiation induced DNA damage.

The image processing algorithm is illustrated in Figure. 3.8. The same top-hat filter with 7X7 pixels used before is also applied to remove background in gamma-h2ax image. Also the same algorithm is used to convert both the GFP image and the background removed gamma-h2ax image into binary images, and extract all image-derived parameters. The latency of the algorithm is about 6.7ms/cell.

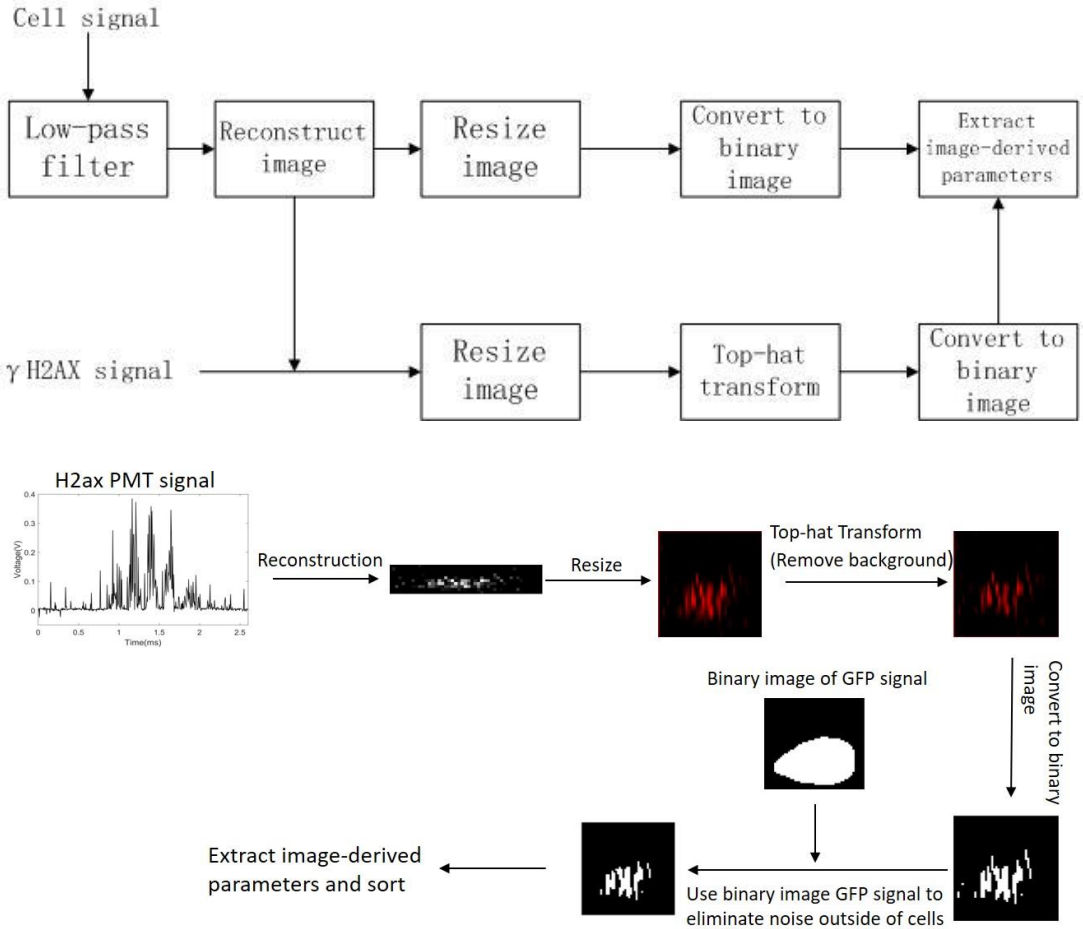


Figure 3.8 (a) Flow chart of real-time image processing algorithm for sorting human glioblastoma cells by the extent of radiation induced DNA damage. (b) Illustration of image processing algorithm for sorting human glioblastoma cells by the extent of radiation induced DNA damage.

3.4 Demonstrate the system by sorting different cell samples

3.4.1 Sorting cells by spatial distribution of specific protein

Spatial distribution of certain protein or organelles such as lysosomes and mitochondria carry important biological information. Image guided cell sorter is, to our knowledge, the only tool that can capture cells of sufficient quantity and purity based on such information. Here we demonstrate such functionality using pEGFP-GR plasmids translocated HEK-297T cells and un-translocated HEK-297T cells. pEGFP-GR expresses the eGFP protein fused to the N-terminal end of the glucocorticoid receptor.

Nuclear import and export of glucocorticoid receptor are important cellular processes related to numerous cancers, chronic inflammatory diseases and developmental disorders.[33-35] Protein translocation does not necessarily change the overall fluorescent or light scattering intensity of cells. Hence conventional flow cytometer is unable to distinguish translocated cells from un-translocated cells.

In the experiment, HEK-293T cells are transfected with GR-GFP and separated into 2 plates. One plate of cells is treated with dexamethasone that causes migration of GR-GFP protein from cytoplasm to nucleus. The other plate of cells is untreated so the GR-GFP protein stays in cytoplasm. The example microscope cell images are shown in Figure. 3.9(a). The mixture of both types of cells are flown through the system and imaged, and the interested subgroup can be isolated based on the real-time captured cell images.

For supervised machine learning, treated and untreated cells are flown through the system and all image-related parameters from their fluorescent and bright field

images are obtained. These extracted image parameters are used to generate gating criteria for real-time image guided cell sorting.

Figure 3.9(b) shows typical reconstructed cell images with bright-field image defining the cell boundary and the fluorescent image delineating the spatial distribution of GR-GFP protein. All image-derived parameters are also extracted to generate criteria for real-time image-based cell sorting. In this example, the image-derived parameters include fluorescent area, bright-field area, perimeters, shape factors (i.e. area/perimeter ratio), etc.

Next all image-derived parameters are ranked by their Receiver Operating Characteristics (ROC) score, which quantifies each parameter's ability to distinguish translocated cells from un-translocated cells. The three highest ranked parameters are selected for image-guided sorting parameters by default unless users enter extra inputs for the system to adopt different sorting criteria. In this experiment, the top 3 parameters are the area ratio between fluorescent and bright field signals, perimeter ratio of fluorescent and bright field signals, and total fluorescent area.

Next a 3D hyperplane separating two cell populations based on the selected top 3 parameters is formed by Support Vector Machine (SVM). This hyperplane (Figure 3.9 (c)) defines the criteria for real-time image-guided cell sorting.

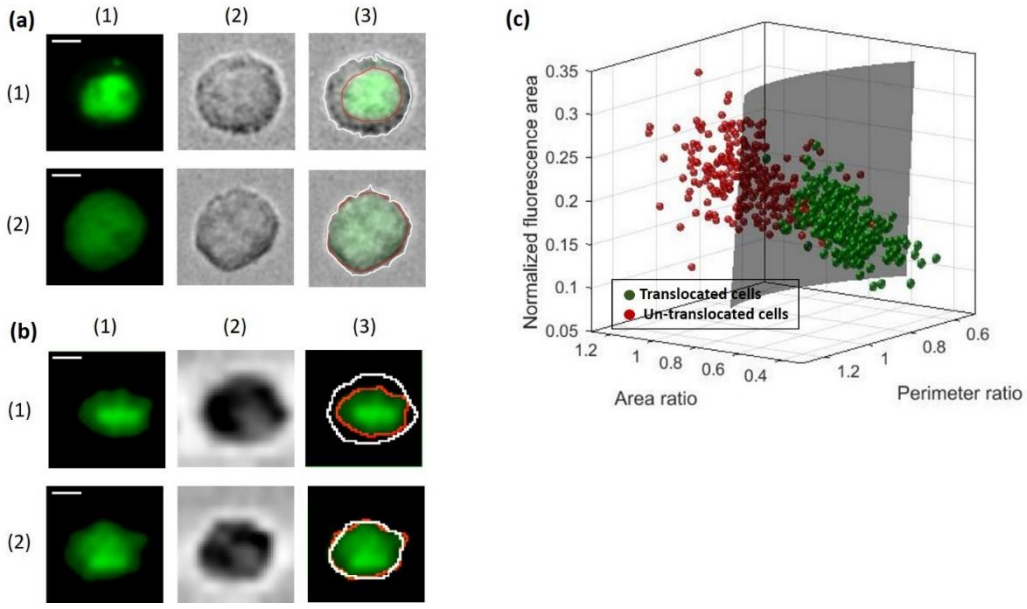


Figure 3.9 Sorting cells by spatial distribution of specific protein. (a) Example of microscope cell images. Row(1) shows translocated cells, and row(2) shows un-translocated cells. Column(1) shows fluorescent images, column(2) shows bright-field images, and column(3) shows overlaid images with their respective contours defined by the computer-generated red and white curves. (b) Example cell images generated by our system. Row(1) shows translocated cells, and row(2) shows un-translocated cells. Column(1) shows fluorescent images, column(2) shows bright-field images, and column(3) shows fluorescent images overlaid with bright-field images with their respective contours defined by the computer-generated red and white curves. (c) Hyperplane formed by SVM. A 5μm scale bar is shown in each row of micrographs.

3.4.2 Sorting cells according to particle binding on cell membrane

To show the capabilities of isolating cells based on surface markers, we sort MDCK cells based on the number of fluorescent particles bonded to the cell membrane. Fluorescent polystyrene beads (1μm diameter) functionalized with carboxylic groups can be adsorbed to almost any membrane proteins. By adjusting the concentration of the beads and cells in the mixture, the test system can produce a large variety of the number of beads bonded to each cell. Here we use a 20X objective lens

to increase the depth of focus, and the dimension of optical spatial filter is adjusted accordingly. The fluorescent signals of cells (520nm wavelength) and beads (645nm wavelength) are detected by 2 PMTs. The image processing algorithm is similar to the previous cases, including generation of fluorescent images (50x50 pixels) of the cells and the beads, noise suppression with a digital filter, finding image contours by converting gray scaled images into binary images with defined thresholds, and extraction of image-derived features (See more details in Methods section). Examples of the reconstructed images are shown in Figure 3.10(a) from which one can unambiguously enumerate the number of particles on the cell surface. Figure 3.10(b) is the histogram of normalized fluorescence area from the beads. One can see a strong correlation between the fluorescent area and the number of beads. The hyperplane generated by SVM separating cell populations bonded with different number of beads is shown in Figure 3.10(c). The top 3 parameters used to generate the 3D hyperplane are normalized fluorescent area from the beads, normalized perimeter of the beads and the net intensity within the central area of beads. The "net intensity within central area of beads" is different from the overall fluorescent intensity of beads because it includes signals only from the region of the highest intensity in each spot, thus minimizing the effects of background noise, blurring, and color bleeding in the signal of conventional FACS systems.

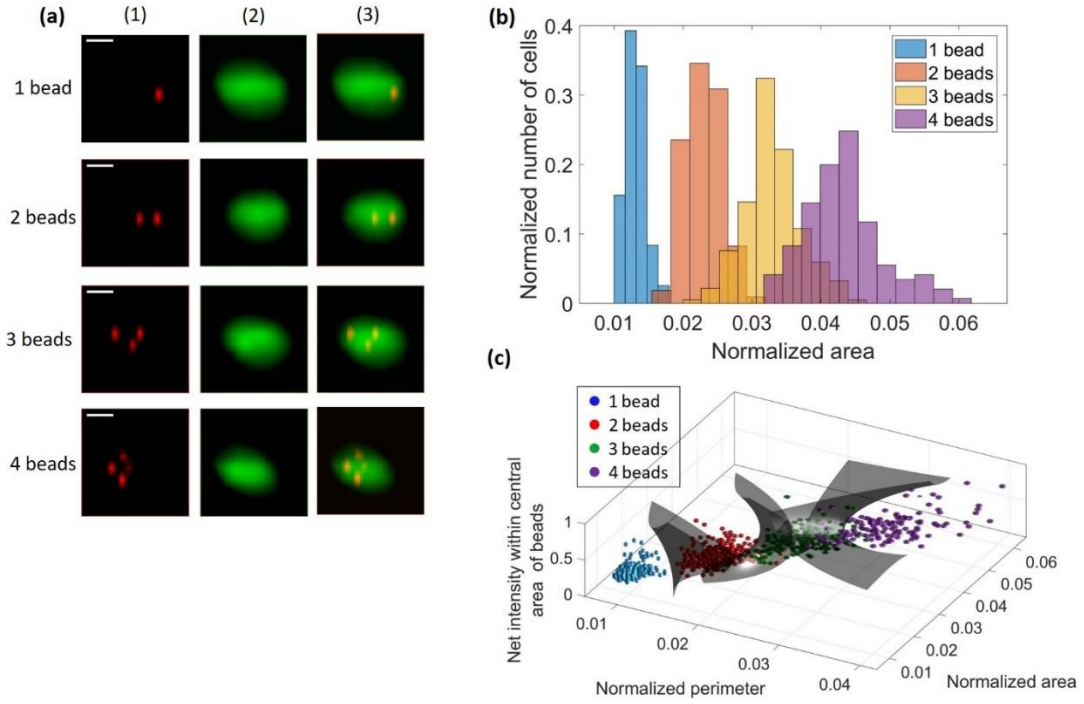


Figure 3.10 Sorting cells according to particle binding on cell membrane. (a) Example cell images generated by our system. Column(1) shows fluorescent images at 645nm, column(2) shows fluorescent images at 520nm, and column(3) shows overlaid images. (b) Histogram for normalized fluorescent area from beads. (c) Hyperplane formed by SVM. A 5μm scale bar is shown in each row of micrographs.

3.4.3 Sorting cells by the extent of radiation damage

Next we report experiment of sorting human glioblastoma cells by the extent of radiation induced DNA damage. The number of gamma-h2ax foci is a key parameter measuring the number of DNA double-strand breaks (DSB) induced by cytotoxic agents including ionizing radiation.[36,37] Currently, fluorescent microscopy equipped with high throughput image processing is used to count gamma-h2ax foci number as a measure of DNA damage.[38] However, besides the lower than desired throughput for microscopy, no technique can efficiently isolate cells according

to cell's resistance to radiation or cytotoxic damage to support downstream molecular analysis over the targeted cell groups. The throughput of laser microdissection is much too low to support such studies. Image-guided cell sorter fills the technology gap by using machine learning and real time image processing to sort cells based on the foci count of gamma-h2ax that directly delineates the sections of broken DNAs by ionizing radiation.

GFP transfected human glioblastoma cells (GBM-CCC-001) are treated with 6Gy irradiation, and then the cells are fixed by paraformaldehyde. The fixed cells are stained with primary antibody (mouse-anti-gH2AX) and secondary antibody conjugated with fluorophore (PerCP-Cy5.5). The GFP signal delineates the area of the entire cell and the distribution of fluorescent signals from gamma-h2ax represents the fragments of double strand DNAs broken by radiation.

As stated previously, the spatial resolution of the raw image of our system is 2μm by 0.4μm. Hence image processing is required to enhance the effective resolution to resolve the gamma-h2ax foci smaller than 1μm diameter. The typical images of the fluorescent spots from gamma-h2ax foci are shown in Figure 3.11 (a). Here off-line compressive sampling algorithm is used to produce higher spatial resolution (0.4μm x 0.4μm) images from the lower resolution (2μm x 0.4μm) raw images.[39,40] Such off-line processed higher resolution images are only displayed to users, but not used for real time image-derived parameter extraction.

The compressive sampling reconstruction can be described by equation 5, where $I_{measure}$ is the background removed image, $f(x, y)$ is the spatial mask, and

$\psi(x, y)$ is the point spread function of objective lens (NA=0.55) approximated by a Gaussian function with 0.35μm root-mean-square (RMS) width, and I_{Foci} is the gamma-h2ax fluorescence distribution. I_{Foci} is solved using ℓ_1-Regularized Least Squares Solver.[41] Using the solved I_{Foci}, the image $I_{Objctive}$ formed by objective lens (NA=0.55) is calculated by equation 6. The image $I_{Objctive}$ has higher resolution compared to the raw image $I_{measure}$, which has 2 μm by 0.4 μm asymmetric resolution equivalent to a 10x50 pixel image. The example cell images reconstructed by compressive sampling algorithm are shown in Figure 3.11(b).

$$I_{measure} = f(x, y) \otimes \psi(x, y) \otimes I_{Foci} \quad (5)$$

$$I_{Objctive} = \psi(x, y) \otimes I_{Foci} \quad (6)$$

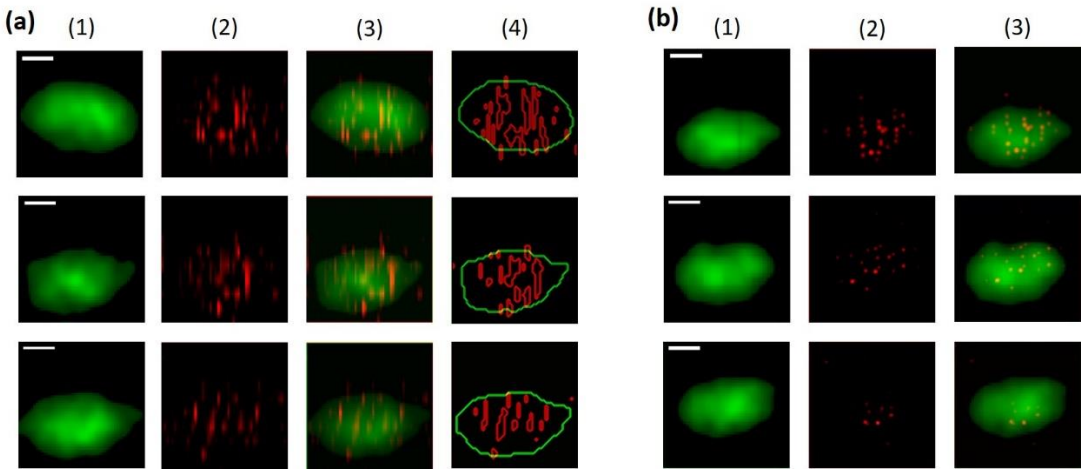


Figure 3.11 Example images of irradiated cells. (a) Example cell images generated by real-time algorithm. Column(1) shows GFP images, column(2) shows gamma-h2ax images with background removed , column(3) shows overlaid images, and column(4) shows image contours, with green contour for GPF image and red contours for gamma-h2ax images. (b) Example cell images generated by off-line processing for human vision. Column(1) shows GFP images, column(2) shows gamma-h2ax images, and column(3) shows overlaid images. A 5μm scale bar is shown in each row of micrographs.

Although compressive sampling algorithms reconstruct images with high resolution, it is not used for image-guided sorting in our experiment because of the limited processing speed of our current hardware. In the experiment reported here we still use the lower resolution raw images in Figure 3.11(a) to guide cell sorting. Two image-derived parameters extracted from the real-time reconstructed images are used to estimate the foci count: the total perimeter of gamma-h2ax images and the net gamm-h2ax intensity within the central area of each focus area. The latter is different from the overall fluorescent intensity of gamm-h2ax because it includes signals only from the region of highest intensity in each focus area, thus minimizing the effects of background noise, blurring, and color bleeding in the signal of conventional FACS systems. Figure 3.12(a-c) shows how conventional intensity-based sorting is compared against image-guided sorting with the above image-derived parameters.

To obtain the ground truth for each cell being investigated, we apply off-line processing to resolve high-resolution images of 1800 cells, and then find how the total fluorescent intensity and real-time image-derived parameters are related to the ground truth images. The actual foci count is derived from the ground truth images, and Poisson regression is used to predict the foci count based on total fluorescent intensity and real-time image-derived parameters. The scatter plot of predicted foci count and actual foci count are shown in Figure 3.12(a) and (b), and the respective R-squared values are calculated. Using total fluorescent intensity to predict the foci count, the R-squared value is found to be 0.632±0.029. In contrast, using real-time image-derived

parameters to predict foci count, the R-squared value is increased to 0.881±0.008, significantly greater than intensity-based sorting. Figure 3.12 (d) and (e) shows histograms of foci count predicted by total gamma-h2ax fluorescent intensity and predicted by image derived parameters. Here bin1 includes cells with 16~17 foci and bin2 includes cells with 25~31 foci. Consistent with the conclusion from Figure 3.12(a-c), Figure 3.12 (d) and (e) indicates that intensity based sorting between bin1 and bin2 has greater overlap (i.e. ambiguity) than image based sorting. The result can be more quantitatively represented from the Receiver Operating Characteristic (ROC) analysis in Figure 3.12 (f). The ROC analysis indicates superior performance of image-guided sorting compared to conventional intensity-based sorting. Also notably in Figure 3.12 (c), there are about 1% "outliers" in the scatter plot of total gamma-h2ax fluorescent intensity versus foci count, indicating the presence of a small population of cells with exceptionally high gamma-h2ax intensity but not particularly large foci count. If cells of such properties are to be studied, one can use sorting criteria that combine intensity and image-derived parameters to isolate such rare population. On the other hand, using only intensity for sorting criteria as in conventional FACS, we would misclassify those high-intensity but moderate foci count cells, presenting the risk of missing vital biological information. This is just another example of the unique capability image-guided cell sorting can bring to the research community.

To quantitatively evaluate the performance of image-guided sorting and intensity-based sorting, we have compared sorting purity and yield by isolating cells with more than 23 foci using both techniques. Aw shown in Figure 3.13 (b), image-

guided sorting with 2µm x 0.4µm spatial resolution has consistently shown superior performance than conventional intensity-based sorting.

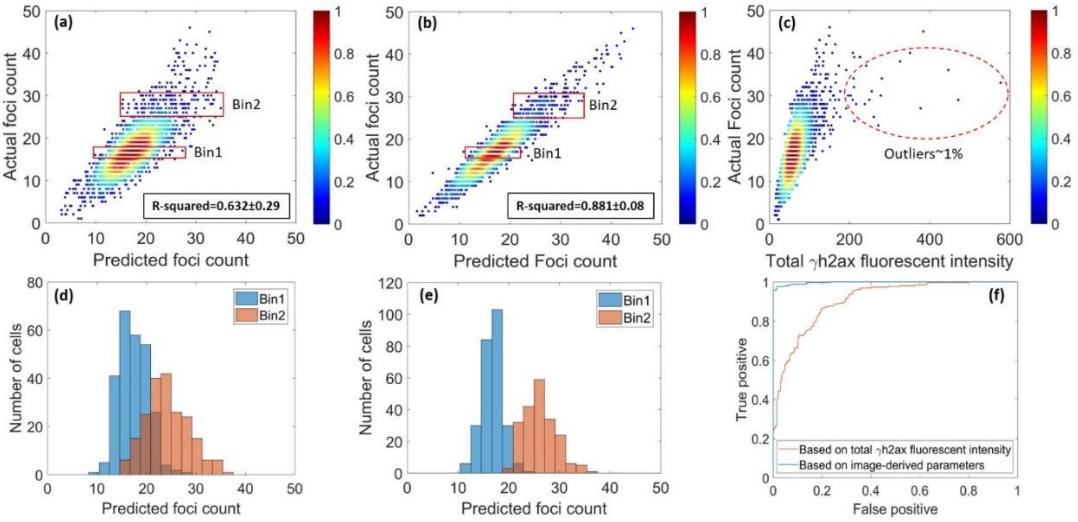


Figure 3.12 Estimation of foci count based on image-derived parameters and total fluorescent intensity. (a) Scatter plot of predicted foci count based on total gamma-h2ax intensity versus actual foci count (b) Scatter plot of predicted foci count based on image-derived parameters versus actual foci count (c) Scatter plot of total gamma-h2ax intensity versus actual foci count including outliers (d) Histogram of predicted foci count based on total gamma-h2ax intensity of bin1 and bin2 (e) Histogram of predicted foci count based on image-derived parameters of bin1 and bin2 (f) Receiving Operating Characteristic(ROC) analysis of bin1 and bin2, showing superior performance of image-guided sorting than conventional intensity-based sorting.

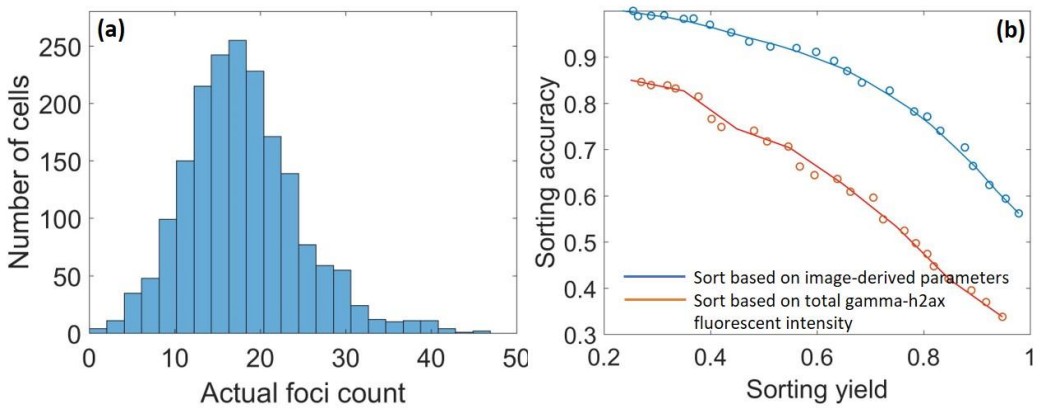


Figure 3.13 Sorting purity and yield. (a) Histogram of actual foci count for Gamma-ray irradiated cells. (b)Sorting yield versus sorting accuracy for isolation of cells with greater than 23 foci using image-guided and intensity-based methods.

Chapter 3 is based on and mostly a reprint of: **Yi Gu**, Alex Ce Zhang, Yuanyuan Han, Jie Li, Clark Chen, and Yu-Hwa Lo. "Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification." **Cytometry Part A** (2019); **Yi Gu**, Rui Tang, Alex Ce Zhang, Yuanyuan Han and Yu-Hwa Lo, "Image-guided microfluidic cell sorter with machine learning", **CLEO** (2019)， Control Number: 3080140; Yuanyuan Han, **Yi Gu**, Alex Ce Zhang, and Yu-Hwa Lo. "Imaging technologies for flow cytometry." **Lab on a Chip** 16, no. 24 (2016): 4639-4647. The dissertation author was the primary author of the work.

Chapter 4 Image-Guided Cell Sorting Using Laser Point Scanning

4.1 System Design of Image-Guided Cell Sorter Using Laser Point Scanning

The overall work flow of the 2D image guided cell sorter is shown in Figure. 4.1. First, as cells flow through the microfluidic channel, each cell image is converted to a temporal waveform produced by the fluorescent or scattering/transmission signal resulted from the scanning laser excitation beam. Second, the temporal fluorescent or optical scattering/transmission waveforms are detected by photomultiplier tubes (PMTs) to become electronic waveforms. These electronic waveforms will then be reconstructed to cell images. Third, the features of the constructed cell images are extracted and available to users to allow users to define or modify the gating criteria for cell isolation. Fourth, the imaging features of each cell passing the interrogation area are calculated in real time and those cells with features meeting the sorting criteria (i.e. within the regime of defined gating) will be sorted.

The optical setup, image reconstruction, imaging feature extraction, gating strategy and real-time processing module will be described in detail in the following sections. The design can support a throughput of over 1000 cell/s, which is several orders of magnitude greater than any microscopy-based cell isolation systems such as laser micro dissection systems.

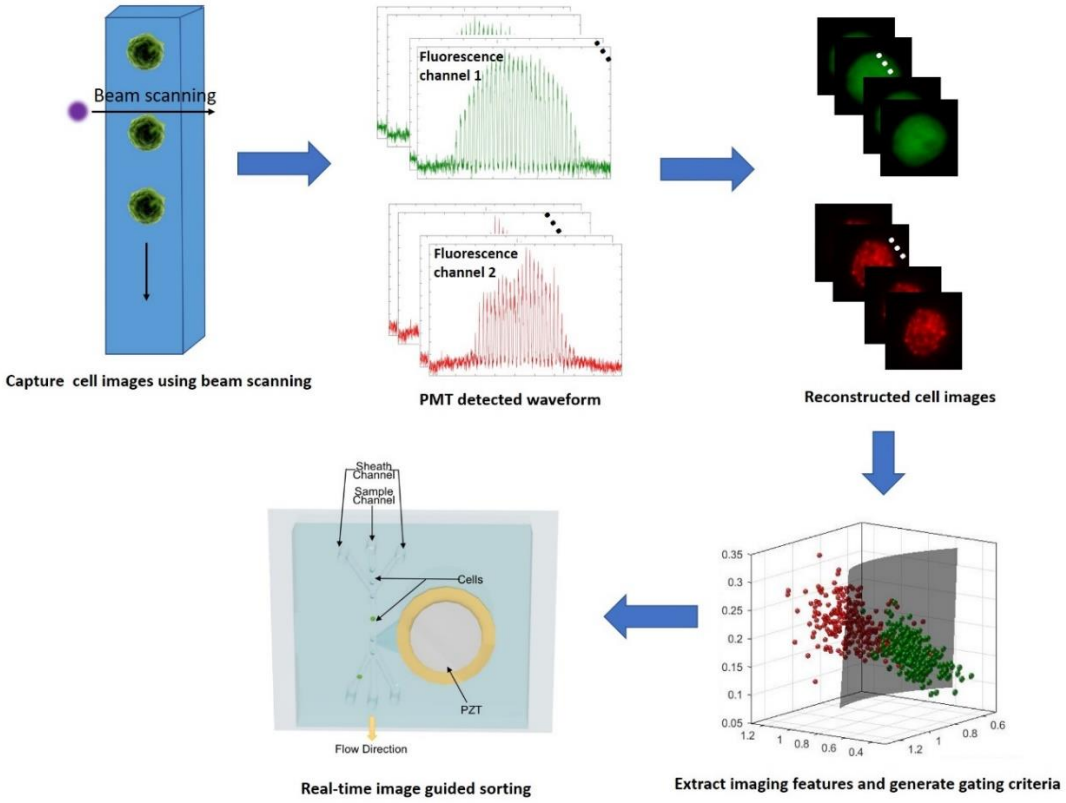


Figure 4.1 Schematic overview of the work flow for an image-guided cell sorter

4.2 Optical Imaging Setup

As shown in Figure. 4.2(a), suspended cells flow in a microfluidic channel. Sheath flow is used to hydro-dynamically focus the travelling cells to the center of the microfluidic channel. The flow direction is along the z-axis in Figure 4.2(a). At the optical interrogation zone, each flowing cell is illuminated simultaneously by a 455nm wavelength LED and a 488nm wavelength laser. The 455nm LED is used to measure the travelling speed of each cell as such information is needed for cell image construction. The 488nm wavelength laser is coupled to an AOD modulator to create a scanning beam with a small (<1 um diameter) spot size for generation of optical

transmission, scattering, and fluorescent signals through fluorophores. In this setup, 10X objective lenses are used for both laser illumination and imaging to obtain high depth of focus across the cell thickness. Besides the laser beam used for producing cell images, the system also has an LED light source. The purpose of the LED is to help us measure the travel speed of each individual cell. As to be explained later, knowing the cell speed is critical to correct construction of cell images without distortion.

Here we explain how to use the LED signal to measure cell speed. The LED light passes the sample and the objective lenses, is reflected by the first dichroic mirror (DM1), and transmits through the spatial mask1 before reaching the high-speed photodetector (PD) to produce a photocurrent. The spatial mask1 has a design of two slits separate in the z-direction (i.e. cell flow direction). The transmitted LED light through the slit produces a dip in the light intensity each time a cell travels through the light path due to light scattering or absorption. The time interval between the two photocurrent dips produced by these two slits can thus give rise to the information of cell travel speed since we know the physical separation of the two slits and the magnification factor of the optical system. The detailed design of spatial mask 1 is shown in Figure. 4.2(c). The spatial mask1 is placed at the image plane of the optical system and consists of 2 parallel slits that are 10μm wide and 1mm long. The center-to-center distance between these 2 slits is 200 μm.

The 488nm wavelength laser illumination is used to generate transmission images and fluorescence images via excitation of fluorophores. The laser illumination

is focused to a diffraction-limited light spot by the objective lens on the left side of the sample. As shown in Figure. 4.2(e), the laser spot is scanned in the y-direction driven by an acoustic optical deflector (AOD). To be explained next, this fast laser scanning system will enable the production of two-dimensional cell images over the y-z plane where y-axis is the beam scanning direction and z-axis is the cell flow direction.

At the image plane for the laser beam, there is a spatial mask2 with a 500μm by 500μm optical window. The actual field-of-view at the image plane for the scanning laser is designed to be 200μm by 200μm, hence in principle, the spatial mask2 can be treated as an optical window. The purpose of spatial mask 2 is to block tray light to enhance the sensitivity of the system.

Dichroic mirrors are used to route the desired emission bands to their respective PMTs. The field of view is determined by the scanning range of AOD and the signal recording time period. In this setup, the field of view is chosen to be 20μm by 20μm, which covers the size of most biological cells. The resolution is determined by diffraction limit of the objective lens. In this setup, the resolution is 1μm.

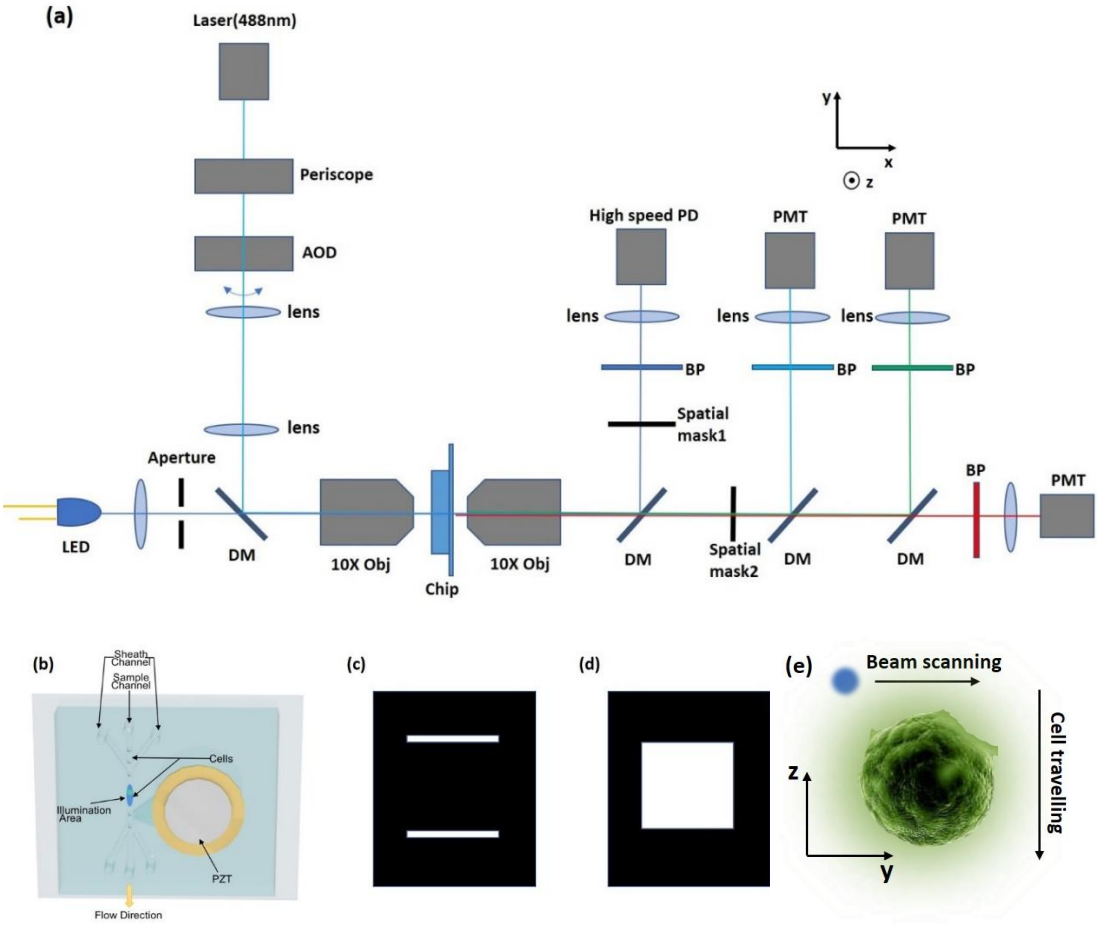


Figure. 4.2 (a) The schematic diagram of the optical setup. As cells flow through the microfluidic channel (along the z-axis), the AOD scans the laser beam spot to produce time domain signals at the output of PMTs. The PMT output signals are acquired and processed by a real-time processing module (e.g. a FPGA) to make the sorting decision. Then the piezoelectric PZT actuator is actuated to sort cells based on the sorting decision. PD is photodetector, BP is bandpass filter, DM is dichroic mirror, PMT is Photomultiplier tube. (b) Microfluidic device with an on-chip piezoelectric PZT actuator to deflect selected cells in the microfluidic channel for image-guided cell sorting. (c) Design of spatial mask1. (d) Design of spatial mask2. (e) Illustration of beam scanning. Cell is travelling in z-axis and beam is scanning in y-axis

4.3 Real-Time image processing and cell classification

4.3.1 Real-Time processing system design

Sorting cells from a flow requires the abilities to compute the image features of each cell, compare these parameters against the sorting criteria, gating, and activate the actuator to isolate the selected cells from the stream; and all of these processes have to be completed rapidly to achieve high throughput, which is typically 1000 cells/sec for flow cytometers. Therefore, real-time processing for computation of cell image features are needed. To accelerate processing speed, one preferred embodiment is a FPGA-CPU hybrid design. The diagram of real-time processing system is illustrated as Figure 4.3. The PMT signals and the voltage signal modulating the AOD are acquired and digitized, and the LED signal from a photodetector is acquired by an Analog Input (AI) module. First, the FPGA reconstructs the cell images from the acquired signals. Then, the reconstructed cell images are transferred to the CPU for imaging feature extraction. After the image features are computed, the features are transferred back to the FPGA. Sorting decision is made based on the extracted imaging features compared against the defined sorting criteria. To sort a cell, a voltage pulse is applied to the on-chip piezoelectric PZT actuator, which instantaneously bends the bimorph PZT disk to deflect the cell away from the central flow into the sorting channel.

The integrated piezoelectric actuator offers a desirable design to sort cells due to is low cost, easy operation, and gentleness that results in high cell viability, but

other mechanisms, including electrophoresis, dielectrophoresis, electric static, mechanical, acoustic, optical tweezers, etc. can also be used to sort cells.[42-45]

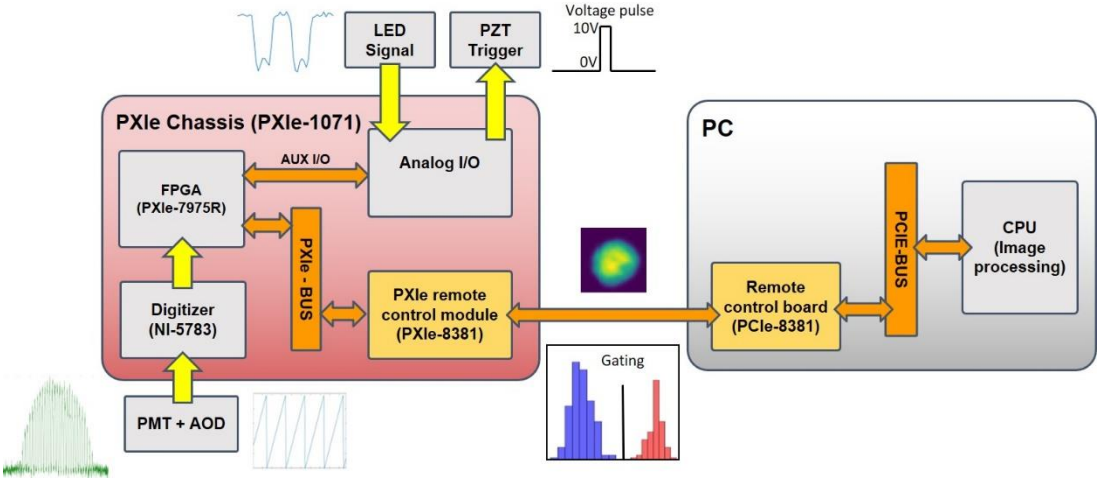


Figure. 4.3 Schematic diagram of real-time processing module

4.3.2 Image reconstruction algorithm

Image construction process contains three steps: (1) conversion of the temporal waveform of PMT into an uncorrected 2D cell image, (2) correction of effect caused by a time delay between the modulating voltage applied to the AOD and the corresponding laser beam spot, which we call correction of "phase shift" below, and (3) correction of image distortion due to the effect of cell travelling speed, which we call "image resizing" as errors in cell speed measurements can result in contraction or stretch of the cell image. In the following we describe each step in detail:

(1) Conversion of the temporal waveform of PMT output into an uncorrected 2D cell image.

As shown in Figure. 4.4(a), we first find the troughs and peaks of the voltage waveform that modulates the AOD. For simplicity, we normalize the waveform plot to set the peak to be 1 and trough to be zero. The waveform that modulates the AOD rises linearly with time from trough to peak and drops rapidly from peak to trough. Such waveform continues over time to form a periodic sawtooth waveform. Within each sawtooth period, the AOD scans the laser spot from one extreme position to another at a uniform speed along the y-axis, and then the laser spot returns quickly back to the starting position for the next scan. In the current design, we choose a liner scan considering the simplicity of computation and the properties of acoustic optic crystal. Other scanning waveforms other than the sawtooth waveform can also be used.

Each time the AOD scans the laser excitation beam spot in the y direction, the resulting fluorescent or transmission signals registered by the PMTs are registered to form a 1D slice of the cell image along the y direction. Since the cell is also travelling in the z-direction during laser scanning, the laser y-scanning actually produces a line scan of the cell image with a small angle ($\theta \sim \frac{\mathrm{v}_{cell}}{v_{scan}}$; v_{cell}: $cell\ flow\ speed$, v_{scan}: $laser\ beam\ scanning\ speed$)

with respect to the frame of the cell. Due to the much faster laser beam scanning speed (e.g. 500cm/s for v_{scan} and 20cm/s for v_{scan}), such effect is rather small and can be neglected or easily corrected, if needed.

Ignoring the small angle between the y-axis for the laboratory frame and the cell frame, one can consider that each time the laser beam makes a line scan along the

y-axis, the cell image features in the y-axis at a new z-position are recorded. Mathematically, each cell travelling through the optical interrogation area will produces a series of image data registered as $S^{z_1}(y_1), S^{z_1}(y_2) \ldots . S^{z_1}(y_N)$; $S^{z_2}(y_1), S^{z_2}(y_2) \ldots . S^{z_2}(y_N)$; $\ldots S^{z_M}(y_1)$, $S^{z_M}(y_2) \ldots . S^{z_M}(y_N)$. Obviously the above data set can be easily arranged into an MxN matrix, representing a 2D cell image in the y-z plane.

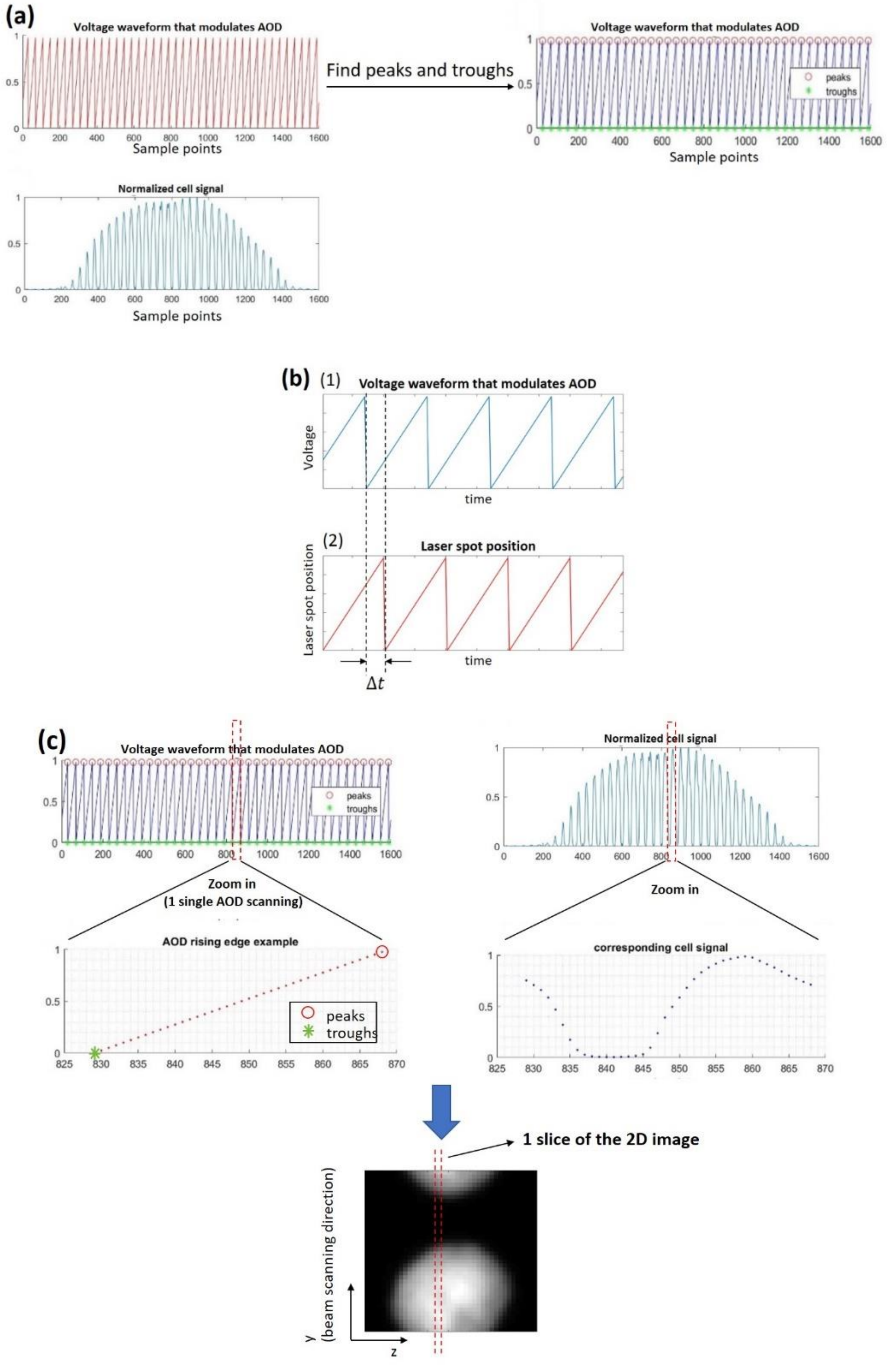


Figure. 4.4 (a) Finding troughs and peaks of the AOD signal. (b) Illustration of the time delay between the modulating voltage applied to the AOD and the corresponding laser beam spot. (c) Form 1D slice of the image based on each rising edge and the corresponding PMT detected signal. And splice the 1D images to reconstruct the 2D cell image.

Correction of the time delay between the modulating voltage applied to the AOD and the corresponding laser beam spot, which we call correction of "phase shift" below. Under DC bias condition, the laser beam spot should match the applied voltage to the AOD, having the beam spot at one extreme position in y-axis at the minimum bias voltage and the opposite extreme position at the maximum bias voltage. However, when a sawtooth voltage waveform is applied to the AOD, there can be a time delay between the actual laser beam spot position and the voltage value due to the electric capacitance of the AOD and parasitic capacitance. The effect of time delay or phase shift is illustrated in Fig. 4.43(b). In Fig. 4.4(b), plot(1) is the voltage waveform that modulates AOD, and plot(2) is the actual laser beam position. As shown in the plot, there is a time delay Δt, which we call "phase shift". Such phase shift will produce a 2D cell image similar to the one in Fig. 4.4(c) where the lower part of the cell appears to be atop the upper part of the cell. To correct such phase shift, we take the image of transmitted signal, which shows the contour of the cell in the y-z plane. By choosing one or a few z-positions and plot the transmitted light intensity along the y- (laser scanning) direction, we obtain an intensity plot similar to Fig. 4.5(c) where the intensity profile does not appear to show a continuous profile but be divided into two regions separated by a section with minimum intensity. By shifting the leftmost region to join the right most region to make a continuous intensity profile and to center the entire intensity profile within the field of view, we can obtain the "phase-shift corrected image" shown in Fig. 4.5(d). The reconstructed image in Fig. 4.5(d) consists of 40 by 40 pixels with a pixel size of 0.5 um.

Note that although we only correct the phase shift for the transmitted image, the same correction for the phase shift is also applicable to the fluorescent signals and other scattering signals since they are all produced by the same scanning laser beam and synchronized with each other.

(3) Correction of the image distortion due to effect of cell travelling speed, which we call "image resizing" as errors in cell speed measurements can result in contraction or stretch of the cell image.

Even with flow confinement, cells may not be at the same position in the cross section of the flow channel. The actual cell positions in the flow channel depend on the cell size, cell stiffness, and the detailed design of the flow channel. For cells in a microfluidic channel with 2D instead of 3D flow confinement, the cell distance from the ceiling (or floor) of the channel can vary appreciably. In a laminal flow where cell speed is determined by the cell position within the channel, the above effects can produce appreciable speed variations among cells. Since the process of cell image construction converts a temporary signal (detected by PMTs) into a spatial signal where cell travelling speed determines how a certain time period is transformed into a certain length in space, variations in cell speed can distort the cell images without correction.

Specifically, we consider a cell moves at a speed v in the flow (z-) direction and for each time interval Δt, the laser spot scans through the y-axis once and producing a series of signal: $S^{z_i}(y_1), S^{z_i}(y_2) \ldots . S^{i}(y_N)$. Then the next scan will produce signal series: $S^{z_{i+1}}(y_1), S^{z_{i+1}}(y_2) \ldots . S^{i+1}(y_N)$

where $z_{i+1} = z_i + \Delta z = z_i + v\Delta t$

For a specific example, for an AOD operating at 400KHz, each scan takes 2.5 μs (i.e. $\Delta t = 2.5\mu s$). If a cell travels at a speed of 20 cm/s, $\Delta z = v\Delta t = 0.5\mu m$. The amount of time of taking 40 scans covers a distance of $\Delta z x 40 = 20\mu m$, which defines the field of view along the z-axis. However, if another cell travels through the optical interrogation area at a speed of 25 cm/s, $\Delta z = v\Delta t$ becomes $0.625\mu m$. Then the same amount of time of taking 40 scans would cover a distance of $25\mu m$ along the z-direction, or effectively the field of view along the z-axis increases with the cell travel speed. If we do not know the travel speed of individual cells and take an average cell speed to construct cell images, then cells travelling faster than the average speed will appear to be smaller along the z-axis than their actual size, and cells travelling at lower than the average speed will appear to be larger than their actual size. Therefore, the process of resizing is necessary to construct the cell image correctly without distortion.

From the above discussion, obviously one needs to know the travel speed of the cell, v, accurately to produce the correct cell image in the y-z plane. The transmitted LED light through the two slits on the spatial mask1 produces the signal that allows us to precisely measure the speed of each single cell, and the knowledge of speed enables us to put the proper length scale in the z-direction for the cell image without distortion.

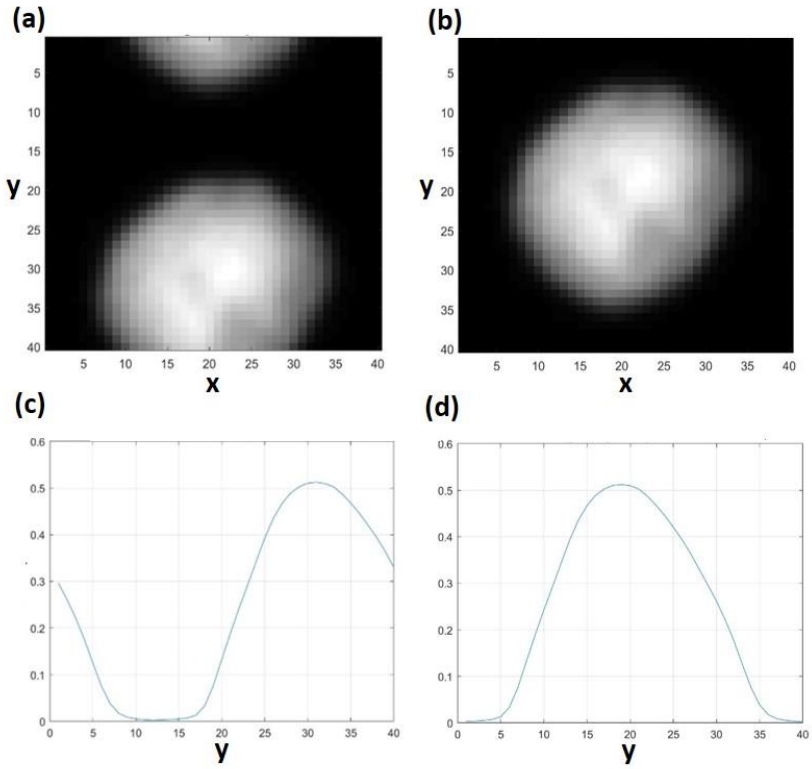


Figure 4.5 (a) Reconstructed image before phase shift correction (b) Reconstructed image after phase shift correction (c) signal integration along x-axis before phase shift correction (d) Signal integration along x-axis after phase shift correction

4.3.3 Feature extraction and gating strategy

Feature extraction While cell images provide intuitive and rich information through visualization, the information can be best managed and utilized by extraction of key features from these images given the extremely large number of images (often well over 1 million cell images) produced in each experiment. Generally the biologically and clinically relevant image features can be divided into two groups: the features from each individual parameter and features from correlations of two or more parameters. The commonly used features from each parameter includes area,

perimeter, shape, major and minor axis length, circularity, concentricity, aspect ratio (major axis length/minor axis length), integrated intensity, mean intensity (intensity divided by area), standard variation of intensity over space, granularity, texture entropy, texture variance, spot count, etc. For relations between two or more parameters, they include area of overlap, correlations in intensity distribution, intensity variance, area, etc. To perform image-guided cell sorting of single cells in a flow environment, it is required that (a) these image features of each cell passing the interrogation zone (i.e. scanning laser beam) need to be computed in real time, typically in less than 1 ms and (b) one needs to properly define "gating" based on some of the image features as the criteria for cell sorting. In the following we will discuss the definition of "gating" and the methods for real time computation of image features for each cell.

Gating strategy In general one can take two approaches to define gating, one based on machine learning or deep learning[32,46], and another based on a methodology of user-interface (UI) and user-experience (UX) interaction. There have been many established machine learning and deep learning algorithms that can be tailored to the cell sorting application since here the image objects are distinct, relatively well defined, and standing out from the background. Here we describe in more detail the second approach involving machine/user interactions as a preferred embodiment from the user perspective. Because this approach is similar to the current method of operating conventional flow cytometer and cell sorter, it is easier to learn

and adaptive to the new image features the new system offers. The operation procedures are outlined in Figure. 4.6.

First, a user selects the interested population based on signal intensities, a method similar to the operation of conventional flow cytometers and familiar to all flow cytometer users. Second, the histograms for each image feature within the selected population are displayed. Users can go through these histograms of image features to decide which features are most relevant to the intended applications. From these histograms, users can easily and intuitively tell if the originally selected cell population can be further divided into subpopulations characterized by these image features (e.g. spatial distribution of the fluorescent intensity, shape or size of cells or organelles). Then users can refine the gate to select one or multiple subpopulations of cells according to a particular set of image features. Thus cell sorting is guided not only by conventional fluorescent or scattering signals but also by image features.

Of course users always have the choice to define the gating (cell sorting criteria) based on the image features (e.g. shape and size of cell body or nucleus) without starting with conventional optical fluorescent or scattering signals. This can be a popular mode of operation for label-free sorting of lymphocytes or CTCs where the cell and nucleus shape/size show distinctive characteristics and rich information compared to traditional forward and side scattering signals.

To assure the image-guided sorting criteria satisfy the user needs or consistent with the user knowledge about their samples, reconstructed cell images inside and

outside the “gated areas” can be displayed, allowing users to check the actual images of cells they choose to sort or exclude. Users can repeat the steps shown in Figure 4.6 (c) and (d) to fine tune and finalize the sorting criteria.

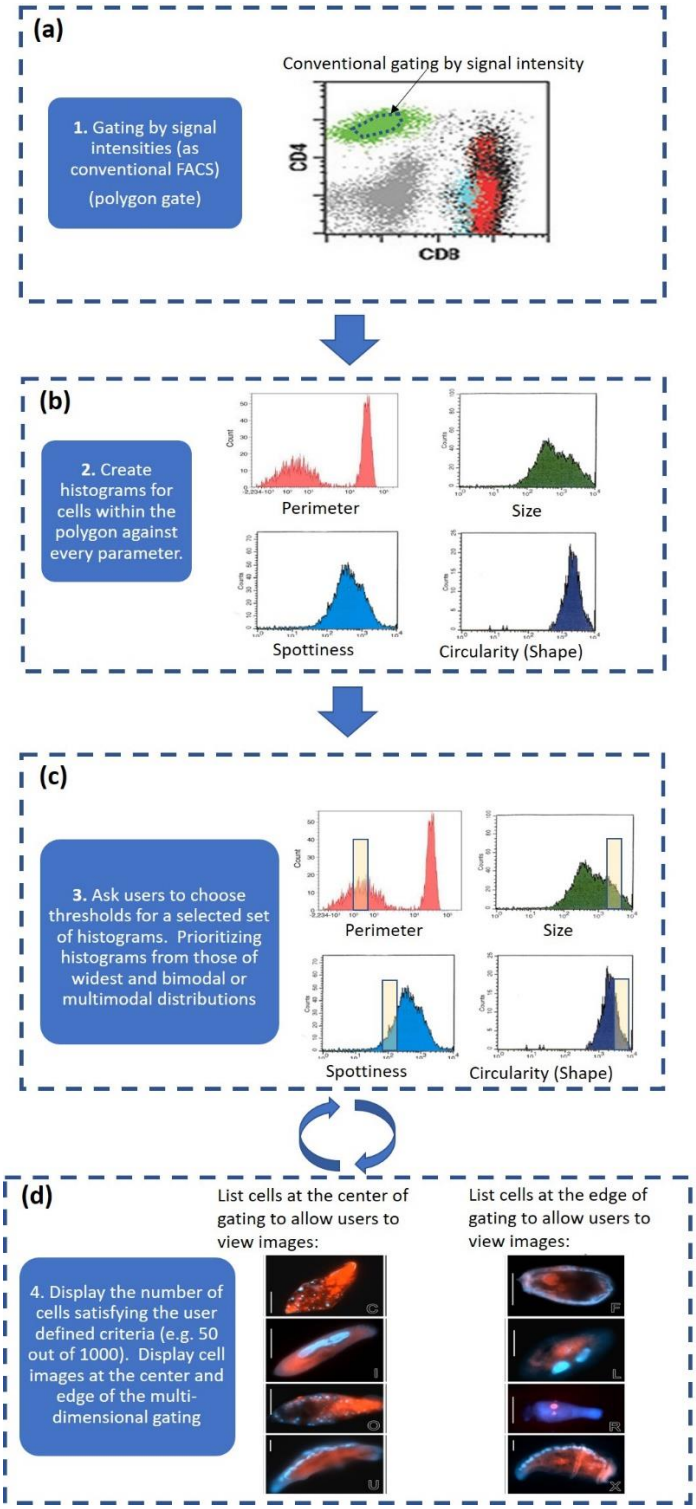


Figure 4.6 Flow chart of gating strategy

4.4 Demonstrate the system by sorting experiments

4.4.1 Sorting of beads by diameter

We demonstrate the image-guided cell sorter by sorting beads by diameter. Polystyrene beads of 2 difference diameters are mixed and isolated by the image-guided cell sorted based on the beads diameter extracted in real-time. The example images generated by the image-guided cell sorter is shown in Figure. 4.7(a), and the diameter histogram is shown in Figure 4.7(b)

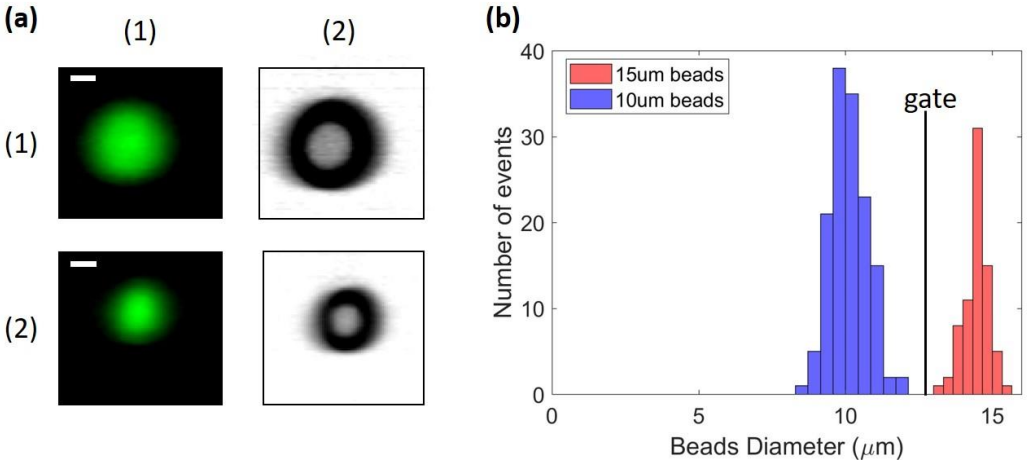


Figure 4.7 Image and histogram of polystyrene beads. (a) Example images of polystyrene beads generated by the Image-Guided Cell Sorter. The first column is fluorescence images, the second column is transmission images. The first is the beads with 15μm diameter, the second row is the beads with 10 μm diameter. (b) Histogram of diameter measured by the Image-Guided Cell Sorter.

In the sorting experiment, polystyrene beads with 15μm and 10μm diameter are mixed at ratio 50:50. Then the mixture is flow through the image-guided cell sorter and only beads with 15μm diameter are sorted and collected. Then the collected beads

are imaged by a fluorescent microscope. Based on the 170 images captured by fluorescent microscope, a 97% purity is calculated.

4.4.2 Sorting of cells according to number of bonded beads

To show the capabilities of isolating cells based on surface markers, we sort HEK cells based on the number of fluorescent particles bonded to the cell membrane. Fluorescent polystyrene beads (1μm diameter) functionalized with carboxylic groups can be adsorbed to almost any membrane proteins. The example cell images generated by image-guided cell sorter is shown in Figure. 4.8(a), and the histogram of beads counted calculated in real-time by image-guided cell sorter is shown in Figure. 4.8(b).

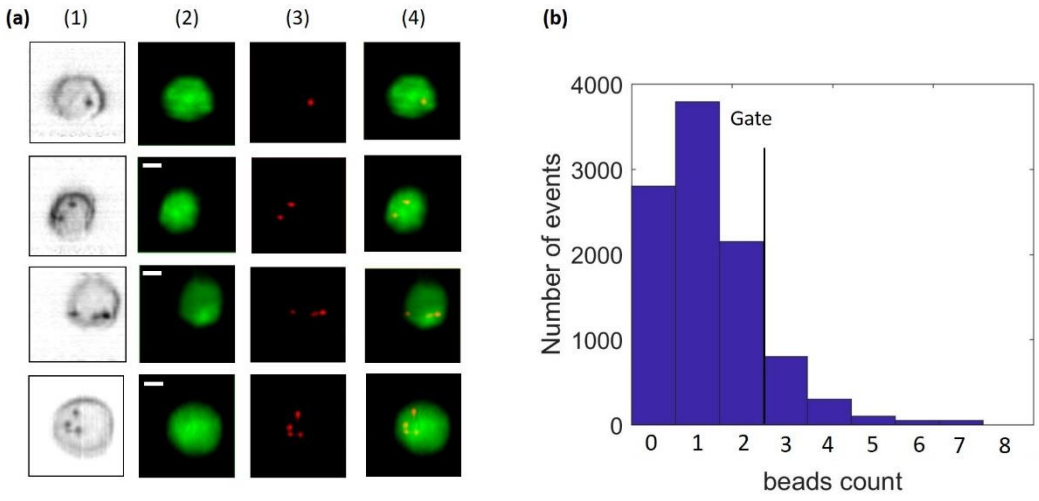


Figure 4.8 Image and histogram of HEK cells bonded with 1μm polystyrene beads. (a) Example images generated by image-guided cell sorter. First column is transmission image, the second column is green fluorescence image, the third column is red fluorescence images, the fourth column is overlay image. (b) Histogram of beads number counted by image-guided cell sorter.

In the sorting experiment, only cells bonded with no less than 3 beads are sorted and collected. The sorted cells are imaged by a fluorescent microscope. A 96% sorting purity is calculated based on 197 microscope images.

4.4.3 Sorting of pEGFP-GR plasmids translocated HEK-297T Human embryonic kidney cells

Spatial distribution of certain protein or organelles such as lysosomes and mitochondria carry important biological information. Image guided cell sorter is, to our knowledge, the only tool that can capture cells of sufficient quantity and purity based on such information. Here we demonstrate such functionality using pEGFP-GR plasmids translocated HEK-297T cells and un-translocated HEK-297T cells.

In the experiment, HEK-293T cells are transfected with GR-GFP and separated into 2 plates. One plate of cells is treated with dexamethasone that causes migration of GR-GFP protein from cytoplasm to nucleus. The other plate of cells is untreated so the GR-GFP protein stays in cytoplasm. The example images generated by image-guided cell sorter are shown in Figure. 4.9(a), and the histogram of area ratio is shown in Figure. 4.9(b). The mixture of both types of cells are flown through the system and imaged, and the interested subgroup can be isolated based on the real-time captured cell images.

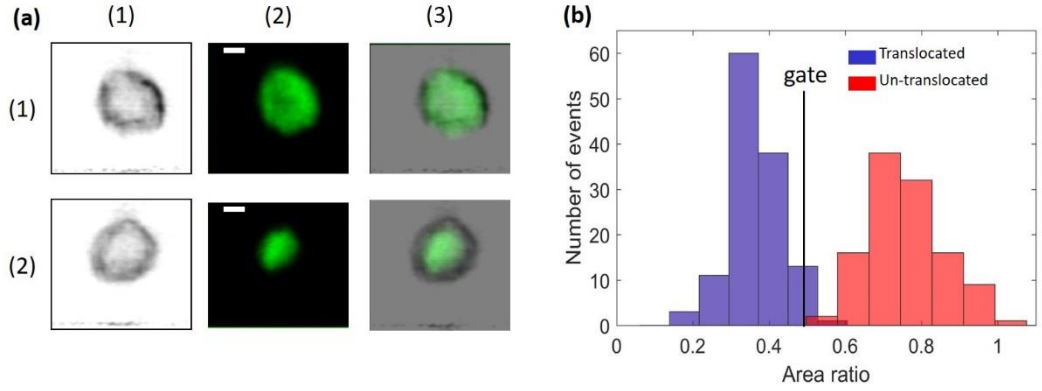


Figure 4.9 Example images and histogram of GFP-GR transfected HEK293T cells. (a) Example images generated by image-guided cell sorter. The first column is the transmission image, the second column is fluorescence image, the third column is overlay image. The first row is un-translocated cell, the second column is translocated cell. (b) Histogram of area ratio.

In the sorting experiment, GFP-GR transfected HEK293T cells treated and un-treated with drug (Dexamethasone) are mixed at ratio 50:50. Only translocated cells are sorted and collected. A fluorescent microscope is used to verify the sorting purity. Based on 130 microscope images, a 100% sorting purity is calculated.

Chapter 4 is based on and mostly a reprint of: **Yi Gu**, Xinyu Chen, Sung Hwan Cho, Chang-Hung Lee, Ivan Gagne, Alex Ce Zhang, Rui Tang, Yuanyuan Han, and Yu-Hwa Lo. "Image-Guided Cell Sorting Using Laser Point Scanning" in preparation. The dissertation author was the primary author of the work.

Conclusion

We demonstrate a image-guided cell sorting and classification system that enables: (1) generation of bright-field and fluorescent images of single cells in real time, (2) generation of image-derived gating criteria with machine learning, (3) generation and scoring of image-guided sorting parameters adaptive to user inputs, (4) cell sorting based on image-guided “gating”.

Importantly, all these additional features are established with essentially the same hardware as conventional FACS, allowing easy transition and upgrade of conventional systems to image-guided cell sorters. For the same token, the system can be easily expanded to include more parameters/colors by leveraging the multi-parameter FACS systems available today, and the throughput of the system can be enhanced with higher speed electronics.

Appendix A

Cell preparation protocols

Translocating pEGFP-GR plasmids to HEK-297T human embryonic kidney cells. First GR-GFP plasmid DNA is obtained from bacterial culture. Then two plates of HEK 293T cells are cultured and transfected with GR-GFP. After transfection, cells are cultured for 2~3 days. Then one plate of cells is treated with dexamethasone. The dexamethasone treatment is supposed to cause migration of **pEGFP-GR** protein from cytoplasm to nucleus. For untreated cells, **pEGFP-GR** protein stays in cytoplasm.

Bond fluorescent beads to MDCK cells. MDCK cells are cultured in a 10cm diameter dish. Then add 100μL solution of 1μm diameter fluorescent beads to the culture dish and keep the cell culture overnight. In the final step, cells are fixed and stained with Carboxyfluorescein succinimidyl ester (CFSE).

Irradiation and antibody staining of Human Glioblastoma Cells. To induce DNA double-strand breaks (DSB), GFP labeled Human Glioblastoma Cells (GBM-CCC-001) are treated with 6Gy irradiation. The treated cells are washed once with phosphate buffered saline (PBS) and fixed with 1% paraformaldehyde 30 minutes post irradiation. The fixed cells are washed with PBS twice. Then 70% ethanol is added to the cells and the cells are incubated on ice for 1 hour. After ethanol treatment,

cells are washed with PBS twice and incubated in 1% TritonX-100 at room temperature for 10 minutes. Then cells are washed with PBS once and incubated in 5% Bovine Serum Albumin (BSA) in PBS for 30 minutes at room temperature on shaker. Then cells are washed with PBS once and incubated in Anti-phospho-Histone H2A.X (Ser139) Antibody, clone JBW301 at 1:300 dilution on ice on shaker for 1 hour. After the primary antibody treatment, cells are washed twice with 5% BSA and incubate in PerCP/Cy5.5 anti-mouse IgG1 Antibody at 1:100 dilution on ice on shaker for 1 hour. At last, the stained cells are washed twice with 5% BSA and resuspended in 1:3 diluted stabilize fixative buffer in MilliQ water.

Bibliography

[1] Herzenberg, L. A., Parks, D., Sahaf, B., Perez, O., Roederer, M., & Herzenberg, L. A. (2002). The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford. *Clinical chemistry*, *48*(10), 1819-1827.

[2] Cho, S. H., Chen, C. H., Tsai, F. S., Godin, J. M., & Lo, Y. H. (2010). Human mammalian cell sorting using a highly integrated micro-fabricated fluorescence-activated cell sorter (μFACS). *Lab on a Chip*, *10*(12), 1567-1573.

[3] Feiguin, F., Ferreira, A., Kosik, K. S., & Caceres, A. (1994). Kinesin-mediated organelle translocation revealed by specific cellular manipulations. *The Journal of Cell Biology*, *127*(4), 1021-1039.

[4] Juan, G., Hernando, E., & Cordon-Cardo, C. (2002). Separation of live cells in different phases of the cell cycle for gene expression analysis. *Cytometry Part A*, *49*(4), 170-175.

[5] Sbarra, A. J., & Karnovsky, M. L. (1959). The biochemical basis of phagocytosis. *J biol chem*, *234*, 1355-1362.

[6] Bolte, S., & Cordelieres, F. P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *Journal of microscopy*, *224*(3), 213-232.

[7] Htun, H., Barsony, J., Renyi, I., Gould, D. L., & Hager, G. L. (1996). Visualization of glucocorticoid receptor translocation and intranuclear organization in living cells with a green fluorescent protein chimera. *Proceedings of the National Academy of Sciences*, *93*(10), 4845-4850.

[8] Li, H., Qian, W., Weng, X., Wu, Z., Li, H., Zhuang, Q., Feng, B., & Bian, Y. (2012). Glucocorticoid receptor and sequential P53 activation by dexamethasone mediates apoptosis and cell cycle arrest of osteoblastic MC3T3-E1 cells. *PLoS One*, *7*(6), e37030.

[9] Newton, A. C. (1995). Protein kinase C: structure, function, and regulation. *Journal of Biological Chemistry*, *270*(48), 28495-28498.

[10] Hawkins, E. D., Oliaro, J., Kallies, A., Belz, G. T., Filby, A., Hogan, T., ... & Seddon, B. (2013). Regulation of asymmetric cell division and polarity by Scribble is not required for humoral immunity. *Nature communications*, *4*, 1801.

[11] Barnett, B. E., Ciocca, M. L., Goenka, R., Barnett, L. G., Wu, J., Laufer, T. M., ... & Reiner, S. L. (2012). Asymmetric B cell division in the germinal center reaction. *Science*, *335*(6066), 342-344.

[12] Yoshida, H., Kawane, K., Koike, M., Mori, Y., Uchiyama, Y., & Nagata, S. (2005). Phosphatidylserine-dependent engulfment by macrophages of nuclei from erythroid precursor cells. *Nature*, *437*(7059), 754.

[13] Konstantinidis, D. G., Pushkaran, S., Johnson, J. F., Cancelas, J. A., Manganaris, S., Harris, C. E., ... & Kalfa, T. A. (2012). Signaling and cytoskeletal requirements in erythroblast enucleation. *Blood*, *119*(25), 6118-6127.

[14] Franzen, C. A., Simms, P. E., Van Huis, A. F., Foreman, K. E., Kuo, P. C., & Gupta, G. N. (2014). Characterization of uptake and internalization of exosomes by bladder cancer cells. *BioMed research international*, *2014*.

[15] Vallhov, H., Gutzeit, C., Johansson, S. M., Nagy, N., Paul, M., Li, Q., ... & Gabrielsson, S. (2011). Exosomes containing glycoprotein 350 released by EBV-transformed B cells selectively target B cells through CD21 and block EBV infection in vitro. *The Journal of Immunology*, *186*(1), 73-82.

[16] Nichols, L. A., Adang, L. A., & Kedes, D. H. (2011). Rapamycin blocks production of KSHV/HHV8: insights into the anti-tumor activity of an immunosuppressant drug. *PloS one*, *6*(1), e14535.

[17] Michel, M. L., Pang, D. J., Haque, S. F., Potocnik, A. J., Pennington, D. J., & Hayday, A. C. (2012). Interleukin 7 (IL-7) selectively promotes mouse and human IL-17–producing γδ cells. *Proceedings of the National Academy of Sciences*, *109*(43), 17549-17554.

[18] Imai, T., Kato, Y., Kajiwara, C., Mizukami, S., Ishige, I., Ichiyanagi, T., ... & Udono, H. (2011). Heat shock protein 90 (HSP90) contributes to cytosolic translocation of extracellular antigen for cross-presentation by dendritic cells. *Proceedings of the National Academy of Sciences*, *108*(39), 16363-16368.

[19] Di Carlo, D., Irimia, D., Tompkins, R. G., & Toner, M. (2007). Continuous inertial focusing, ordering, and separation of particles in microchannels. *Proceedings of the National Academy of Sciences*, *104*(48), 18892-18897.

[20] Di Carlo, D., Edd, J. F., Irimia, D., Tompkins, R. G., & Toner, M. (2008). Equilibrium separation and filtration of particles using differential inertial focusing. *Analytical chemistry*, *80*(6), 2204-2211.

[21] Oakey, J., Applegate Jr, R. W., Arellano, E., Carlo, D. D., Graves, S. W., & Toner, M. (2010). Particle focusing in staged inertial microfluidic devices for flow cytometry. *Analytical chemistry*, *82*(9), 3862-3867.

[22] Godin, J., Lien, V., & Lo, Y. H. (2006). Demonstration of two-dimensional fluidic lens for integration into microfluidic flow cytometers. *Applied Physics Letters*, *89*(6), 061106.

[23] Wu, T. F., Mei, Z., & Lo, Y. H. (2012). Optofluidic device for label-free cell classification from whole blood. *Lab on a Chip*, *12*(19), 3791-3797.

[24] Wu, T. F., Mei, Z., & Lo, Y. H. (2013). Label-free optofluidic cell classifier utilizing support vector machines. *Sensors and Actuators B: Chemical*, *186*, 327-332.

[25] Wu, T. F., Mei, Z., Pion-Tonachini, L., Zhao, C., Qiao, W., Arianpour, A., & Lo, Y. H. (2011). An optical-coding method to measure particle distribution in microfluidic devices. *AIP advances*, *1*(2), 022155.

[26] Deinard, A. S., Fortuny, I. E., Theologides, A., Anderson, G. L., Boen, J., & Kennedy, B. J. (1974). Studies on the neutropenia of cancer chemotherapy. *Cancer*, *33*(5), 1210-1218.

[27] Crawford, J., Ozer, H., Stoller, R., Johnson, D., Lyman, G., Tabbara, I., ... & Smith, R. (1991). Reduction by granulocyte colony-stimulating factor of fever and neutropenia induced by chemotherapy in patients with small-cell lung cancer. *New England Journal of Medicine*, *325*(3), 164-170.

[28] Han, Y., & Lo, Y. H. (2015). Imaging cells in flow cytometer using spatial-temporal transformation. *Scientific reports*, *5*, 13267.

[29] Han, Y., Gu, Y., Zhang, A. C., & Lo, Y. H. (2016). imaging technologies for flow cytometry. *Lab on a Chip*, *16*(24), 4639-4647.

[30] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*, 3-24.

[31] Chen, C. L., Mahjoubfar, A., Tai, L. C., Blaby, I. K., Huang, A., Niazi, K. R., & Jalali, B. (2016). Deep learning in label-free cell classification. *Scientific reports*, *6*, 21471.

[32] Blasi, T., Hennig, H., Summers, H. D., Theis, F. J., Cerveira, J., Patterson, J. O., ... & Rees, P. (2016). Label-free cell cycle analysis for high-throughput imaging flow cytometry. *Nature communications*, *7*, 10256.

[33] McLane, L. M., & Corbett, A. H. (2009). Nuclear localization signals and human disease. *IUBMB life*, *61*(7), 697-706.

[34] Usmani, O. S., Ito, K., Maneechotesuwan, K., Ito, M., Johnson, M., Barnes, P. J., & Adcock, I. M. (2005). Glucocorticoid receptor nuclear translocation in airway cells after inhaled combination therapy. *American journal of respiratory and critical care medicine*, *172*(6), 704-712.

[35] Vandevyver, S., Dejager, L., & Libert, C. (2012). On the trail of the glucocorticoid receptor: into the nucleus and back. *Traffic*, *13*(3), 364-374.

[36] Kuo, L. J., & Yang, L. X. (2008). γ-H2AX-a novel biomarker for DNA double-strand breaks. *In vivo*, *22*(3), 305-309.

[37] Sharma, A., Singh, K., & Almasan, A. (2012). Histone H2AX phosphorylation: a marker for DNA damage. In *DNA repair protocols* (pp. 613-626). Humana Press, Totowa, NJ.

[38] Lapytsko, A., Kollarovic, G., Ivanova, L., Studencka, M., & Schaber, J. (2015). FoCo: a simple and robust quantification algorithm of nuclear foci. *BMC bioinformatics*, *16*(1), 392.

[39] Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on information theory*, *52*(4), 1289-1306.

[40] Coskun, A. F., Sencan, I., Su, T. W., & Ozcan, A. (2010). Lensless wide-field fluorescent imaging on a chip using compressive decoding of sparse objects. *Optics express*, *18*(10), 10510-10523.

[41] Kim, S. J., Koh, K., Lustig, M., Boyd, S., & Gorinevsky, D. (2007). An Interior-Point Method for Large-Scale l1-Regularized Least Squares. *IEEE journal of selected topics in signal processing*, *1*(4), 606-617.

[42] Chou, H. P., Spence, C., Fu, A., Scherer, A., & Quake, S. (1998, June). Disposable microdevices for DNA analysis and cell sorting. In *Proc. Solid-State Sensor and Actuator Workshop, Hilton Head, South Carolina* (pp. 11-14).

[43] Fiedler, S., Shirley, S. G., Schnelle, T., & Fuhr, G. (1998). Dielectrophoretic sorting of particles and cells in a microsystem. *Analytical chemistry*, *70*(9), 1909-1915.

[44] Schmid, L., Weitz, D. A., & Franke, T. (2014). Sorting drops and cells with acoustics: acoustic microfluidic fluorescence-activated cell sorter. *Lab on a Chip*, *14*(19), 3710-3718.

[45] Wang, X., Chen, S., Kong, M., Wang, Z., Costa, K. D., Li, R. A., & Sun, D. (2011). Enhanced cell sorting and manipulation with combined optical tweezer and microfluidic chip technologies. *Lab on a Chip*, *11*(21), 3656-3662.

[46] Eulenberg, P., Köhler, N., Blasi, T., Filby, A., Carpenter, A. E., Rees, P., ... & Wolf, F. A. (2017). Reconstructing cell cycle and disease progression using deep learning. *Nature communications*, *8*(1), 463.