Three-dimensional Morphometric Biosignatures Of Cancer

By Automated Analysis Of Transmission-mode Optical Cell CT Images

by

Vivek Nandakumar

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2013 by the
Graduate Supervisory Committee:

Deirdre R. Meldrum, Chair
Alan C. Nelson
Lina J. Karam
Jieping Ye
Roger H. Johnson
Kimberly J. Bussey

ARIZONA STATE UNIVERSITY

August 2013

ABSTRACT

Despite significant advances in digital pathology and automation sciences, current diagnostic practice for cancer detection primarily relies on a qualitative manual inspection of tissue architecture and cell and nuclear morphology in stained biopsies using low-magnification, two-dimensional (2D) brightfield microscopy. The efficacy of this process is limited by inter-operator variations in sample preparation and imaging, and by inter-observer variability in assessment. Over the past few decades, the predictive value quantitative morphology measurements derived from computerized analysis of micrographs has been compromised by the inability of 2D microscopy to capture information in the third dimension, and by the anisotropic spatial resolution inherent to conventional microscopy techniques that generate volumetric images by stacking 2D optical sections to approximate 3D.

To gain insight into the analytical 3D nature of cells, this investigation explores the application of a new technology for single-cell optical computed tomography (optical cell CT) that is a promising 3D tomographic imaging technique which uses visible light absorption to image stained cells individually with sub-micron, isotropic spatial resolution. This dissertation provides a scalable analytical framework to perform fully-automated 3D morphological analysis from transmission-mode optical cell CT images of hematoxylin-stained cells. The developed framework performs rapid and accurate quantification of 3D cell and nuclear morphology, facilitates assessment of morphological heterogeneity, and generates shape- and texture-based biosignatures predictive of the cell state. Custom 3D image segmentation methods were developed to precisely delineate volumes of interest (VOIs) from reconstructed cell images. Comparison with user-defined ground truth assessments yielded an average agreement (DICE coefficient) of 94% for the cell and its nucleus. Seventy nine biologically relevant morphological descriptors (features) were computed from the segmented VOIs, and statistical classification methods were implemented to determine the subset of features that best predicted cell health.

The efficacy of our proposed framework was demonstrated on an *in vitro* model of multistep carcinogenesis in human Barrett's esophagus (BE) and classifier performance

using our 3D morphometric analysis was compared against computerized analysis of 2D image slices that reflected conventional cytological observation. Our results enable sensitive and specific nuclear grade classification for early cancer diagnosis and underline the value of the approach as an objective adjunctive tool to better understand morphological changes associated with malignant transformation.

DEDICATION

To all those who contributed

ACKNOWLEDGEMENTS

My doctoral education has been a very eventful experience that has greatly molded my mindset from that of a conventional engineering graduate to one with a multidisciplinary skill set. I sincerely thank my adviser and committee chair - Prof. Deirdre Meldrum, for giving me the opportunity to work on a cutting-edge multidisciplinary project in a top-tier research laboratory setting at the Center for Biosignatures Discovery Automation (CBDA), and for her unwavering support throughout the course of my research. I am also grateful to Prof. Alan Nelson, Prof. Lina Karam and Prof. Jieping Ye for serving on my committee and providing valuable advice in their respective areas of subject matter expertise. I greatly appreciate committee members Dr. Roger Johnson and Dr. Kimberly Bussey for their invaluable day-to-day mentorship, for stoking my interest in translational health science research, and guiding my efforts on high-resolution micro-imaging and cancer biology.

Teamwork is an integral component of successful multidisciplinary research, and I experienced this first-hand during my predoctoral stint. My research would not have come to fruition if not for the efforts of several of my lab colleagues. First, a big thank you to team members in my research project: Kathryn Hernandez, Nanna Hansen, Miranda Slaydon, Stephanie Helland, Jessica Han, and Atma Thompson for their untiring efforts at the bench. I am also very grateful to staff colleagues Dr. Laimonas Kelbauskas, Dr. Brian Ashcroft, Jeff Houkal, Dr. Honor Glenn, Dr. Thai Tran, and Juan Vela for their guidance and timely assistance with different aspects of my research. Cheers to the rest of the staff and student colleagues in CBDA.

I consider myself very fortunate to have been able to pursue my research under the auspices of the unique NCI Physical Sciences in Oncology (PSOC) consortium. The network has been instrumental in shaping and enhancing my multidisciplinary skill set, perspectives, and scientific vision. I am deeply grateful to Prof. Paul Davies - the principal investigator at ASU, senior investigators in the network, and the management at the NCI. Finally, my deepest gratitude to my family for their patience, perseverance, sacrifice, and support during my years away from home. I will forever remain indebted to them.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Motivation

Cancer is one of the leading causes of death worldwide. A minimum of 1.6 million new cases and half a million deaths were estimated for 2012 in the United States alone [1]. The disease also poses a major economic burden due to medical expenses or lost productivity. The American Cancer Society estimated an expenditure of $200 billion in 2012 [2]. The high mortality rate associated with cancer is attributed to its complexity. In contrast to most other diseases, cancer cannot be attributed to any single infectious microbial agent or enzyme disorder. Instead, it is a heterogeneous conglomerate of consequences that are triggered by myriad host and environmental factors operating over a multitude of length and time scales.

Extensive research has been undertaken over the last several decades to unravel the origins of cancer and characterize its pathophysiologies in an effort to develop more efficacious treatment strategies. The widely accepted theme that has emerged from this vast body of scientific literature describes cancer as the result of a multistep and multiscale breakdown in biological regulatory mechanisms that damage the genome and eventually effect a permanent alteration in normal cellular function. Hanahan and Weinberg have excellently consolidated the vast majority of research accomplishments in the field in the form a conceptual framework on cancer biology that describes the major functional attributes of the disease [3, 4].

Recent breakthroughs in high throughput sequencing technologies have greatly increased domain knowledge about the molecular mechanisms of cellular function and their aberrations in cancer [5, 6]. These new discoveries have spurred the development of numerous targeted molecular therapies that repair aberrant cellular function by modulating specific molecular drivers [7, 8]. Such therapies have been shown to provide a promising alternative to traditional surgical resection and chemotherapeutic treatment of tumors [9]. However, the unavailability of complete information about the spatial and temporal

dynamics of signaling mechanisms, their interactions with regulatory networks, the existence of crosstalk between molecular pathways, inherent systemic redundancies in function, and genotypic heterogeneity limit the efficacy of molecular diagnostics [10, 11]. Research endeavors are being undertaken by way of rational combination therapies [12, 13] and pharmacokinetic modeling to unravel problems such as acquired drug resistance [14], but several other challenges remain to be addressed before the merits of personalized oncology can be fully realized in the clinic [15–17]. Consequently, definitive cancer diagnosis in the clinic remains largely centered around recognizing the morphological manifestations of the disease, often referred to as malignancy associated changes (MACs). Malignant transformations in cells or tissue are associated with chaotic tissue architecture, increased mitotic activity in cells, and abnormal morphology of cells, their nuclei, and DNA packing within the nuclei [18–20]. These factors, along with ectopic occurrence of malignant cells, constitute a major component of universally adopted grading and staging schemes used to diagnose cancer or its precursors in various tissue and organ systems [21, 22].

The conventional procedure followed in clinics to diagnose cancer is a qualitative, manual grading of morphological abnormalities in whole cell smears (cytology) or 5-micron thick sections of tissue (histology) stained with absorption dyes hematoxylin and eosin (H&E) and mounted on glass slides. Trained pathologists visually assess the specimen using two-dimensional (2D) brightfield microscopy at magnifications of 10x, 20x, or 40x. Common pathology practice is to first scan the sample at 10x and then proceed to observations of intriguing cells or regions of tissue at the higher magnifications. The high depth of field of a 10x objective lens (approximately 8.5 microns) is leveraged to infer morphological information along the depth of the specimen. Certified pathologists develop their domain knowledge about morphological aberrations by observing 2D staining patterns in several thousands of H&E stained cytological or histological samples over many years. Despite being the standard of care, current pathological practice poses several technical problems for precise diagnosis.

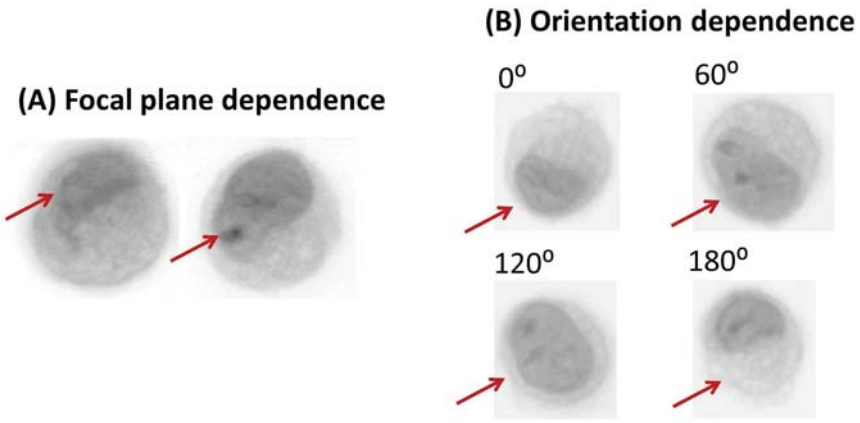

Figure 1.1: 2D brightfield microscopy assessment of cells poses several problems (indicated by red arrows in each sub-figure) that could skew an observer's judgment. (A) Different focal planes from the same cell present different morphologies. (B) Different angular orientations of the same cell present different morphologies. Cell is stained with hematoxylin dye.

As shown in Figure 1.1, factors such as focal plane selection and sample orientation on the slide may bias the outcome of the diagnosis due to obscuration or incomplete feature detail. The arrows in Figure 1.1(A) highlight the changes in morphological detail at two different focal planes within the same cell. Similarly, Figure 1.1(B) illustrates morphological variations at different viewing angles of the same cell. Intra- and inter-personnel variations in sample preparation and judgment may further compound the risk of biased outcomes. Sample degradation with archival or slide breakage during transfer are other limiting concerns.

The need for precise morphological assessment in cancer diagnosis has triggered the development of innovative 2D and 3D biomedical imaging technologies with spatial resolutions that span a multitude of length scales ranging from millimeters to nanometers [23–25]. Technological breakthroughs in molecular labeling enable the study of various functional (such as metabolic, genomic, transcriptomic, and proteomic) profiles, their heterogeneity, and temporal dynamics in biological specimens at these various length scales. Some volumetric imaging technologies such as x-ray computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and ultrasound

(US) now constitute a part of standard of care practices for screening, detection and management of cancer [26]. Advances in automation sciences, image digitization, visualization, archival, and retrieval have greatly enhanced the utility of biomedical imaging methods by facilitating a comprehensive quantitative phenotypic characterization of the underlying specimen [27]. Software commonly referred to as CAD (computer-aided detection) tools have leveraged the advances in computer vision and data mining techniques to perform rapid and objective detection and a quantitative annotation of target object morphologies in a semi-automated or fully-automated manner.

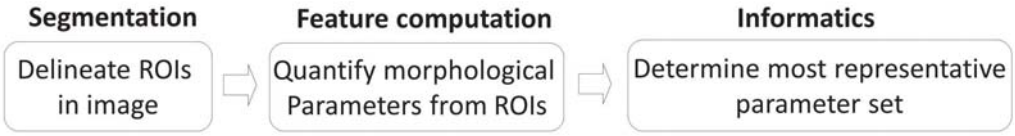

Figure 1.2: Sequence of methods adopted by CAD tools.

Figure 1.2 illustrates the typical methodological sequence followed by CAD tools. First, computer vision algorithms are applied to delineate (segment) regions of interest (ROIs) in acquired digital images. Diagnostically relevant morphological parameters (features) are then measured in the delineated ROIs. The procedure is repeated for all digital images to be assessed. The resulting quantitative feature sets are subject to data mining techniques to derive biological signatures composed of the subset of features that best identify or stratify the specimen according to its morphology.

The development of robust CAD tools for a variety of imaging modalities has been a subject of extensive research over the past two decades. Diagnostic radiology practice routinely uses CAD tools optimize detection and treatment strategies for suspect or malignant tumors [28]. The advent of digital microscopy has resulted in the development of a plethora of CAD tools that detect and grade cancer by analyzing digitized micrographs of H&E-stained cell or tissue samples. Methods have been proposed for numerous cancer types including blood [29], skin [30], brain [31], oral [32], thyroid [33], esophagus [34],

lung [35], breast [36], liver [37], kidney [38], colon [39], ovary [40], prostate [41], and cervix [42]. Digital pathology is an integrated technological thrust that leverages the advances in automation and CAD tool development in an effort to assist pathologists in their morphological assessment of cytological and histological samples [43]. The technology provides pathologists with digitized 2D whole slide images of H&E stained samples along with CAD solutions for quantitative morphometry and objective grading. Some platforms also provide 3D visualizations of the samples at multiple magnifications by taking a select number of 2D focal planes through the specimen volume.

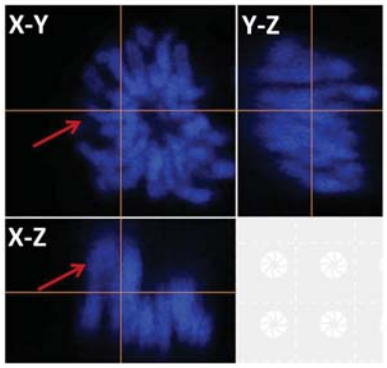

Figure 1.3: Serial-section reconstruction based 3D microscopy possess anisotropic spatial resolution. Red arrows in the orthogonal cross-sections of a 3D image stack of a mitotic cell fluorescently labeled with DAPI dye illustrate the decrease in resolution along the X-Z and Y-Z image planes relative to the X-Y plane.

Digital pathology thus provides several tangible benefits relative to conventional practice, including increased productivity, quantitative assessment, digital archival and ease of data transfer for telepathology. However, several hurdles remain to be addressed with regard to protocol standardization and federal regulation before digital pathology can be incorporated into primary pathology practice [44]. In addition to regulatory policy considerations, CAD schemes that use 2D micrographs are still limited in their efficacy by technical constraints such choice of focal plane, sample orientation, overlapping cells, and the lack actual information about the third dimension. To overcome the limitations associated with 2D microscopy, CAD analysis has been undertaken using 3D microscopy

images acquired by popular imaging techniques such as confocal microscopy [45]. Conventional volumetric microscopy techniques generate 3D images by serially stacking 2D focal plane images acquired along the vertical axis of the specimen. Volumetric images generated by such serial-section acquisition suffer from anisotropic spatial resolution that results in a 2-4 fold decrease in spatial resolution along the vertical (z-)axis relative to the lateral (x-, y-) axes (see Figure 1.3). Interpolation schemes commonly used to generate isotropic voxels in images can only approximate the actual morphological information in the specimen and may thus decrease the sensitivity of derived morphometric signatures.

1.2 Approach

This research is motivated by the intuitive fact that accurate quantitative characterization of cell and nuclear morphology by 3D analyses of high contrast, high resolution 3D imagery with isotropic resolution would facilitate a better assessment of morphological changes associated with malignancy.

Optical microscopy techniques capable of rapid volumetric imaging with isotropic spatial resolution have just begun to be realized in prototype form [46]. Micro-imaging methods that apply tomographic reconstruction from projections overcome the spatial anisotropy limitation by virtue of their imaging physics and thus provide volumetric images of whole, intact cells with isotropic spatial resolution. Optical projection tomography (OPT) [47] was among the earliest of such methods and enables 3D imaging of multicellular specimen with an isotropic spatial resolution of few microns. Soft x-ray tomography (SXT) [48] produces 3D images of unstained individual cells with 60 nm isotropic resolution. However, it requires very expensive infrastructure with a large footprint for a low to moderate throughput, and is thus not affordable by most research labs.

Optical CT is a cellular imaging technique that uses computed tomography principles and white light excitation to generate 3D cell images with an isotropic resolution of 350 nm [49, 50]. Each image is generated by tomographic reconstruction from 500, equi-angular pseudoprojection images acquired over a 360 degree rotation of a stained cell suspended in an index-matched optical gel within a glass capillary.

A pseudoprojection image is generated by integrating widefield focal plane information over the cell volume using a 100x, 1.3 NA, oil immersion objective lens. Acquired pseudoprojection images are denoised, registered and subject to reconstruction algorithms to generate the volumetric cell image. Cells are imaged either in transmission-mode or epi fluorescence-mode depending on the dye used for staining.

The Cell-CT^{TM} (VisionGate Inc., Phoenix, AZ) is a benchtop instrument that applies optical CT to rapidly image cells in true 3D and thus provides a low-cost, high-throughput alternative to SXT for 3D single-cell imaging. The Cell-CT^{TM} is utilized to address the 3D cell imaging requirement for this dissertation research. Precise, automated 3D analysis of transmission-mode optical cell CT images generated using the Cell-CT^{TM} and the subsequent application of 3D morphological analysis for early cancer detection is the central theme of the dissertation.

1.3 Dissertation objectives

The overarching goal of this research is to develop robust 3D morphological biosignatures indicative of neoplastic progression in epithelial cells by analysis of volumetric cell images generated by transmission-mode optical cell CT. This dissertation aims to develop an efficient, automated data-processing framework that can (a) precisely quantify three-dimensional cell and nuclear morphology from Cell-CT ^{TM} images of hematoxylin-stained cells, and (b) compute a morphological biosignature composed of the set of morphological parameters that can best distinguish two or more classes of cells with differing health state.

1.4 Summary of contributions

The following contributions are made in this dissertation:

1. A modular, automated computational framework is developed to perform high-throughput, 3D morphological analysis of volumetric images of transmission-mode Cell-CT^{TM} images.

2. Custom 3D image processing methods are provided to (a) accurately delineate volumes of interest (VOIs) in the volumetric images, and (b) compute biologically relevant morphological and texture parameters from the segmented VOIs.

3. The efficacy of proposed segmentation techniques is validated relative to manually traced ground truth volumes.

4. The proposed image processing sequence is applied in conjunction with data mining techniques to investigate the existence and power of a 3D morphometric biosignature to distinguish metaplastic, dysplastic, and adenocarcinoma esophageal epithelial cells from normal squamous epithelial cells.

5. The relative merit of 3D morphological analysis over 2D analysis is investigated.

1.5 Dissertation organization

This chapter provided an overview of relevant theoretical background information germane to the scope of this dissertation. The rest of this dissertation is organized as follows: Chapter 2 provides an overview of three-dimensional optical CT imaging of individual hematoxylin-stained cells using the Cell-CT^{TM} instrument. Chapter 3 details the challenges posed for precise segmentation of transmission-mode optical cell images and describes fully-automated methods to segment the cell, nucleus, and ultra-high density (UHD) clumps in the nucleus from the image data. The chapter also details image enhancement methods to be applied prior to segmentation methods. Chapter 4 details the datasets and procedures to generate ground truth volumes to assess the efficacy of proposed

segmentation methods. It concludes with a performance analysis of the proposed segmentation methods relative to user-derived ground truth. Chapter 5 details the set of 3D morphological and texture features computed from the cell images. It also summarizes the suite of methods implemented to facilitate the generation of quantitative biosignatures. Chapter 6 describes the application of the proposed computational framework in conjunction with statistical data mining methods to generate 3D morphological biosignatures that can distinguish the four stages of esophageal adenocarcinoma progression with high sensitivity and specificity. The chapter includes an analysis of the relative efficacy of 3D morphological analysis over 2D analysis on images representative of acquisition with a 10x objective lens. Chapter 7 summarizes this dissertation research and provides an outlook on potential future research directions.

Chapter 2

3D SINGLE-CELL IMAGING BY OPTICAL CT

This chapter provides an overview of the cellular-scale optical computed tomography technique used to image single cells in true 3D. It also details the setup, sample preparation, and image acquisition procedures for the Cell-CT^{TM} instrument that applies optical CT to tomographically image stained individual cells in brightfield mode.

2.1 Optical CT

Optical CT is a cellular-scale imaging technique that incorporates computed tomography principles into widefield optical microscopy to generate three-dimensional single-cell images with sub-micron, isotropic spatial resolution [49–51]. The 3D cell image is generated by mathematical reconstruction from five hundred equiangular, 2D optical projection images of a stationary, stained cell suspended in an optical gel within a glass capillary. Optical cell CT has been successfully demonstrated with multiple widefield microscopy modalities such as brightfield, darkfield, phase-contrast, and epi fluorescence [52].

2.2 The Cell-CT^{TM} instrument

Setup

Brightfield or transmission-mode optical cell CT technology has been commercialized by VisionGate, Inc. (Phoenix, AZ) as a diagnostic tool to perform three-dimensional cytopathology. The instrument – Cell-CT^{TM}, consists of an inverted optical microscope (IX71, Olympus, San Diego, CA) fitted with a 100x, 1.3 NA, oil immersion objective lens (UPlanFluor, Olympus, CA), a 1400$\times$1100 pixel, monochrome CCD camera (Prosilica1650, Burnaby, BC), and a custom mechanical assembly that incorporates a motorized stage, gadgetry for precision injection and rotation control, and a specialized, portable cartridge which houses the tomographic imaging chamber. A 100 μl glass syringe (Hamilton, Reno, NV) containing stained cells embedded in a special carrier gel (SmartGel, Nye Lubricants, Fairhaven, MA) is mechanically coupled to the injection control system at one end and the cartridge at its other end. The cartridge is mounted on the stage with its

imaging chamber positioned between the microscope's condenser and piezo-actuated objective lens. LabView software (National Instruments, Austin, TX) is used to automate the image acquisition process.

Cartridge design

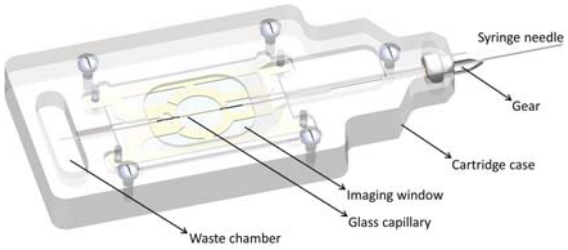

Figure 2.1: Schematic for the Cell-CT^{TM} cartridge

The imaging cartridge (see Figure 2.1) houses a glass microcapillary (inner diameter of 50 μm, outer diameter 150 μm) to transport cells, a proximal syringe needle to facilitate coupling with the glass syringe, a gear to mechanically rotate the capillary, and a distal waste reservoir to collect the extruded cell sample after imaging. A teflon gasket is used to provide an airtight coupling between the cartridge and the syringe at the needle end. The imaging chamber is in the middle of the cartridge and is formed by an immersion oil or optical gel-filled glass coverslip assembly that sandwiches the capillary. The capillary is superglued in the syringe needle orifice and the needle is rotated by the gear. The capillary glass, the coverslip sandwich, the gel for cell transport, and the immersion oil for the objective lens are all refractive index matched at about 1.45 to minimize spatial distortions during projection image acquisition.

Cell sample preparation

To accomplish a scenario analogous to x-ray CT where the 3D image captures variations in object (electron) density, biological cells need to be stained with an absorption dye that imparts optical contrast proportional to the density of the biological material. Hematoxylin is an absorption dye used in clinical practice that possess this capability. It predominantly stains nuclear DNA by virtue of its affinity for acidic moieties. Standard cytological

protocols are adopted for staining. Suspended cells are fixed for one hour with CytoLyt (Cytyc, Malborough, MA) and then smeared onto a clean microscope glass slide (VWR, West Chester, PA) coated with 0.01% Poly-L-Lysine (PLL) solution (Sigma Aldrich, St. Louis, MO). Prior to coating with PLL, the slide is washed with 2% Neutrad liquid detergent (Fisher Scientific, Fair Lawn, NJ), then rinsed with de-ionized water. Hematoxylin stain solution is freshly prepared with filtered tap water. Cells are stained for a few minutes (cell type dependent) in aqueous 6.25% w/w Gill's hematoxylin (Electron Microscopy Sciences, Hatfield, PA) solution, followed by bluing reagent (Fisher Scientific, Fair Lawn, NJ) for 30 seconds after washing thrice with filtered tap water. The cells are subsequently dehydrated through an ethanol series (50%, 95%, and 100%) and two washes with 100% xylene. Lastly, the stained cells are embedded in the carrier gel which is scraped off the glass slide surface and loaded into a 100μl glass syringe (Hamilton, Reno, NV).

Hematoxylin staining results in a bluish nucleus and a lighter cytoplasm. Optimal staining is key to precise tomographic imaging of cells. Since proper staining is highly dependent on experimental conditions (including pH of the water used), multiple trials are usually needed to optimize the concentrations of reagents and durations of protocol steps.

Projection image acquisition

Stained cells embedded in the carrier gel are transported through the rotating glass capillary by forward actuation of the syringe plunger that pressurizes the gel and causes it to flow. When a cell flows into the field of view of the microscope, pressure is released off the gel which causes it to stop flowing and solidify. As a result, the cell becomes stationary in the field of view. Operator-based, manual cell selection is performed to ensure that only well-stained, intact interphase cells are imaged. For every selected cell, 500 projection images are acquired at angular intervals of 0.72° around the cell (see Figure 2.2) by rotating the capillary at a constant speed with the gear assembly that is coupled to the cartridge. At each angular increment, an optical projection image is generated by sweeping the objective lens through the cell volume and integrating the resultant infinite focal plane information on the camera chip [49]. The carrier gel that transports cells in the capillary

also supports them to avoid tumbling or drift as the capillary is rotated during image acquisition. To account for the precession of the capillary due to mechanical wear and tear, a registration step is performed track the cell's trajectory prior to projection acquisition to ensure the objective lens correctly sweeps the cell volume at every angular increment. After projection acquisition, a spatially registered stack of well-focused images spanning the entire cell is acquired for subsequent 2D cytological assessment. The imaged cell is finally transported to the distal waste reservoir by advancing the syringe plunger, and the next cell is moved into the field of view.

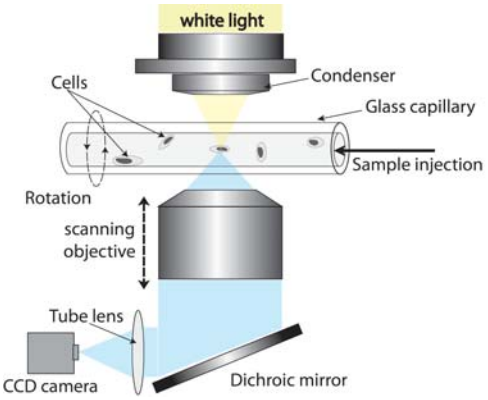

Figure 2.2: Optical path for transmission-mode projection image acquisition in the Cell-CT^{TM}

3D cell image generation by tomographic reconstruction

The acquired projection image data aligned and subject to mathematical reconstruction algorithms to generate the 3D image of each cell. Prior to alignment, a background subtraction routine is performed to eliminate pattern noise emanating from inconsistent illumination intensity, optical path discrepancies, and imperfections in the detector (CCD camera). Projection images are aligned based on center of mass and the aligned projections are subject to filtered back projection reconstruction using a custom ramp filter to obtain the volumetric cell image having an isotropic spatial resolution of 350 nm. Reconstructed volumes are stored as 2D image stacks at bitdepths of 8 and 16 bits. Intensities in the image inversely correlate with hematoxylin stain density in the cell.

Chapter 3

ALGORITHMS FOR 3D IMAGE SEGMENTATION

Image segmentation refers to process of accurately delineating regions of interest (ROIs) from images. Although manual segmentation is considered as the gold standard to identify an object, computerized methods are becoming increasingly popular accessories to enable faster and repeatable segmentation of ROIs. The goal of this dissertation is to develop rapid, fully-automated algorithms to accurately segment volumes of interest (VOIs) from optical cell CT images. Target VOIs are the cell (VOI 1), nucleus (VOI 2) and ultra-high density (UHD) clumps inside the nucleus (VOI 3). We define UHD clumps as regions with high intensity. These include nucleoli and dense agglomerations of chromatin, since the hematoxylin dye stains both of them. Optimal segmentation of these VOIs is a crucial precursor for accurate morphological analysis.

A common useful practice prior to segmentation is to pre-process the images to either reduce noise or enhance the target object. The terms “denoising” or “enhancement” are used, often interchangeably, to refer to the process of reinforcing voxel intensities of the target object relative to the background. Numerous techniques exist to enhance images, and the efficacy of a technique depends on the application at hand.

3.1 Related prior work

Algorithm development for segmentation of biological images remains a subject of active and extensive research since the last twenty five years. A vast number of interactive and fully-automated segmentation techniques have been proposed for numerous imaging modalities. Segmentation algorithms are broadly divided into either intensity-based, edge-based, or region-based methods. For purposes of brevity, this section will focus on segmentation methods developed for delineation of cellular ROIs in volumetric microscopy images.

Numerous semi- and fully automated segmentation schemes have been developed for a multitude of microscopy modalities. To enable optimal performance, methods for cell

and nuclear segmentation from 3D microscopy images are usually customized to the imaging mode (transmission or fluorescence) and image characteristics such as signal-to-noise ratio (SNR), number of volumes of interest (VOIs) in the imaged volume, and their proximity to one another. Li et al. automatically segmented fluorescently labeled cell nuclei from 3D confocal image stacks by adaptively thresholding similar intensity regions of the image after gradient vector flow diffusion and clustering [53]. Zanella et al. proposed a method based on subjective surfaces guided by predefined parameters to automatically segment fluorescently labeled cell and nuclear membranes from 3D time-lapse confocal images of zebrafish embryos [54]. Chinta et al. automatically segmented cell nuclei from 3D confocal images of live drosophila cells by evolving multiple narrow band level set surfaces [55]. Initial masks for nuclei were computed by a combination of K-means classification and a weighted 2-class expectation-maximization (EM) algorithm. Nandy et al. applied an automated multistage watershed segmentation followed by supervised artificial neural networks to accurately identify fluorescently labeled nuclei from 3D widefield images of cancerous breast tissue sections [56]. Electron tomography images pose a relatively greater challenge for segmentation due to their inherently low signal to noise ratios (SNRs). Bartesaghi et al. segmented HIV particles from electron tomographic images of human macrophages using a semi-automated method that essentially determined optimal minimum energy surfaces around manually selected seed points in the image volume [57]. Nguyen and Ji automatically segmented biological membranes from electron tomographic images using a shape-driven 3D watersnake method [58]. Moussavi et al. proposed an automated slice-wise recursive algorithm based on a global shape model and shape-dependent features to recover cell volumes from cryo-electron tomographic images [59]. Rigort et al. performed automated template matching with a predefined cylindrical element followed by a custom tracing algorithm on the resulting template field to segment actin filaments from cryo-electron tomographic images of eukaryotic cells [60].

3.2 Challenges for segmentation of transmission-mode optical cell CT images

Little is known about optimal segmentation strategies for single-cell optical CT images. Segmentation methods used for soft x-ray tomography (SXT) images are not extendable since they determine threshold intensity values for an VOI based on the extent of energy absorption by the molecules in that VOI [61]. Transmission-mode optical cell CT images pose the following challenges that complicate the segmentation procedure:

- Hematoxylin stain biochemistry and the cell staining procedure:

  Absorption stains serve as the source of signal for image reconstruction in transmission-mode optical cell CT. Assuming faithful reconstruction, voxel intensities in the cell directly depend on the stain density. We use hematoxylin dye to impart stain contrast to cells, primarily motivated by its widespread use in diagnostic pathology. Hematoxylin binds to acidic moieties in the cell. Consequently, it preferentially stains the cell nucleus relative to the rest of the cell due to the abundance of acidic components in DNA. However, hematoxylin is a non-stoichiometric stain, which implies that stain density may not correlate with the amount of its binding target. Artifactual (over or under) staining of cells due to operator inconsistency would result in an incorrect representation of optical densities in the reconstructed image which in turn could bias the segmentation. The non-specific binding tendency of hematoxylin may contribute to artifacts similar to partial volume effects in conventional CT or MRI imaging.

- Factors related to imaging system:

  Optical CT imaging is a relatively nascent imaging modality compared to conventional volumetric options such as widefield or confocal microscopy. The unique physics of image formation from optical projections in optical cell CT precludes the application of noise models used in conventional widefield microscopy.

Signal degradation in transmission-mode optical cell CT images may be attributed to the following reasons:

- Optical path noise:

  Several components in the optical setup may contribute to the degradation of the optical signal that ultimately constitutes projection images. Particulate debris along the optical path, fluctuations in the excitation source, misalignments of mechanical components along the optical path, improperly functioning imaging cartridges that cause excessive capillary precession, and dark noise in the CCD camera are common factors that may decrease signal intensities in the acquired projection images.

- Unavailability of a point spread function (PSF):

  The availability of a PSF for an optical microimaging setup greatly improves the ability to mitigate optical path discrepancies through the process of deconvolution. However, in contrast to conventional microscopy, the unique image formation process in optical cell CT poses technical challenges that greatly complicate the computation of a consistent analytical or experimental PSF. Micron-scale calibration phantoms would facilitate a better estimation of system noise but the design and fabrication of such phantoms is beyond the scope of this dissertation.

- Image reconstruction artifacts:

  The generation of volumetric cell images from 2D optical projections involves mathematical algorithms to subtract the background, align projections, and finally reconstruct from projections. Errors in the algorithm sequence will hamper faithful reconstruction of the imaged cell. Errors in background subtraction or projection alignment propagate through the reconstruction algorithm, and result in increased image noise and reconstruction artifacts. The choice of reconstruction algorithm and its constituent filters also govern the noise pattern and its magnitude in the reconstructed volume.

- Sub-optimal staining of cellular specimen:

  The efficacy of optical cell CT largely depends on the contrast derived by staining the cellular specimen to distinguish it from its background. Despite the application of the best available reconstruction algorithm, poorly stained cells result in noisy reconstructions due to their low contrast relative to the background.

- Choice of reconstruction algorithm:

  Although the reconstruction algorithm aims to faithfully reproduce optical densities in the cell, the type of reconstruction algorithm applied in optical cell CT governs the complexity of segmentation. Analytical reconstruction algorithms such as filtered backprojection differ from one another in their ability to retain or accentuate low or high frequency components. Iterative reconstruction methods tend to be sensitive to high frequency components.

- Cellular specimen under investigation: Despite a faithful reconstruction, the underlying cell biology of the specimen also influences the efficacy of segmentation in several ways. Notable biological factors are:

  - Variation in cell and nuclear morphology as a function of cell cycle: Cellular structure and content dynamically changes according to the cell's current phase (G1-, S-, G2-, or M-) in the cell cycle. For instance, protein production changes cytoplasmic volume. Increases in DNA content during progression from G1 phase to G2 directly alter nuclear morphology. Hematoxylin staining thus differs from cell to cell as a result of these variations.

  - Variation in cell morphology as a function of malignancy: Cells exhibit numerous variations in their structure as a consequence of abnormalcy. In addition to the widely known shape and size variations with cancer progression, cells have been observed to become softer and more motile. Altered cell elasticity causes changes in the thickness of the cell membrane and nuclear envelope.

Non-nuclear organelles and proteins undergo changes in their abundance and localization. The position of the nucleus within the cell (referred to as polarity) changes with the onset of malignancy. These biological variations directly influence hematoxylin staining and consequently, the reconstructed image. Since a complete three-dimensional annotation of such biological traits is not yet available due to technological limitations, statistical shape models commonly adopted for segmentation may have limited efficacy.

- Phenotypic cellular heterogeneity: It refers to cell-cycle independent variations observed in the physical attributes of cells sharing the same genome. Heterogeneity in the 3D shape of cell nuclei, especially upon observation with images generated using the Cell-CT^{TM}further limit the efficacy of popular segmentation methods that use statistical shape priors or machine learning techniques.
- Suspended cell morphology used for optical cell CT: Epithelial cells typically exhibit a scrambled egg morphology while growing in their native state. However, cells used for optical cell CT need to be resuspended in solution prior to the staining procedure. Resuspension of cells predominantly results in a spherical morphology, which in turn causes a realignment of non-nuclear organelles such as mitochondria, endoplasmic reticulum, or golgi bodies in the cytoplasm. Spherical cell morphology may also affect nuclear structure, but this has not yet been definitively established. The altered cytoplasmic localization of non-nuclear organelles that are also stained with hematoxylin may hinder accurate segmentation of the nucleus.

3.3 Approach

To overcome the limitation of non-existent prior work on segmentation methods for transmission-mode optical CT image data, an extensive manual exploration of the images was first undertaken to understand hematoxylin staining patterns, its variation across cell type, and the effects of different reconstruction algorithms on cell images. The overall

strategy was to develop algorithms suitable for the target application rather than exhaustively test algorithms from literature solely on the basis of their popularity. The following observations were made and leveraged to develop segmentation algorithms for the target VOIs:

- High signal to background in correctly reconstructed images: Cursory manual interrogation of the image data with Volview software (Kitware, Clifton Park, NY) revealed that images with minimal errors in the reconstruction process exhibited a high intensity gradient between the stained cell and its background.

- Sequential segmentation of VOIs: Although hematoxylin is a non-stoichiometric dye, its extent of staining a cellular region is loosely proportional to the voxel intensities of that region. Thus, nucleus is preferentially stained relative to the cytoplasm and cell membrane, and UHD clumps within the nucleus are preferentially stained over other nuclear regions. Consequently, boundary intensities of target VOIs increased in the order: cell $<$ nucleus $<$ UHD clumps. Therefore, an incremental segmentation strategy was adopted to delineate our target VOIs in every cell (see Figure 3.1). Accordingly, the cell membrane was first delineated, then the nuclear envelope from the cell VOI, and finally the UHD clump boundaries from the nucleus VOI.

- Correlation between VOI boundary delineation and intensity histogram profile: Manual interrogation of the image data revealed that intensity thresholds sufficed to accurately segment the cell and UHD clump VOIs with the caveat that the threshold values varied from cell to cell. Additionally, optimal thresholds for these VOIs were observed to be located proximal to dominant peaks (modes) of the intensity histogram of the 3D image. The existence of a single cell per volumetric image implied that all the information in the intensity histogram corresponded to information about only that cell and its background.

- Correlation between image contrast and nucleus VOI boundary delineation: Manual interrogation of the image data also revealed a correlation between the ease of

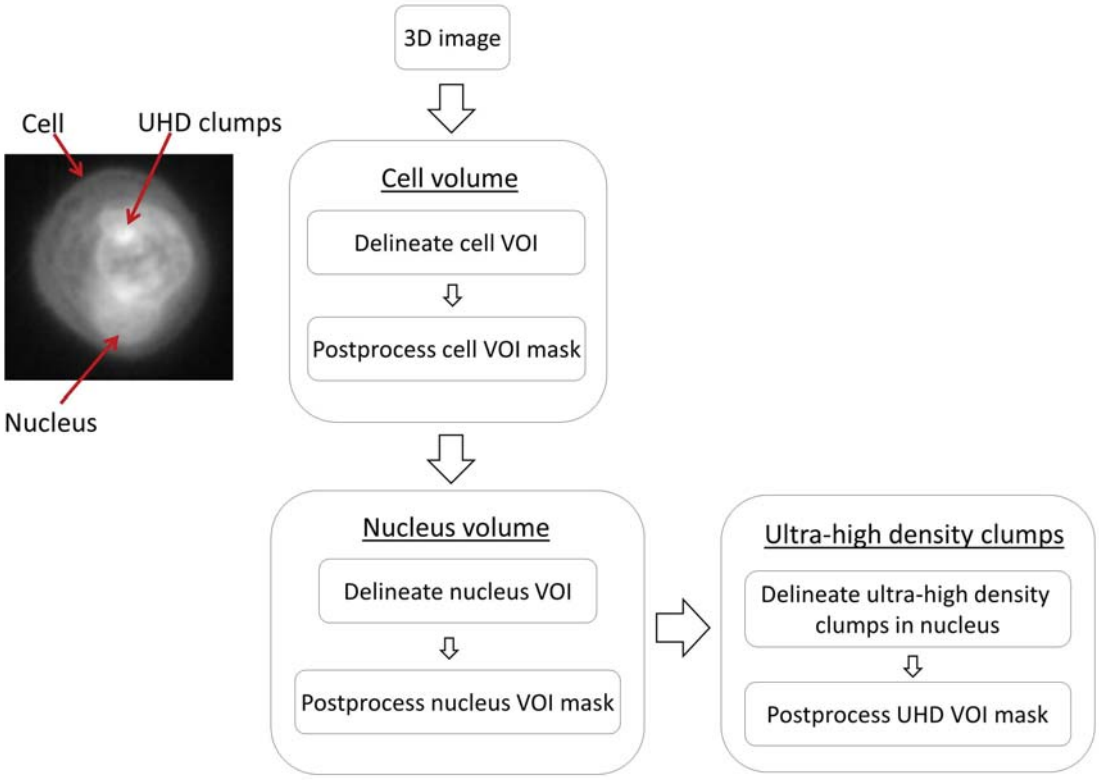

Figure 3.1: Sequential segmentation strategy for target VOIs in optical CT images.

nucleus VOI delineation and the contrast between the nucleus and cytoplasm resulting from hematoxylin staining. Intensity thresholds sufficed to segment the nucleus in cells exhibiting high contrast but were inadequate for cells with low contrast between nucleus and cytoplasm.

- Similarity in intensity histogram profile across bitdepth: The observations above on correlation between histogram profiles and VOI characteristics were roughly consistent across 8- and 16-bit reconstructions of the cell (see Figure 3.2). Hence, a decision was made to operate on the 8-bit image for the segmentation process due to its relatively smooth profile and its smaller size.

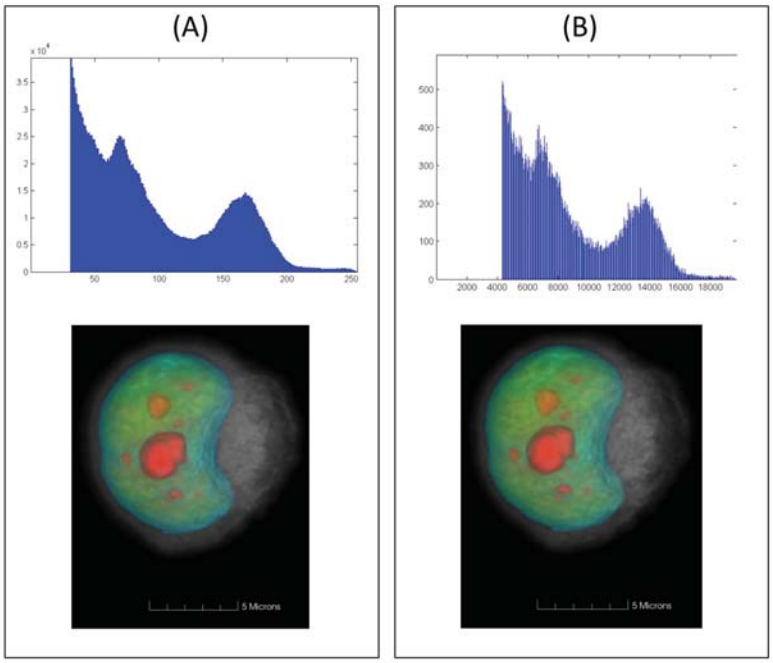

Figure 3.2: Intensity histogram profile of cell volume is similar between 8-bit (A) and 16-bit (B) grayscale images. Top images illustrate histogram profile of the segmented cell volume. Bottom images are manual renderings of the cell volume to further indicate similar morphologies

Based on the above observations, the proposed strategy was to first determine optimal intensity thresholds for the target VOIs sequentially from the intensity histogram and then refine the VOI boundaries using region-based techniques based on contrast considerations. Two options were utilized to rapidly identify optimal thresholds for the VOIs:

- The three orthogonal 2D maximum intensity projection (MIP) images of the 3D image: A MIP image along a particular direction is formed by preserving the highest intensity value along each parallel ray traversing the image volume (see Figure 3.3). The motivation for using MIP images is attributed to their inherently high SNR that naturally deemphasizes background noise. MIP images also tend to naturally accentuate VOI boundaries. The intensity threshold for each VOI in the 3D grayscale image was the minimum among the values computed in the 3 MIP images.

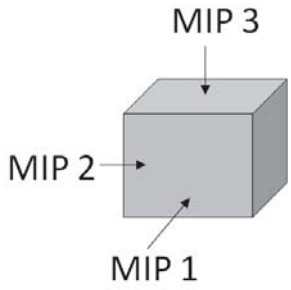

Figure 3.3: Illustration of orthogonal 2D MIP image formation from the 3D image

- The original 3D grayscale image: The motivation for this approach was its proximity to reality than the 2D MIP images.

For each VOI, a binary segmentation mask was generated by performing connected component labeling with a 26-neighbor connectivity constraint followed by a postprocessing step that smoothed the mask after removing spurious volumes based on an empirically-determined size constraint.

3.4 Image enhancement prior to segmentation

Caution is necessary during manipulation of tomographically reconstructed images due to the inherent connection between voxel intensity and the density at that location in the actual specimen. The micron order length scale of biological objects in optical CT images necessitates further discretion with the use of filters. For example, large filter sizes may obliterate thin structures such as the nuclear envelope or minute agglomerations of DNA inside the nucleus. The motivation for the application of image enhancement techniques was to only accentuate the VOIs for purposes of accurate segmentation. Operations were performed on the original grayscale image for feature computation after computing the necessary segmentation masks.

Traditional filtering techniques such as Gaussian, median, or mathematical morphology were not effective for segmenting target VOIs in optical CT images since they only caused a further blurring of object boundaries despite judicious choice of filter size.

Traditional image sharpening methods such as contrast stretching, log tranformation, or power transformation were relatively more effective in enhancing the cell relative to the background in properly reconstructed images. However, their efficacy was mitigated in noisy reconstructions due to their sensitivity to noise. Elevated noise streaks in noisy reconstructions would also get enhanced along with actual object voxels. Anisotropic diffusion is a pre-processing technique based on partial differential equations [62, 63] and is used to smooth image regions while retaining objects of interest such as edges. The extent of diffusion and its strength can be tailored to the object characteristics in the image. Therefore, anisotropic diffusion was adopted, followed by image sharpening to accentuate the target VOIs prior to segmentation. Since the log transform and power transform provided similar results, contrast stretching (CS) and power transform (PT) methods were evaluated in conjunction with anisotropic diffusion (AD). For optical CT images, 3 iterations of the anisotropic diffusion process with a diffusion coefficient of 5 sufficed to smooth background noise. To accomplish contrast stretching adaptively, the lower limit was set as the value obtained by applying Rosin's method [64] to the unfiltered image. The upper limit was set to the highest intensity of the image. The power law transformed version of the original image was obtained as follows:

$$I_{PT} = I^{1.5} \tag{3.1}$$

The power coefficient(1.5) was estimated empirically to suppress the background in optical CT images.

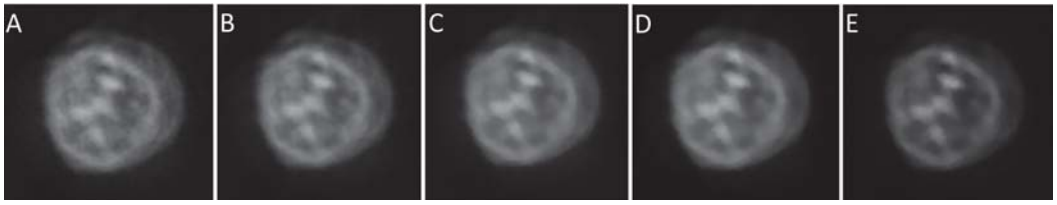

Figure 3.4: Comparison between image enhancement techniques on 2D slice of reconstructed image. (A) Original image, (B) $5 \times 5 \times 5$ Gaussian, (C) Anisotropic Diffusion, (D) Anisotropic Diffusion + Contrast stretching, (E) Anisotropic Diffusion + power law transformation

Figure 3.4 illustrates the effect of image enhancement techniques on a reconstructed image.

3.5 Proposed algorithms to segment the cell VOI

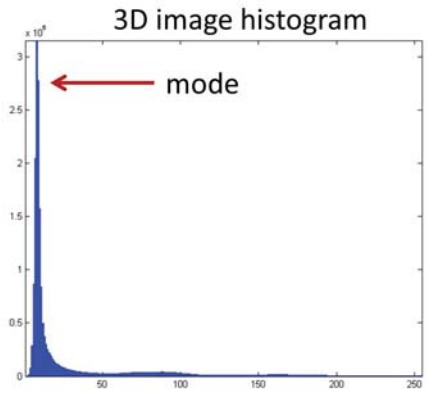

Figure 3.5: Distinct unimodal histogram profile of 3D image

Precise tomographic reconstruction of stained cells provides images with a marked contrast between the cell and its background. The distinct unimodal profile of the intensity histogram in the cell image (see Figure 3.5) motivated the application of Rosin's method [64] to identify the intensity threshold for the cell boundary.

2D MIP-based adaptive threshold

Rosin's method was applied to the intensity histograms of all 3 orthogonal MIP images and the minimum among the three values was selected as the intensity threshold for the 3D grayscale image.

3D intensity histogram-based adaptive threshold

The increased contrast inherent to MIP images tended to oversegment the cell boundary. Hence the analysis was modified to directly operate on the original 3D image. The image histogram was first smoothed using a 1D Gaussian filter, the mode of the histogram was identified, all intensities below the mode were thresholded out, and then Rosin's method was applied on the remainder of the histogram to obtain the threshold intensity for the cell boundary.

3.6 Proposed algorithms to segment the nuclear volume

2D MIP-based adaptive threshold

The threshold intensity for the nuclear envelope was computed using a gradient-based method on the cell segment of each MIP. The grayscale 2D MIP image was first normalized by the highest intensity value. The gradient magnitude was then computed at each point in the image from the $x-$ and $y-$ gradients. The resulting gradient magnitude image exhibited a distinct bimodal histogram. Otsu's thresholding method [65] is a technique that provides an optimal threshold for bimodally distributed data. Hence, it was applied and the gradient magnitude image was thresholded at the computed value to obtain the boundary of the nuclear envelope.

This boundary was superposed on the 2D grayscale MIP image and then then threshold intensity was calculated to be the average of all pixel intensity comprising the boundary. The minimum among the three threshold intensities (one per MIP) was set as the intensity threshold for the nucleus VOI.

Generalized Otsu threshold on 3D image histogram

The 2D MIP based method provided limited effectiveness in segmenting nuclei with low contrast or boundary inhomogeneities. To avoid oversegmentation of nuclei, techniques that utilized information from the original 3D image were evaluated. In light of the observations of intensity thresholds being proximal to the modes of the intensity histogram, the first intuition was to apply a threshold computed as a generalized formulation of the bimodal method originally proposed by Otsu.

Liao et al have provided such a method as part their proposed multi-level thresholding scheme [66]. First, the intensity histogram corresponding to the segmented grayscale cell volume was computed. A typical histogram has multiple (≥ 2) peaks or modes, depending on cell type (see Figure 3.6). The histogram was smoothed with a 1D Gaussian filter and then the intensity threshold for the nucleus VOI was set as the highest value determined by the method proposed by Liao et al.

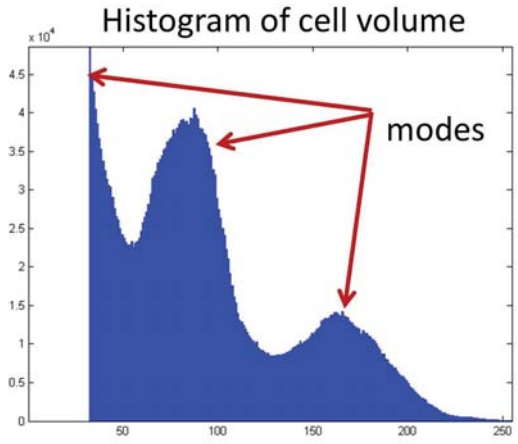

Figure 3.6: Multimodal histogram profile of cell volume

Heuristic k-means clustering on 3D image histogram

The generalized Otsu method was better than the 2D MIP option but still caused an oversegmentation of nuclei due to its inherent dependence on the variance of subpopulations in a mixture model. To circumvent this problem, a strategy was designed to be biased to the modes of the histogram. The heuristic k-means method (see Figure 3.7) seeks to locate the minima of the histogram by iteratively finding the geometric mean of the modes of the intensity histogram. The intensity histogram corresponding to the grayscale cell volume was computed, smoothed with a 1D Gaussian filter, the histogram modes were identified, and then the heuristic k-means clustering method was applied to determine the optimal threshold for the nucleus VOI. Iterations for the algorithm were discontinued when the limit ϵ became lesser than 0.5.

Heuristic contrast-driven adaptive thresholding

The heuristic k-means method was effective for cells with high contrast between the nucleus and cytoplasm, but it also resulted in oversegmentation of nuclei in low contrast images. The cause of this problem was identified as the insensitivity of these methods to contrast between the nucleus and cytoplasm, since contrast is directly reflected in the width of the histogram modes. This observation prompted the incorporation of additional

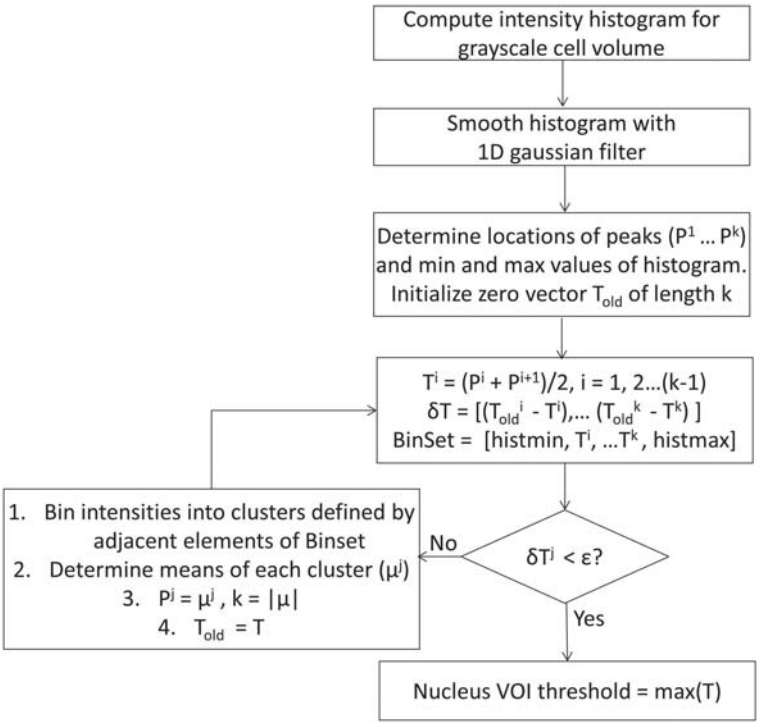

Figure 3.7: Flowchart illustrating k-means clustering method to determine threshold intensity for nucleus VOI

complexity by explicitly incorporating contrast into the proposed thresholding scheme. However, there is no straightforward quantitative metric to robustly define contrast between the nucleus and the cytoplasm. The problem is further complicated by the fact that the contrast varies from cell to cell. Therefore, a heuristic binary contrast function based on the histogram profile was defined as follows. First, the grayscale cell volume was derived by superimposing the binary cell mask onto the grayscale cell image. Then the graylevel intensity histogram was computed for the segmented cell VOI, smoothed with a one-dimensional Gaussian filter of size 10, and bin locations corresponding to the dominant peaks and minima were identified. A binary contrast function CF (CF=1 $\Rightarrow$ high contrast) was then derived from the histogram profile. For a histogram with m modes, CF is governed by three conditions $C1$, $C2$, and $C3$ that effectively reflect contrast between the nucleus and cytoplasm with information solely from the histogram.

Conditions $C1$, $C2$, and $C3$ are defined as follows:

$$CF = \begin{cases} 1, & C1 \wedge C2 \wedge C3 = 1 \\ 0, & C1 \wedge C2 \wedge C3 = 0 \end{cases} \tag{3.2}$$

where,

$$C1 = \begin{cases} 1, & max[histminima] > max[mlevel\ Otsu] \\ 0, & otherwise \end{cases} \tag{3.3}$$

$$C2 = \begin{cases} 1, & max[histpeak] > max[mlevel\ Otsu] \\ 0, & otherwise \end{cases} \tag{3.4}$$

$$C3 = \begin{cases} 1, & \frac{counts[histpeak_{m-1}]}{counts[histminima_m]} > 1 \\ 0, & otherwise \end{cases} \tag{3.5}$$

Parameters $histpeak$ and $histminima$ denote one-dimensional integer arrays of bin locations corresponding to the peaks and valleys of the intensity histogram. $m - levelOtsu$ denotes the Otsu thresholds for a histogram with m peaks.

The flowchart in Figure 3.8 illustrates the heuristic thresholding scheme proposed to segment the nucleus VOI. As described in the lower-right section of figure 3.8, the proposed method essentially identifies the bin location with the minimum count which precedes the highest intensity mode of the smoothed cell histogram. The heuristic was designed to threshold cell volumes more aggressively if high contrast existed between the nucleus and cytoplasm. A conservative definition was adopted to segment images with low contrast between the nucleus and the cytoplasm.

The intuition behind the use of conservative thresholds was to prevent an oversegmentation of the nucleus VOI. For cell volumes with low contrast but with more than 2 modes in their histograms, the histogram minimum in the vicinity of the 2-level Otsu threshold was identified. Cell volumes with low contrast and 2 dominant modes in their histogram are most segmented most conservatively in light of the difficulty in delineating the nuclear envelope.

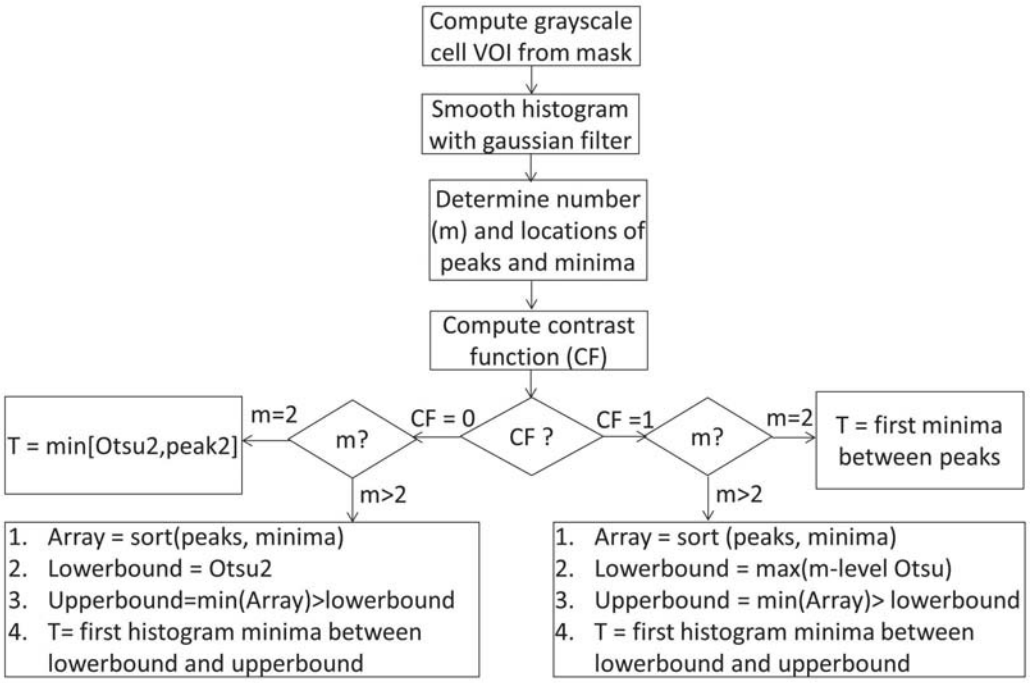

Figure 3.8: Flowchart describing the heuristic to segment the nucleus VOI using the 3D image histogram

Hybrid segmentation using heuristic adaptive thresholding and 2D GVF snake

Intensity thresholding schemes do not always succeed in correctly segmenting cell nuclei. Figure 3.9) illustrates a classic instance of this problem. There is no single intensity that can provide an accurately segmented mask. The problem was empirically observed to consistently arise in cell images that exhibit low contrast between nucleus and cytoplasm, and such cells as referred to as “problematic” cells.

Model-based segmentation techniques are better suited than adaptive thresholding schemes to accurately segment the nuclei of “problematic cells”. Model-based techniques for biological imaging applications have been a subject of extensive research since the 90s, especially after the development of the interactive active contour method (known as ’snake’) by Kass et al. [67]. Numerous classes of model-based segmentation methods have been developed for a variety of bioimaging modalities. Active contour methods and

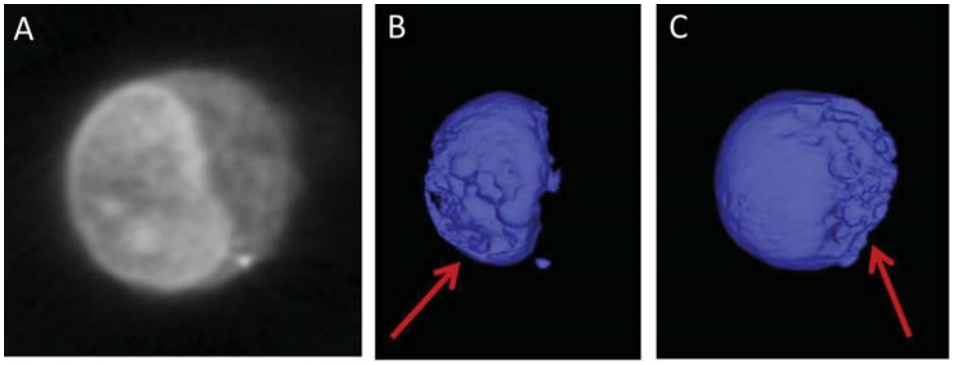

Figure 3.9: Representative image panel demonstrates the limitations of thresholding schemes to precisely delineate the nucleus boundary. (A) shows a 2D slice from a cell image. (B) is a volumetric rendering of conservative (under) segmentation. (C) is the volumetric rendering of aggressive (over) segmentation

their higher dimensional equivalents are the most commonly applied model-based techniques to segment ROIs in images. These methods accomplish segmentation by essentially letting a closed object (contour or a surface) modify itself within its physical constraints and under the influence of the underlying image characteristics. 3D or higher-dimensional images can be segmented either in a slice-wise manner using 2D models or by directly modifying surfaces or solid volumes. The latter options provide the advantage of smoothness but at the cost of markedly increased computational complexity. Active contour methods are broadly classified as either parametric or geometric methods depending on their formulation. A segmented mask using parametric techniques is the end result of an iterative procedure that minimizes the total energy of the model, wherein the total energy is expressed as a weighted sum of internal and external energy terms. Traditional energy-based formulations for parametric active contour segmentation are limited by their need to be initialized close to the actual object boundary and their inability to negotiate boundary concavities.

One popular alternative to circumvent these limitations is the Gradient Vector Flow (GVF) snake method proposed by Xu et al. [68]. GVF snakes use a force field-based formulation for the external energy term and is extendable to multi-dimensional images.

Despite their distinct advantages over traditional GVF-based techniques still require the right numerical coefficients for the constituent terms to achieve accurate segmentation. Further, 3D GVF snake models are not suitable for “problematic” cell nuclei due to intensity inhomogeneities over the nucleus boundary that are in turn caused by variations in the underlying staining. Hence, the propagation of a 2D GVF snake contour that deforms to arrive at the final segmentation mask for the cell nucleus was adopted. The segmentation result derived by the contrast-based heuristic described in section 3.6 was used as the initial mask.

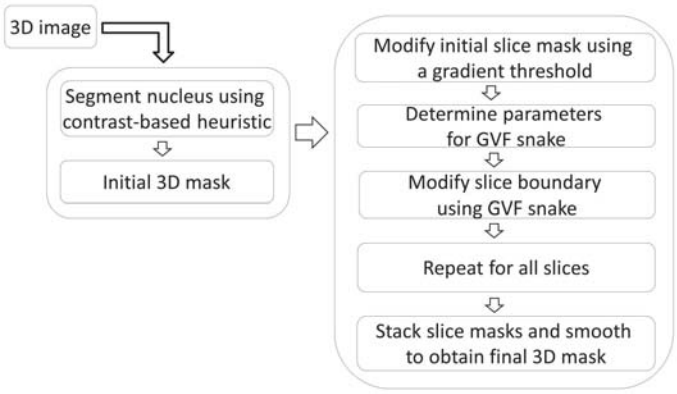

Figure 3.10: Flowchart describing the methodology to better segment the nucleus of low contrast cells using the 2D GVF snake

Figure 3.10 illustrates the proposed fully-automated segmentation methodology using 2D GVF snakes. For each image slice, the initial nucleus boundary mask was first refined to bring it in closer proximity to the actual nucleus boundary. This step was accomplished by thresholding the slice at the integer gradient value above which a contiguous closed contour does not exist i.e. when the actual nuclear membrane is breached. Checks were incorporated to ensure a stable but conservative refined boundary contour. The next challenge was to automatically identify the parameters for the GVF snake. Since the emphasis is on deforming the snake, the only requirement was to define the parameters for the external force terms and the number of iterations. Our hypothesis was that the values for these variables were related to the strength of the actual nucleus

boundary for that slice. Hence, the coefficients for the external forces were set to the gradient threshold T, and the number of iterations as T^2. The parameters associated with the internal forces were set to a constant value of 0.1 to ensure that the contour itself has maximum flexibility while still maintaining continuity.

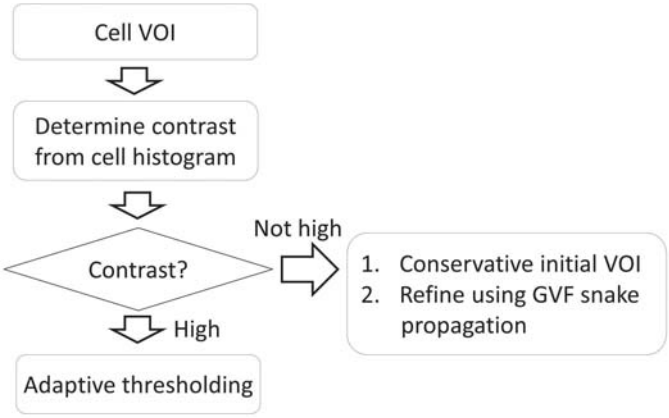

Figure 3.11: Hybrid fully-automated segmentation scheme to segment cell nuclei

The flowchart in Figure 3.11 illustrates the proposed strategy to segment cell nuclei based on their stain contrast relative to their cytoplasm. Adaptive histogram-based intensity thresholding schemes sufficed for nuclei with high contrast while a slice-wise 2D GVF snake propagation was necessary to segment nuclei that exhibited low contrast.

3.7 Proposed algorithms to segment ultra high density (UHD) clumps in the nucleus

Preliminary manual exploration of images indicated that UHD clumps typically assumed intensity values in the highest quartile of the image's dynamic range but differed in their individual threshold values (see Figure 3.12). Gradient-based segmentation methods are ineffective for this problem due to the low contrast between UHD clumps and the rest of the nucleus. Therefore, a conservative adaptive thresholding scheme was designed to ensure inclusion of the UHD clump with the lowest intensity value. To circumvent the possibility of spurious UHD clump segments, a size filter was applied to the initial binary mask. Only those clumps that contained more than 100 voxels (empirically determined) were retained as UHD clumps.

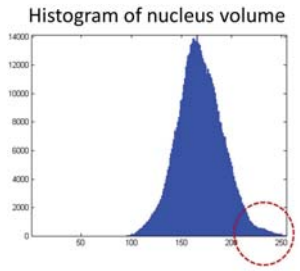

Figure 3.12: Intensity histogram of a segmented nucleus volume. UHD clumps typically exhibit intensities in the circled region

Two methods were developed to delineate UHD clumps.

2D MIP-based adaptive threshold

The grayscale nucleus segment in each MIP image was first computed, the intensity image was then inverted so that the UHD clumps were represented by intensities lower than the remainder of the nucleus. The intensity histogram of the resulting inverted image exhibited a distinct bimodal profile. A temporary intensity threshold was then computed for UHD clumps in the inverted MIP image by applying Otsu's method [65] to the histogram of the inverted image. The actual threshold intensity for UHD clumps was then calculated by subtracting the temporary threshold value from the maximum intensity of the MIP image. This procedure was performed for all MIP images and the threshold intensity for UHD clump VOI in the 3D grayscale image was defined as the minimum among the three threshold intensities.

3D intensity histogram-based adaptive threshold

The 2D MIP method for UHD clumps resulted in occasional oversegmentation, similar to the case with the cell and nucleus VOIs. Hence, the 3D image histogram was directly used to obtain a threshold for the UHD clumps. First, the intensity histogram of the nucleus volume was computed and smoothed with a 1D Gaussian filter. The mode of the histogram was identified, all intensities below the mode were thresholded out, and Rosin's method was applied on the remainder of the histogram to obtain the threshold intensity for the cell boundary.

Chapter 4

PERFORMANCE EVALUATION OF PROPOSED 3D SEGMENTATION ALGORITHMS

The efficacy of segmentation algorithms is typically evaluated by comparing their output to manual delineation of the same region(s) of interest by domain experts. Manually traced boundaries are collectively referred to as “ground truth”. A vast amount of ground truth data is available for conventional 2D and 3D microscopy techniques but was required to be generated for optical CT imaging. This chapter briefly outlines the establishment of ground truth data for transmission mode optical CT and then details the quantitative evaluation of proposed 3D segmentation algorithms developed in this dissertation research.

4.1 Establishment of ground truth

Image datasets for ground truth

A set of twenty faithfully reconstructed images comprising 20 correctly stained normal and cancer cells from either the human esophagus or breast epithelium were randomly selected to assess the efficacy of proposed segmentation techniques.

Ground truth establishment

A collaborative endeavor spearheaded by a laboratory colleague was undertaken to develop software to facilitate the establishment of volumetric ground truth for optical CT images acquired in transmission mode. The custom, interactive software interface was developed in Visual C# to record and archive tracings drawn with a stylus on a tablet display (Cintiq 12wx, Wacom, Vancouver, WA).

The software provided users with the choice of drawing contours along any of the three orthogonal axes (*XY*, *YZ*, *XZ*) of the 3D image along with options for manipulating image brightness and contrast. Volumetric ground truth data was generated by three domain experts who independently traced boundary contours of the target VOIs in every selected image. The exercise was restricted to the cell and nucleus VOIs since the complex 3D shape and spatial localization of UHD clumps posed significant technical challenges for accurate tracing and archival. The three users were initially trained on the software prior to the start of the exercise. To reduce the time consumed for establishing ground truth, the

gradual variation between VOI boundaries was leveraged and all users were instructed to mark contours on every alternate slice along the chosen orientation except at the perceived beginning or end of VOI boundaries. At these junctions, they were instructed to contour every slice. The software then approximated the intermediate (untraced) contours for the VOI by linearly interpolating neighboring contours. Lastly, the software generated binary masks for every VOI by setting all voxels enclosed within boundaries to 1 and archived the volumetric ground truth data in the form of binary image stacks.

Six of the twenty datasets were eliminated upon observing either errors or significant inter-user variability in their ground truth contours. The remaining 14 datasets were composed of 7 normal and 7 cancer cells. Cell and nucleus VOI volumes and their inter-user variability in the 14 ground truth datasets are listed in Tables 4.1 and 4.2. Inter-user variability was observed to be considerably larger for the cell than the nucleus.

Table 4.1: Inter-user variance in ground truth assessment of cell VOI

Cell Index	Cell VOI volume (μm^3)				
	User1	User2	User3	Mean	Standard deviation
1	2228.8	1499.8	2296.2	2008.5	441.9
2	2413.8	2385.9	2201.6	2333.8	115.3
3	1933.7	1372.0	1528.3	1611.3	289.9
4	1540.0	1041.3	1300	1293.8	249.4
5	2883.4	2285.9	2516.2	2561.8	301.3
6	2706.1	2119.4	2330.6	2385.4	297.1
7	1119.3	887.6	1077.7	1028.2	123.5
8	892	876.8	900.1	889.6	11.8
9	1572.2	1067	1243	1294	256.5
10	1566.7	1441.0	1623.8	1543.8	93.5
11	815	630	656.2	700.4	100.1
12	1123.8	1107.8	1110.4	1114	8.6
13	1408.1	1361.4	1584.2	1451.2	117.5
14	844.3	729.9	848	807.4	67.1

Table 4.2: Inter-user variance in ground truth assessment of nucleus VOI

Cell Index	Nucleus VOI volume (μm^3)				
	User1	User2	User3	Mean	Standard deviation
1	689.8	627	646.3	654.3	32.2
2	602.2	668.4	669.1	646.2	39
3	472.5	491.5	451.7	471.9	19.9
4	414.9	422.2	419.9	419.1	3.7
5	465.8	558.2	546.4	523.5	50.3
6	740.4	658.8	724.8	708	43.3
7	330	353.5	358	347.1	15.0
8	240.7	281.2	298.5	273.5	29.7
9	321.7	313.2	313.9	316.3	4.8
10	338.2	359.9	442.3	380.1	54.9
11	266.2	249.3	263.9	259.8	9.1
12	475.9	488.8	494.7	486.5	9.6
13	633.9	638.6	643.3	638.6	4.7
14	308.2	321.1	324.3	317.9	8.6

Generation of maximum likelihood VOI masks to accommodate inter-user variability

To manage the variability in the shortlisted fourteen datasets, a maximum likelihood (ML) strategy was developed to define a ground truth volume representative of the individual user traces of every VOI. The representative binary mask was called the 'MLMask' and it was computed using a majority voting scheme. A voxel in the MLMask was assigned a value of 1 if the corresponding voxels in two or more user masks had a value of 1. Otherwise, the voxel was assigned a value of 0. The method is summarized in Equations 4.1 and 4.2.

Given volumetric binary masks $Mask_u$,$u \in 1,2,3$ for any VOI,

$$\text{UnionMask(i,j,k)} = \sum_{u=1}^{3} Mask_u(i,j,k) \tag{4.1}$$

$$\text{MLMask(i,j,k)} = \begin{cases} 1, & \text{UnionMask(i,j,k)} \geq 2 \\ 0, & otherwise \end{cases} \tag{4.2}$$

where, i,j, and k are voxel indices along the $x-$,$y-$ and z directions respectively of the image.

The cell and nucleus VOI volumes derived from the maximum likelihood masks for the 14 datasets are listed in Table 4.3.

Table 4.3: Cell and nuclear volumes from maximum likelihood (ML) VOI masks

Cell Index	ML cell volume (μm^3)	ML nuclear volume(μm^3)
1	2176.9	637.1
2	2323.1	636.5
3	1551.2	472.3
4	1314.5	417.3
5	2526.3	532.8
6	2336.7	702.8
7	1054.2	348.2
8	889.4	272.1
9	1247.5	313.8
10	1556.9	370.0
11	666.7	256.8
12	1112.8	489.2
13	1407.7	641.9
14	825.1	311.3

4.2 Performance evaluation of 3D segmentation algorithms

Metrics for performance evaluation

Numerous quantitative metrics are available to evaluate the agreement between segmentation algorithms and ground truth data [69]. For brevity, the DICE coefficient was used in this investigation to assess the efficacy of proposed segmentation algorithms. The metric is defined in Equation 4.3:

$$\text{DICE} = \frac{2 \times |G_t \cap S_a|}{|G_t| + |S_a|} \tag{4.3}$$

where G_t is the ground truth segmentation, S_a is the algorithmic segmentation, $|.|$ denotes the number of voxel elements in the set. A DICE value of 1 represents perfect agreement with ground truth.

Evaluation of algorithms for cell segmentation

To identify the best segmentation method, DICE coefficients were evaluated between the ML ground truth volume and cell masks derived from the unfiltered image. Subsequently, the effect of image enhancement techniques was tested on the accuracy of the best performing segmentation technique. The segmentation results for the two methods proposed to segment the cell VOI are summarized in Table 4.4.

Table 4.4: Performance evaluation of 3D segmentation methods for cell VOI (unfiltered image, 14 cells)

Segmentation method	Average DICE coefficient (%)
2D MIP-based method	89.8
3D histogram-based method	90.9

The 3D histogram-based method was observed to provide a relatively better segmentation of the cell boundary. The disagreement between algorithmic and user-defined boundaries was primarily attributed to the conservative estimation by the algorithm. While users defined the boundary based on visual contrast, the algorithmic threshold was in a shell just outside the perceivable visual contrast. The application of image enhancement methods prior to segmentation reduced the disagreement.

Table 4.5: Performance assessment (% DICE) of 3D histogram-based method on cell VOI after image enhancement

Unfiltered	Gaussian	Median	AD	ADCS	ADPT
90.9	91.1	90.9	90.1	94.5	93.2

The performance of the 3D histogram-based method after the application of image enhancement techniques is summarized in Table 4.5. A significant improvement was achieved in segmentation accuracy by applying conjunctive enhancement methods instead of traditional methods such as Gaussian or median methods. Contrast stretching (CS) and power transform (PT) after anisotropic diffusion (AD) were abbreviated as ADCS and ADPT respectively. The improvement in segmentation accuracy by application of image enhancement methods is illustrated in Figure 4.1.

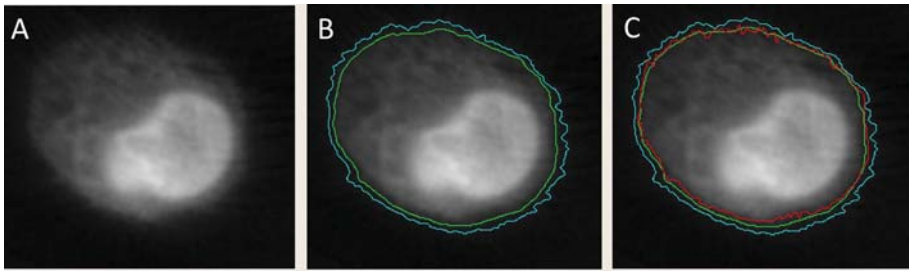

Figure 4.1: Effect of image enhancement on cell segmentation accuracy. 2D boundary contours overlaid on original image slice (A) after running segmentation algorithm on unfiltered (cyan) and filtered (green) images (B). Algorithm performance relative to ground truth contour (red) is shown in (C).

Evaluation of algorithms for nucleus segmentation

The accuracy of proposed methods for segmenting the nucleus VOI before and after image enhancement are summarized in Table 4.6. The presented results are focused on the proposed 3D segmentation methods since they performed better than the 2D MIP-based method. All four proposed methods provided a segmentation accuracy of at least 90% on

Table 4.6: Performance evaluation of 3D segmentation methods for nucleus VOI (% DICE averaged over 14 cells)

Segmentation method	Unfiltered	3D Gaussian	AD	ADCS	ADPT
Generalized Otsu	93.1	93.1	93.4	92.5	71.4
Heuristic k-means clustering	92.5	92.8	93	92.8	88
Heuristic contrast-driven	90	90	89.7	91.3	94.1
2D GVF snake	92.9	92.4	93.5	93.1	93.6

the unfiltered images. The generalized Otsu and heuristic k-means techniques had relatively higher DICE coefficients for unfiltered image but their performance decreased with the use of ADCS or ADPT image enhancement procedures. In contrast, the contrast-driven heuristic method had relatively lower DICE coefficient with the unfiltered image but its performance improved with the use of ADCS or ADPT enhancements. The performance of the 2D GVF snake method did not vary significantly with the use of enhancement techniques.

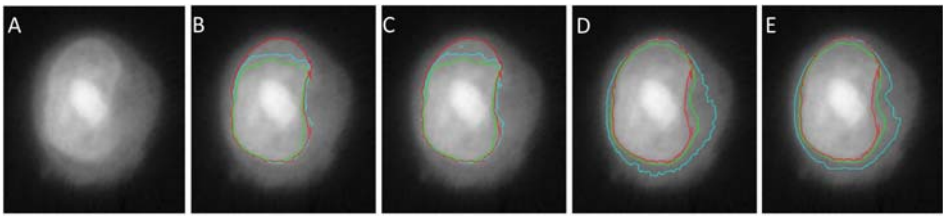

Figure 4.2: Overlaid algorithm boundary contours relative to ground truth (red) illustrate performance and oversegmentation errors of methods proposed for nucleus segmentation before (cyan) and after (green) image enhancement. (A) is the original image slice. (B), (C), (D), and (E) are results from Generalized Otsu, heuristic k-means clustering, contrast-driven heuristic method, and 2D GVF snake-based hybrid method respectively.

Comparative overlays of the four proposed methods relative to user-defined ground truth are shown in Figure 4.2. The 2D GVF snake method provided good performance with unfiltered and filtered images. Close observation of the segmentation overlays revealed the oversegmentation vulnerabilities of the generalized Otsu and heuristic k-means methods, especially in cells exhibiting low contrast between nucleus and cytoplasm. On the other hand, the contrast-driven heuristic method provided a conservative initial boundary which is then refined by the application of image enhancement methods. The limitation posed by oversegmenting methods is masked by the DICE metric since it is less sensitive to oversegmentation than undersegmentation.

Chapter 5

3D FEATURE COMPUTATION AND STATISTICAL ANALYSIS

This chapter details the 3D numerical descriptors (features) that were computed to characterize morphology in the reconstructed 3D cell images and the suite of classification and feature selection methods that were implemented for biosignature discovery.

5.1 Feature computation

The choice of features to be quantified was motivated by their intuitive correlation with the underlying structural biology and their relevance to clinical cytopathology assessment. Thus, the focus of this research was primarily on features that characterize nuclear shape, structure and texture. All features were implemented in Matlab (Mathworks, Natick, MA), and integrated into the automated computational software framework. The framework automatically computed feature data from the segmented VOI volumes and archived the data in Microsoft Excel spreadsheet files at a specified location. The framework also provided an option for computing and archiving equivalent 2D features (derived from a 2D grayscale image) for the list of 3D features.

Morphological features

Morphological features characterized structural and geometric aspects of VOIs. Computed features included the volume and surface area of the cell, nucleus, and ultra-high density (UHD) nuclear clumps.

Table 5.1: List of morphological features

Feature type	Features
Morphological	cell volume, cell surface area, nuclear volume, nuclear surface area, nucleus to cytoplasm (NC) volume ratio, number of UHD clumps, total volume, average volume, and variance in volume of UHD clumps, average margination and variance in margination of UHD clumps from nucleus center, nuclear sphericity, eccentricity, aspect ratio, major axis length and minor axis length of nucleus

The volumetric nucleus to cytoplasm (NC) ratio was determined from the cell and nuclear volumes. Cell polarity was implemented as the physical distance between the geometric centroids of the cell and nucleus. The structural geometry of the nucleus was characterized by computing its sphericity, largest eccentricity, aspect ratio, and the longest major and minor axis lengths. Lastly, the number and size of UHD clumps were computed and the heterogeneity in their size and spatial location was estimated using descriptive statistics (see Table 5.1).

Texture features

Texture features characterized the organization of coarse chromatin within the nucleus. Three types of texture features were implemented: descriptive, discrete and Markovian.

Descriptive texture features

Table 5.2: List of descriptive texture features

Feature type	Features
Descriptive	integrated optical density, mean optical density, variance in optical density, skew in optical density, kurtosis in optical density

Descriptive texture features characterized nuclear content heterogeneity by computing global statistics from the nuclear voxel intensity histogram. The total, mean, and higher order moments were computed to serve as descriptive texture features (refer Table 5.2).

Discrete texture features

Discrete texture features characterized the variations in density of nuclear regions and their relative spatial locations. To compute these features, the nuclear interior was divided into three distinct regions representing low, medium and high optical density. The intensity thresholds separating low- and high-density regions from medium-density regions were set at a single standard deviation around the mean nuclear intensity. The number, size (volume), compactness, relative proportion, and localization of each density region relative

to the other were also computed. The relative spatial location of density regions were estimated by their geometric centroids and corresponding inter-centroidal distances. The complete list of discrete texture features is presented in Table 5.3.

Table 5.3: List of discrete texture features

Feature type	Features
Discrete	number of objects in low, medium and high condensation states (3), volume fraction of each condensation state (3), content (optical density) fraction of each condensation state (3), smoothness of transition between condensation states (3), compactness of each condensation state (3), fractal dimension of each condensation state (3), total surface area of each condensation state (3), total volume of each condensation state (3), heterogeneity in volume of each condensation state (3),average distance of center of each condensation state from the nucleus center (4)

Markovian texture features

Markovian features characterized the directional variations in intra-nuclear densities at different granularities. In this research, Markovian features were derived using the method based on gray level co-occurence matrices (GLCM) proposed by Haralick [70]. For a grayscale image with bit depth k, a GLCM is of size $2^k \times 2^k$ and annotates the frequency of occurrence at a physical distance 'd' (voxels) of all voxel intensities spanning the range $i \in [0, 2^k]$. 'd' represents the granularity of the texture. Eleven Markovian features (see Table 5.4) were computed from 26 gray level co-occurrence matrices (GLCMs) (corresponding to the 26-voxel neighborhood) centered at the geometric center of mass of every segmented nucleus. The 26 values were averaged to obtain one representative value for each feature at the selected offset. Voxel offsets of 7 and 14 were used to capture coarse density variations.

Table 5.4: List of Markovian texture features

Feature type	Features
Haralick	energy, contrast, correlation, homogeneity, sum of squares variation, sum average, sum variance, sum entropy, difference variance, difference entropy, information measures of correlation (2)

5.2 Classification and feature selection

Classification and feature selection are critical computational components necessary to define robust biosignatures. In this dissertation, classification refers to the process of predicting a cell type on the basis of its numerical feature set computed from optical CT images. It involves the development of predictive computational models from the given feature data. Feature selection is the process of identifying the subset of features that possess the most discriminatory power when used in a classifier. It eliminates redundant features and thus facilitates the development of efficient classifier models.

Methods for classification and feature selection have been the subject of extensive research in the past decade. Classification methods are broadly divided into supervised or unsupervised learning methods, depending on the strategy used to build the predictive model from the feature data. An additional, more recent taxonomy among supervised learning methods is the use of stand-alone versus ensemble strategies. While the stand-alone strategy utilizes a single learning method for classification, ensemble methods use an array of individual classifier models to perform classification. Feature selection methods are also broadly divided into either filter, wrapper, or embedded techniques depending on their mode of operation [71]. Filter methods select features purely on the basis of some correlation metric on the feature data prior to classification. Their performance is thus independent of the classification technique with which they are applied. Wrapper methods select features on the basis of their performance with a classifier (usually assessed by computing the misclassification error). Embedded methods are hybrid strategies that utilize classifiers to select features as part of the learning process. Wrapper and embedded feature selection methods require more computational power and time than filter methods.

This dissertation applied a suite of popular statistical classification and feature selection methods that have been shown to perform excellently for a wide variety of biological applications. For classification, the focus was predominantly on stand alone and

ensemble supervised learning methods. The option of building classifier models with either k-fold crossvalidation or leave *N* out methods are provided in the computational framework. The methods currently available in the computational framework are listed in Table 5.5 and Table 5.6.

Table 5.5: List of classification methods implemented for biosignature discovery

Stand-alone methods	Ensemble methods
Naive Bayes	Decision trees
Discriminant analysis	Bagged trees
Support vector machines	Boosted trees
neural networks	
logistic regression	
K nearest neighbor	

Table 5.6: List of feature selection methods available for biosignature discovery

Method type	Method list
Filter methods	mRmR, mutual information criterion, T-test
Wrapper methods	forward and backward feature selection with stand-alone learners
Embedded methods	forward and backward feature selection with ensemble learners

Numerous performance metrics to evaluate the efficacy of classifiers were also implemented. These are listed below:

1. Global metrics: mean accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and their respective standard deviations

2. ROC-based metrics: ROC plot with 95% confidence intervals, area under the curve (aROC or AUC) with 95% confidence intervals, and optimal ROC point determination. The optimal ROC point is calculated as the 2D point located on the ROC plot and which is most proximal to the ideal ROC point of (0,1).

Chapter 6

3D MORPHOLOGICAL BIOSIGNATURES FOR ESOPHAGEAL ADENOCARCINOMA PROGRESSION

6.1 Motivation

Esophageal adenocarcinoma (EA) is among the fastest growing cancers in the western world [72]. The poor survival rate associated with the disease necessitates robust early detection for a good prognosis. The most well-known precursor for EA is Barrett's Esophagus (BE): a condition characterized by replacement of normal squamous epithelium by metaplastic columnar epithelium in the distal esophagus [73]. The onset of BE has primarily been attributed to factors such as gastroesophageal reflux disease (GERD), obesity, and smoking [74]. Patients with BE are considered to be at a considerably high risk for EA. Neoplastic progression in BE typically proceeds from non-dysplastic metaplasia to low-grade and high-grade dyplasia, and finally to invasive adenocarcinoma. Endoscopic examination of the esophagus is the most commonly adopted screening method to identify BE. Patients with BE are usually recommended to regular endoscopic surveillance programs that check for potential progression to adenocarcinoma. A major translational research objective is to determine the subset of metaplastic BE patients who will progress onto dysplasia or EA. The definitive diagnostic procedure to determine the existence of BE or EA in patients is still slide-based pathology of tissue biopsied by a surgical endoscopist. The limitations of this approach have been detailed in chapter 1. Further, cost-effective alternatives to endoscopy are required for surveillance and patient management.

Computer-aided 2D morphometric analysis of digitized H&E stained BE tissue samples has been pursued in the past to predict progression from metaplasia to dysplasia [34] and adenocarcinoma [75]. These studies reported variations in nuclear area, shape and texture with progression. This dissertation aims to derive sensitive and specific three-dimensional morphological biosignatures predictive of progression in BE by 3D image analysis of optical cell CT data. The research also aims to investigate the relative efficacy of 3D biosignatures over 2D morphological signatures derived from digital images

that approximate conventional pathological observation with a lens having an optical magnification of 10x. Optical CT imaging was applied and the computational framework developed as part of this dissertation to were used to facilitate the investigation.

6.2 Methods

Cell culture

Immortalized esophageal epithelial cell lines representative of normal squamous (EPC2, [76]), metaplastic (CPA) and dysplastic (CPD) BE [77], and adenocarcinoma (FLO-1, [78]) were used in this study. All cell lines were grown per prescribed protocols as monolayers to approximately 85% confluence in 25-cm^2 flasks with vented caps (BD Falcon, San Jose, CA). Serum-free keratinocyte medium was used to grow EPC, CPA, and CPD cells. FLO-1 cells were grown in DMEM. Cultured cells were trypsinized with trypsin (Mediatech, Manassas VA) and then resuspended in PBS in preparation for optical cell CT imaging. EPC2, CPA, and CPD cells needed 0.05% trypsin while FLO-1 cells required 0.25% concentration.

Sample preparation and tomographic imaging

The procedure detailed in chapter 2 was followed to prepare stained cell samples and image them individually by optical CT. Briefly, suspended cells from all four lines were fixed, stained with Gill's 1 hematoxylin (Electron Microscopy Sciences, Hatfield, PA), and imaged individually by optical CT using the Cell-CT^{TM}(VisionGate, Inc., Phoenix, AZ). 250 images were acquired for every cell type and used for morphometric analysis.

3D image processing

The combination of best performing methods detailed in chapter 3 were used to segment the cell, nucleus, and UHD clumps in reconstructed 3D images. Prior to segmentation, the images were enhanced using anisotropic diffusion and power transformation in sequence (refer chapter 4). 79 numerical features (refer chapter 5) were computed to characterize the 3D morphology of each cell and its nucleus.

2D image processing

Since it was not feasible to retrieve the exact cells that were imaged with the Cell-CT™, a 2D image was simulated to approximate acquisition using a 10x objective lens in brightfield mode. A 2D image was derived from the reconstructed 3D image by averaging intensity information over a distance equal to the depth of field (8.5 microns) provided by a 10x lens. The 8.5 micron depth of field in a 10x lens corresponded to an averaging of a slab of 115 image slices that was centered at the center of mass (centroid) of the nucleus in the reconstructed image.

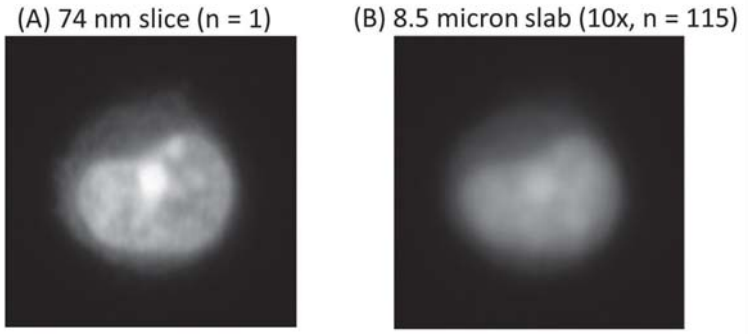

Figure 6.1: Comparison between a 74 nm thick 2D slice in a reconstructed 3D cell image (A) and a 2D slice simulating widefield acquisition with a 10x objective lens (B).

A representative 10x simulation alongside a slice of the reconstructed image is shown in Figure 6.1. The simulated 2D image is considerably blurred compared to the 3D image. The poor contrast in the simulated image precluded the development of computer vision methods that could reliably segment the target regions of interest (ROIs). Therefore, the 3D segmentation masks derived earlier were leveraged to obtain 2D masks for target ROIs in the simulated 2D image. To obtain segmentation masks for each 2D ROI, the 3D binary masks were averaged over the 115 slices and then binarized the resulting image at a value of 0.1 which was empirically determined by manual exploration. Lastly, 79 2D features were computed from the segmented VOIs in the 2D images. These 2D features were equivalent to the 3D features estimated from the volumetric images.

Statistical analysis for biosignature discovery

The supervised learning and feature selection methods listed in chapter 5 were investigated to derive morphological biosignatures that distinguished normal squamous (EPC2), metaplastic (CPA), dysplastic (CPD), and EA (FLO-1) cells with the highest sensitivity and specificity. The most sensitive and specific 2D and 3D biosignatures were derived for each of the following eight pairwise comparisons between the four cell types:

1. Normal squamous (EPC2) versus adenocarcinoma (FLO-1)
2. Normal squamous (EPC2) versus dysplastic Barrett's (CPD)
3. Normal squamous (EPC2) versus metaplastic Barrett's (CPA)
4. Normal squamous (EPC2) versus all abnormal cells (CPA + CPD + FLO-1)
5. Control (EPC2 + CPA) versus diseased (CPD + FLO-1)
6. Metaplasia (CPA) versus dysplasia (CPD)
7. Metaplasia (CPA) versus adenocarcinoma (FLO-1)
8. Dysplasia (CPD) versus adenocarcinoma (FLO-1)

For each of the above comparisons, the best classifier models was first identified using all 79 morphological features. A total of 16 classifier models were derived from corresponding sets of 2D and 3D morphological feature data.

The best classifier model was defined as the one which provided the highest combination of accuracy, sensitivity, specificity and area under the ROC curve (aROC or AUC) averaged over the 10 cross-validation runs. The positive and negative predictive values (PPV, NPV) were also computed for each comparison to assess comparisons where the other metrics did not suffice to enable a clear decision. Feature selection methods were then applied to the best performing classifier models in order to further streamline their

performance and concurrently identify the set of most discriminatory features between two given cell types. The minimum-redundancy maximum-relevance (mRMR) method [79] was applied with the best classifiers that were derived for each comparison. The mRMR method is a filter-based technique that essentially ranks features according to their relevance after eliminating redundant feature data. The method has been shown to be very effective for a wide variety of biomedical applications. To determine the best set of most powerful features for each comparison, the best classifier previously identified for that comparison was incrementally trained using the set of k ($k \in [1, 5, 10, 15...75, 79]$) best features identified by the mRMR method. The value of k that maximized classification accuracy was selected to be the optimal feature set for that comparison. The resulting streamlined classifier would have the minimum number of most discriminatory features necessary to obtain the best classification accuracy. This procedure was repeated for all 16 classifiers to determine the variation in the degree of learning required for 2D and 3D morphological analysis. Lastly, the effect of sample size on classifier performance was analyzed as a form of retrospective power analysis to estimate the minimum number of cells required for good classification performance.

6.3 Results

Measurable alterations in 3D cell and nuclear structure accompany progression to esophageal adenocarcinoma

Appreciable differences were observed in 3D cellular morphology between EPC2, CPA, CPD, and FLO-1 cell lines (see Figure 6.2). Cell size, nuclear size, nucleus-to-cytoplasm (NC) ratio, and the clumpiness of nuclear content within the nucleus differed as cells progressed from normal to cancer. Histograms of some visually appreciable and diagnostically relevant morphological features are shown in Figure 6.3. As evidenced by the histogram distributions, single-cell optical CT enabled a better appreciation of the marked intra-population phenotypic heterogeneity in measured morphological parameters. Normal squamous cells expectedly exhibited the smallest cell and nuclear size among the four cell types. Similarly, cancer cells exhibited the highest NC ratio, followed by dysplastic cells. The large size of metaplastic cells could be an artifact from immortalization.

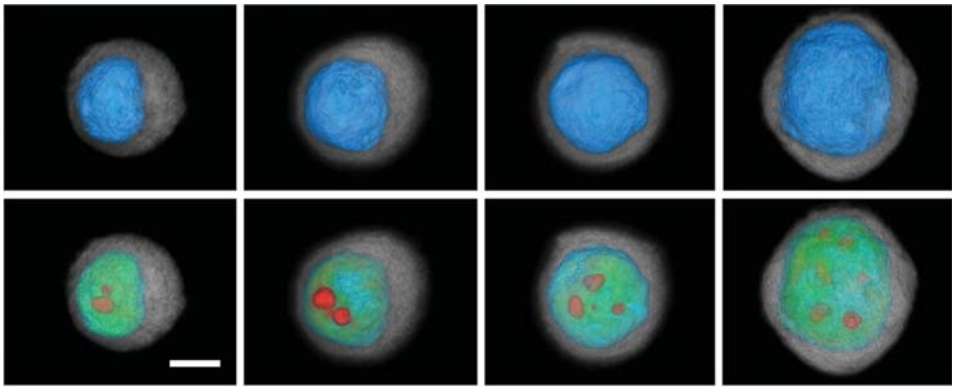

Figure 6.2: Cellular morphology varies with progression from normal squamous to esophageal adenocarcinoma. (Left to right) Pseudocolor volume renderings show distinct changes in three-dimensional cell and nuclear structure that accompany progression from normal squamous through metaplasia, dysplasia and to esophageal adenocarcinoma. Top row shows a surface rendering with the cytoplasm in gray and the nucleus in blue. Bottom row illustrates corresponding transparent views to provide information about the nuclear interior. Increasing density of intra-nuclear content is color coded from green to red. Scale bar = 5 microns.

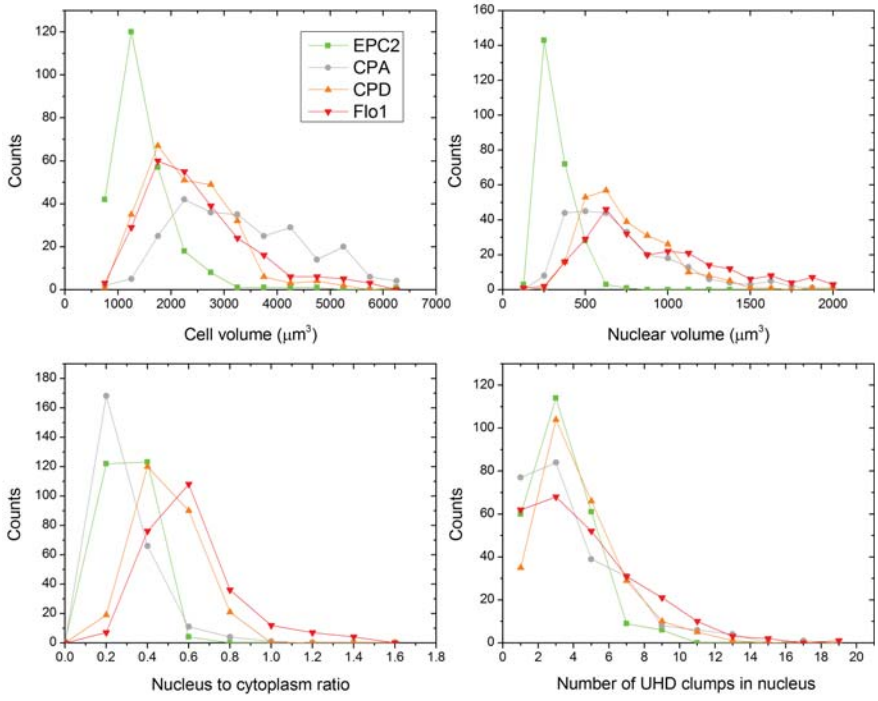

Figure 6.3: Histograms of visually apparent 3D morphological features illustrates the appreciable intra-population phenotypic heterogeneity.

Development of morphological biosignatures

Quantitative 3D morphological analysis indeed provided sensitive and specific biosignatures to distinguish normal, metaplastic, dysplastic, and cancer cells. The best classifier models for each comparison are listed in Table 6.1. Linear discriminant analysis (LDA) was observed to work well for our application, especially for the 2D analysis. This result indicated the existence of gaussian or near gaussian distribution in feature data. Ensemble methods such as Adaboot or logistic boost also proved to be useful classifiers for a subset of the comparisons.

Table 6.1: Best classifier models for each comparison

Comparison	3D analysis	2D analysis
EPC vs FLO-1	LogitBoost	LDA
EPC vs CPD	LDA	LDA
EPC vs CPA	LDA	LDA
EPC vs Abnormal	LDA	LDA
(EPC,CPA) vs (CPD,FLO)	LDA	LDA
CPA vs FLO	LDA	LDA
CPA vs CPD	LogitBoost	LDA
CPD vs FLO	Adaboost	LDa

Performance statistics using the best classifier models are detailed in Table 6.2. Normal squamous cells could clearly be distinguished from cancer(99% sensitivity and 97% specificity) and dysplastic cells (95% sensitivity and 96% specificity). Fairly robust classification was also achieved to independently distinguish cancer cells from metaplastic (94% sensitivity and 93% specificity) and dysplastic cells (94% sensitivity and 92% specificity). Metaplastic cells could be distinguished from dysplastic cells with fairly good accuracy (91% sensitivity, 93% specificity) as well. The poorest performance was observed between squamous and metaplastic cells (91% sensitivity, 86% specificity). This result reinforces our qualitative visual observations of the cells and could be a fall out of their immortalization. Normal cells could be distinguished from metaplastic, dysplastic and cancer cells with high specificity (95%) and moderate sensitivity (86%). The high specificity indicates the efficacy of this comparison for screening applications. The relatively low

sensitivity could be attributed to the class imbalance and the weak classification between normal and metaplastic cells. The comparison between the more clinically relevant combination of normal and metaplastic cells versus dysplastic and cancer cells yielded good classification (93% sensitivity, 91% specificity). Similar trends were observed with 2D analysis. However, the results in Table 6.2 indicate that 3D morphological analysis consistently outperformed the equivalent 2D analysis for all eight comparisons. Figures 6.4 and 6.5 illustrate comparative ROC plots along with their upper and lower 95% confidence intervals (CIs), mean area under the ROC curve (AUC) and optimal operating point for each of the 8 comparisons. The 95% CIs are plotted in dotted lines. Results from the ROC plots reinforce the conclusions drawn from Table 6.2 about the relative merit of 3D analysis over 2D analysis to distinguish the four cell types. The 95% CIs for the ROC plots associated with 3D analysis were relatively tighter than the corresponding 2D plots. Tight CIs (especially the lower interval) indicate low variance in performance and thus show promise for clinically relevant tests. The weak classification between normal and metaplastic cells is also evident from the ROC plot of Figure 6.4C.

Table 6.2: Comparative performance summary of best classifier models using all 79 features, 250 datasets, 10-fold crossvalidation. Acc = Accuracy, Sens = Sensitivity, Spec = Specificity, PPV = Positive Predictive Value, NPV = Negative Predictive Value, AUC = Area under ROC Curve

Comparison	3D feature analysis						2D feature analysis					
	Performance metrics(%)						Performance metrics (%)					
	Acc	Sens	Spec	PPV	NPV	aROC	Acc	Sens	Spec	PPV	NPV	aROC
EPC vs FLO-1	98 ± 0.06	99 ± 0.04	97 ± 0.2	97 ± 3.5	98 ± 2.4	99.45	94 ± 3	97 ± 2.9	91 ± 6.4	91 ± 5.5	97 ± 2.8	97.88
EPC vs CPD	96 ± 2	95 ± 2.3	96 ± 3.3	96 ± 3.5	95 ± 4.2	99.14	95 ± 4.3	95 ± 4.5	94 ± 6.4	95 ± 5.8	94 ± 5.4	97.68
EPC vs CPA	89 ± 3.7	91 ± 6.7	86 ± 6.8	87 ± 4.9	91 ± 7.2	95.48	88 ± 4	90 ± 4.1	87 ± 6.5	88 ± 4.2	88 ± 7.6	95.17
EPC vs Abnormal	93 ± 2.6	86 ± 6	95 ± 3.1	85 ± 9.1	95 ± 2.5	96.77	92 ± 2.5	82 ± 8	96 ± 2.7	86 ± 8.8	94 ± 3.1	95.89
(EPC,CPA) vs (CPD,FLO)	92 ± 2.2	93 ± 3.8	91 ± 4.2	94 ± 4.4	93 ± 3.5	97.36	89 ± 3.1	90 ± 3.7	88 ± 5.1	88 ± 5.6	90 ± 3.5	94.63
CPA vs CPD	92 ± 5.9	91 ± 7.1	93 ± 5.1	92 ± 6.5	92 ± 5.8	97.39	89 ± 3.9	90 ± 5.8	89 ± 5.6	89 ± 4.5	89 ± 6.3	94.87
CPA vs FLO-1	94 ± 0.04	94 ± 0.1	93 ± 0.4	94 ± 4.2	94 ± 4	97.33	90 ± 3.3	92 ± 5.7	88 ± 3.5	88 ± 4.2	92 ± 6.2	95.56
CPD vs FLO-1	93 ± 0.1	94 ± 0.4	92 ± 0.4	93 ± 4.5	94 ± 6.9	98.52	87 ± 3.1	91 ± 4.9	82 ± 5.3	84 ± 4	90 ± 4.4	93.55

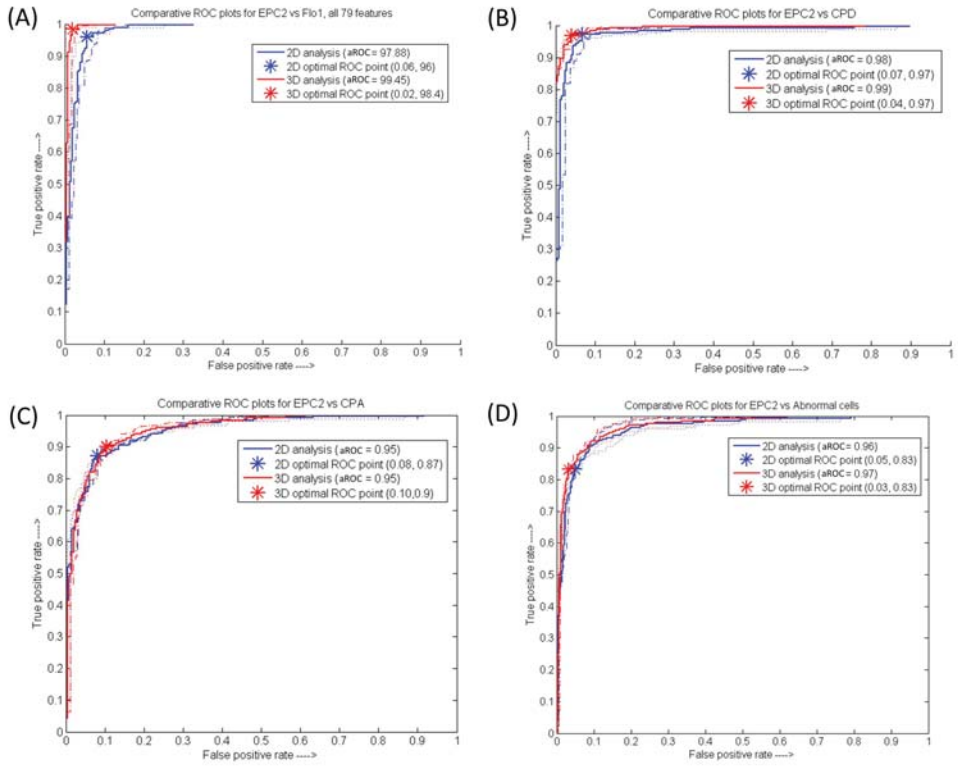

Figure 6.4: Comparative ROC plots of best classifier models between normal squamous and (A) cancer cells, (B) dysplastic cells, (C) metaplastic cells, and (D) metaplastic, dysplastic, and cancer cells using 3D (red) and 2D (blue) morphological data. The mean ROC points lie on the solid curve. The corresponding 95% confidence intervals are indicated by dotted lines. All 79 features were used to train the classifier model.

Feature selection using the mRMR method reduced the training time required to derive best classifier models for all 3D analysis and a majority of the 2D analysis. A concise summary in Table 6.3 indicates that the reduction in features was more appreciable for the 3D features in most of the comparisons. With the exception of the normal-metaplasia and normal-dysplasia comparisons, morphological analysis using 3D features required lesser number of features than analysis on equivalent 2D features. Feature selection did not improve classification with 2D features for comparisons among metaplastic, dysplastic, and cancer cells. Table 6.4 details the classifier performance for each comparison after feature selection. ROC plots for classifier models derived after feature selection are shown in Figures 6.6 and 6.7. These results show that 3D morphological analysis continued to

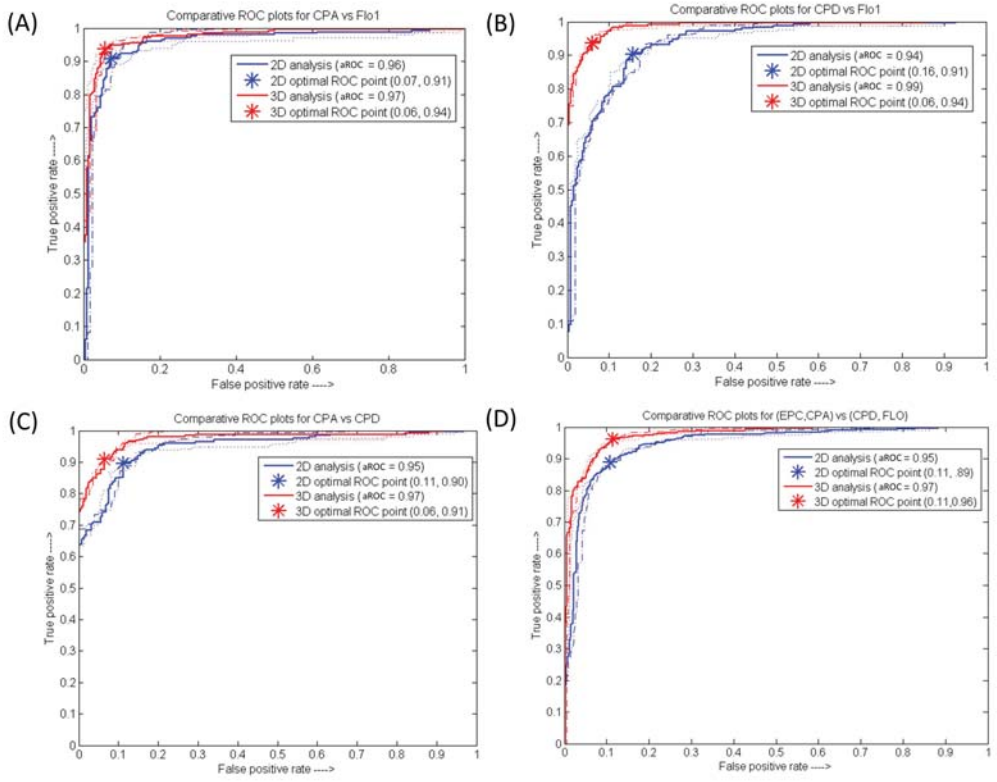

Figure 6.5: Comparative ROC plots of best classifier models between (A) metaplastic Barrett's and cancer cells, (B) dysplastic Barrett's and cancer cells, (C) metaplastic and dysplastic cells, and (d) normal and metaplastic cells versus dysplastic and cancer cells. Red curve represents analysis using 3D morphological data and blue curve indicates analysis with 2D data. The mean ROC points lie on the solid curve. The corresponding 95% confidence intervals are indicated by dotted lines. All 79 features were used to train the classifier.

outperform 2D analysis even with a reduced feature set. The 95% CIs for the ROC plots associated with 3D analysis remained relatively tighter than the corresponding 2D plots.

Table 6.3: Effect of applying mRMR feature selection

Comparison	Number of features required for best performance	
	3D feature data	2D feature data
EPC vs FLO-1	5	60
EPC vs CPD	65	45
EPC vs CPA	50	50
EPC vs Abnormal	45	55
(EPC,CPA) vs (CPD,FLO-1)	60	65
CPA vs FLO-1	50	75
CPA vs CPD	45	79
CPD vs FLO-1	40	79

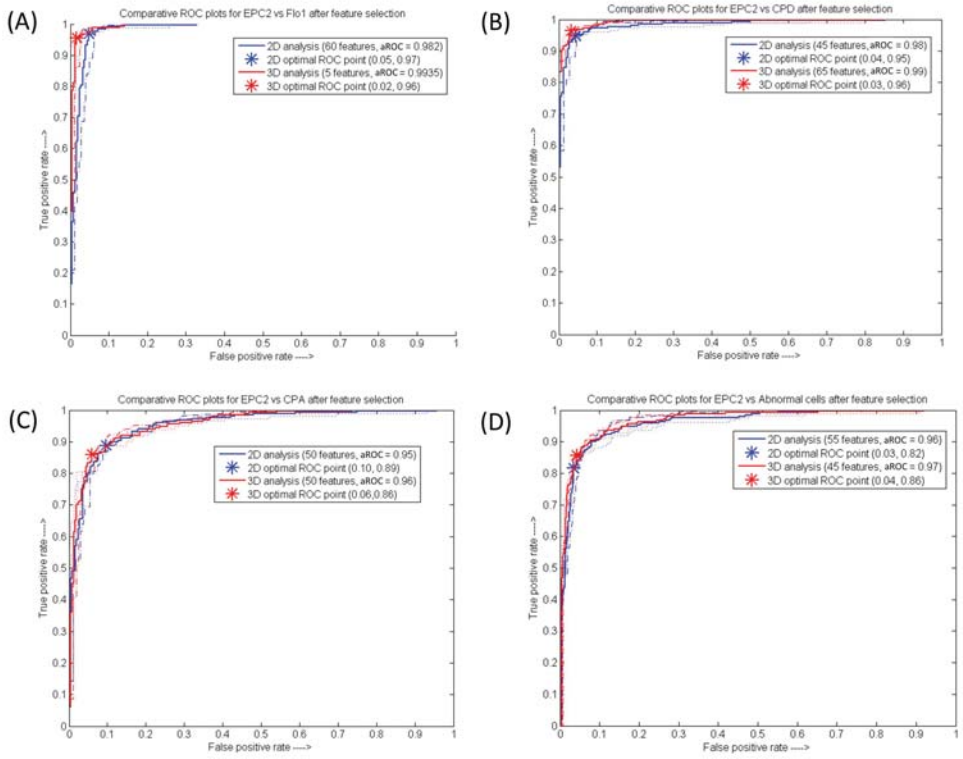

Figure 6.6: Comparative ROC plots of best classifier models after feature selection. Plots comparing 3D analysis (red) and 2D analysis (blue) are shown between normal squamous and (A) cancer cells, (B) dysplastic cells, (C) metaplastic cells, and (D) metaplastic, dysplastic, and cancer cells. The mean ROC points lie on the solid curve. Corresponding 95% confidence intervals are indicated by dotted lines.

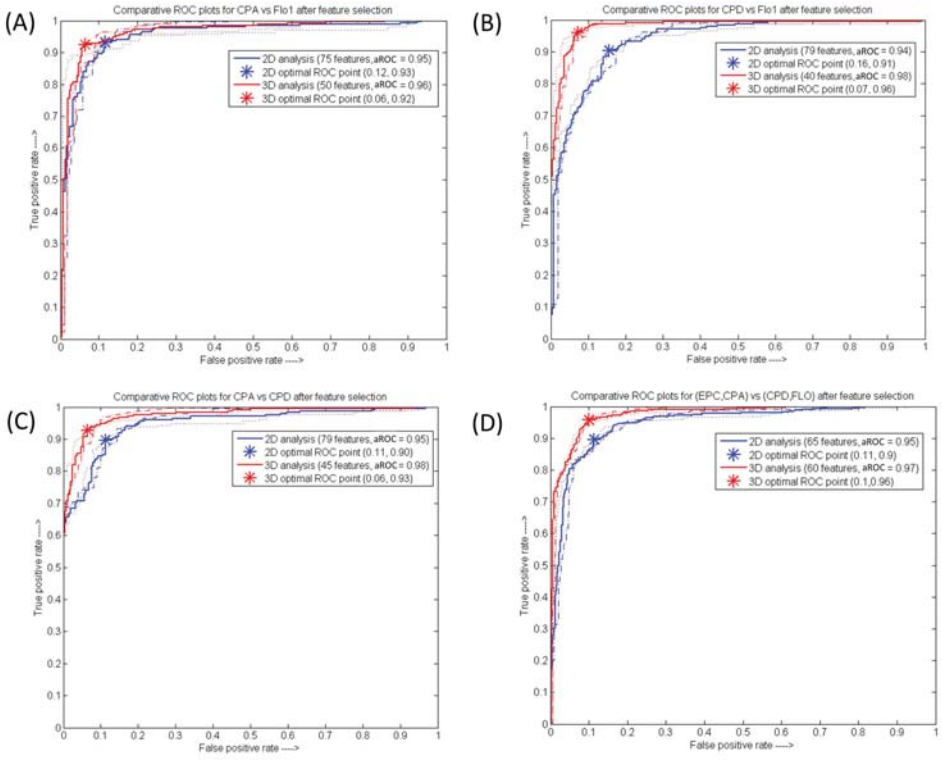

Figure 6.7: Comparative ROC plots of best classifier models after feature selection. Plots comparing 3D analysis (red) and 2D analysis (blue) are shown between (A) metaplastic Barrett's and cancer cells, (B) dysplastic Barrett's and cancer cells, (C) metaplastic and dysplastic cells, and (d) normal and metaplastic cells versus dysplastic and cancer cells. The mean ROC points lie on the solid curve. Corresponding 95% confidence intervals are indicated by dotted lines.

Table 6.4: Comparative performance summary of best classifier models after feature selection, 250 datasets per class, 10-fold crossvalidation

Comparison	3D feature analysis						2D feature analysis					
	Performance metrics(%)						Performance metrics (%)					
	Acc	Sens	Spec	PPV	NPV	aROC	Acc	Sens	Spec	PPV	NPV	aROC
EPC vs FLO-1	97 ± 0.03	97 ± 0.1	96 ± 0.1	96 ± 2.7	97 ± 2.9	99.35	95 ± 3	98 ± 2	93 ± 5	93 ± 5.1	98 ± 2	98.2
EPC vs CPD	97 ± 1.3	96 ± 2	97 ± 3	97 ± 3.2	93 ± 2.8	99.14	95 ± 3.4	95 ± 4	95 ± 5.6	94 ± 5.4	95 ± 4.9	98.38
EPC vs CPA	89 ± 5	90 ± 7.9	88 ± 8.8	88 ± 4	88 ± 9	95.33	90 ± 3.3	89 ± 4.1	90 ± 4.6	90 ± 4.9	89 ± 6	95.18
EPC vs Abnormal	93 ± 2.7	86 ± 8.5	96 ± 3.3	87 ± 9	95 ± 3.2	96.54	93 ± 3.1	82 ± 9.4	96 ± 2.5	88 ± 8.9	94 ± 3.5	95.98
(EPC,CPA) vs (CPD,FLO-1)	93 ± 1.5	93 ± 3	92 ± 2.9	92 ± 3	93 ± 2.8	97.4	89 ± 1.8	90 ± 3.6	88 ± 4.4	88 ± 4.7	90 ± 3.6	94.62
CPA vs CPD	93 ± 4.8	93 ± 5.6	93 ± 5.7	92 ± 6.6	93 ± 4.6	97.5	89 ± 3.9	90 ± 5.8	89 ± 5.6	89 ± 4.5	89 ± 8.2	94.87
CPA vs FLO-1	94 ± 0.1	92 ± 0.4	94 ± 0.3	94 ± 5.9	92 ± 6	96.45	91 ± 2.8	93 ± 5.4	88 ± 4.2	89 ± 4.3	93 ± 6	95.24
CPD vs FLO-1	94 ± 0.1	95 ± 0.2	93 ± 0.4	94 ± 4.6	95 ± 4.3	98.04	87 ± 3.1	91 ± 4.9	82 ± 5.3	84 ± 4	90 ± 4.4	93.55

The results of feature selection provide numerous insights on the relevance of morphological features for each comparison. Table 6.5 lists the set of five highest ranked features obtained with 2D and 3D analysis for every comparison. The results show the differences between features across comparisons. The table also indicates that the most powerful 2D features do not always translate to 3D. Another important outcome of the feature selection process was the consistent high ranking of Markovian texture descriptors (reflective of DNA packing and spatial localization) for all comparisons. Texture descriptors performed better than most of the morphological parameters used for conventional diagnosis. Similar trends on relevance of texture features were obtained with 2D analysis even though the variations were not readily apparent to the eye.

Analysis of sample size on classifier performance

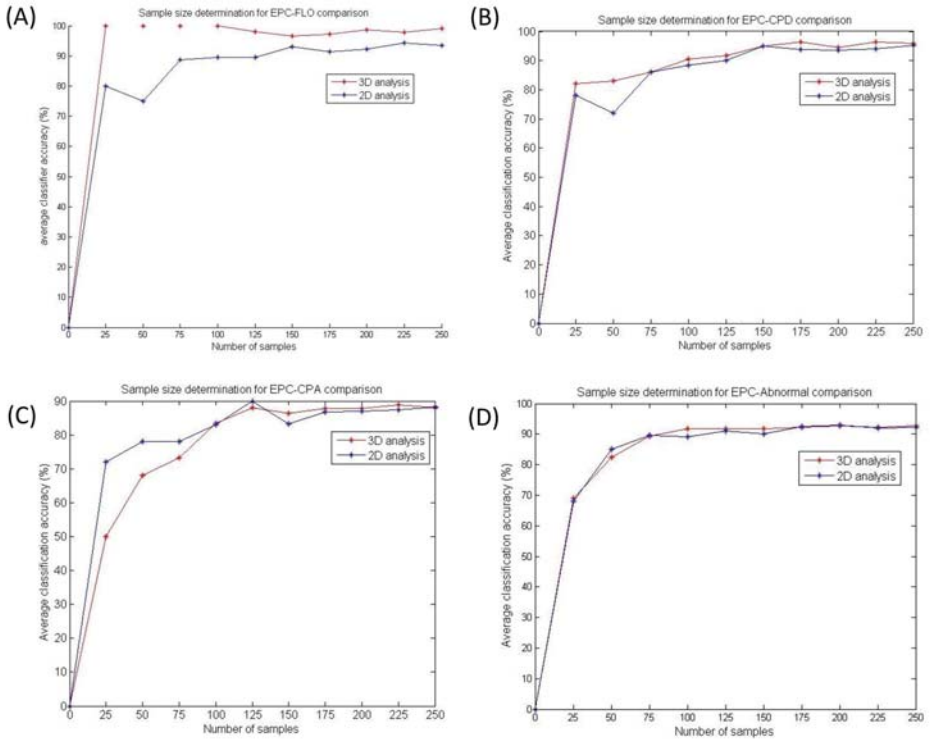

Figure 6.8: Comparative plots showing the effect of sample size on performance of best classifier models using 3D features (red curve) and 2D features (blue curve). Plots are shown comparing normal squamous and (A) cancer cells, (B) dysplastic cells, (C) metaplastic cells, and (D) metaplastic, dysplastic, and cancer cells.

Table 6.5: A comparison of the highest ranked feature to distinguish cell types on the basis of 2D and 3D analysis

Comparison	Most powerful feature	
	3D feature analysis	2D feature analysis
EPC vs FLO-1	Total volume of UHD clumps, Markovian texture at 1μm intervals	Total area of UHD clumps, Markovian texture at 1 μm intervals
EPC vs CPD	Integrated nuclear content (volume), Markovian texture at 1μm intervals	Integrated nuclear content (area), Markovian texture at 1μm and 0.5 μm intervals
EPC vs CPA	volume of high density regions in nucleus	Integrated nuclear content (area)
EPC vs Abnormal	volume of high density regions in nucleus, Markovian texture at 1 μm intervals	Nuclear area, Markovian texture at 1 μm intervals
(EPC,CPA) vs (CPD,FLO-1)	NC ratio (volume), Markovian texture at 1 μm intervals	NC ratio (area), Markovian texture at 1 μm intervals
CPA vs FLO-1	NC ratio (volume), Markovian texture at 1 and 0.5 μm intervals	NC ratio (area), Markovian texture at 1 μm intervals
CPA vs CPD	NC ratio (volume), Markovian texture at 1 and 0.5 μm intervals	NC ratio (area), Markovian texture at 1 and 0.5 μm intervals
CPD vs FLO-1	Total volume of UHD clumps, Markovian texture at 1 μm intervals	Nuclear content per pixel, Markovian texture at 0.5 μm intervals

Figures 6.8 and 6.9 illustrate the effect of sample size on classifier performance for each of the 8 comparisons. The plots show that the approximate saturation in classifier performance corresponds to roughly 200 cells of each cell type. The normal-cancer comparison (Figure 6.8A) expectedly required much fewer samples (approximately 50) to reach maximum performance. With the exception of the normal-metaplasia comparison, analysis with 3D features required relatively fewer samples for a better classification than the equivalent 2D analysis.

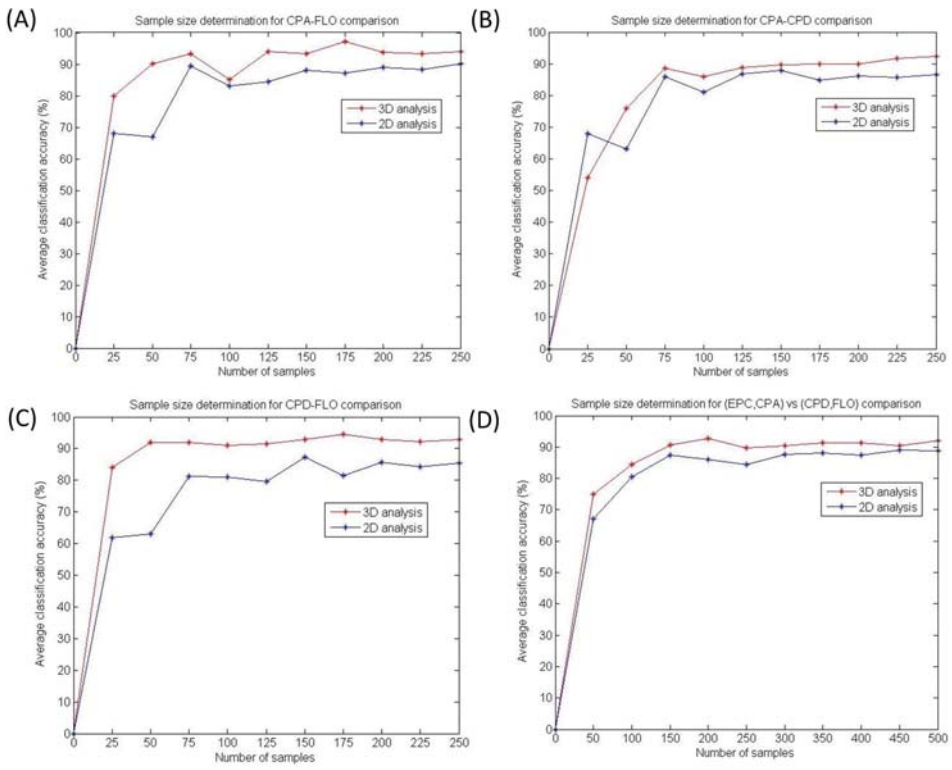

Figure 6.9: Comparative plots showing the effect of sample size on performance of best classifier models using 3D features (red curve) and 2D features (blue curve). Plots are shown comparing (A) metaplastic Barrett's and cancer cells, (B) dysplastic Barrett's and cancer cells, (C) metaplastic and dysplastic cells, and (D) normal and metaplastic cells versus dysplastic and cancer cells.

Choice of optimal ROC operating point

Practical considerations associated with clinical relevance drive the selection of the optimal ROC operating point on the ROC plot. A diagnostic test usually maximizes sensitivity while a screening test maximizes specificity. We discuss this subject using the comparison: non-metaplastic cells vs abnormal cells (dysplastic or cancer). Figure 6.10 illustrates the variation in optimal ROC operating points depending on whether the requirement is for the best possible diagnosis (Fig. 6.10A) or the best possible screening (Fig. 6.10B). For both cases, it is clear that 3D morphometric analysis performs significantly better than equivalent 2D analysis. Figure 6.10 underscores the importance of 3D morphometric

analysis over 2D analysis. Figure 6.10A shows that a diagnostic test can achieve a sensitivity in 3D of 98% with a 22% false positive rate, while the same sensitivity in 2D is compromised with a 47% false positive rate. In other words, at the same high sensitivity, the 2D approach more than doubles the number of false positives compared to the 3D approach. Likewise, Figure 6.10B shows that a screening test can achieve a specificity in 3D of 98% with a sensitivity of 80%, while the same specificity in 2D is compromised with a lower sensitivity of only 42%. In screening, the 2D approach would triple the number of false negatives. Thus, 3D analysis is vastly superior to 2D analysis in either diagnostic applications or screening applications.

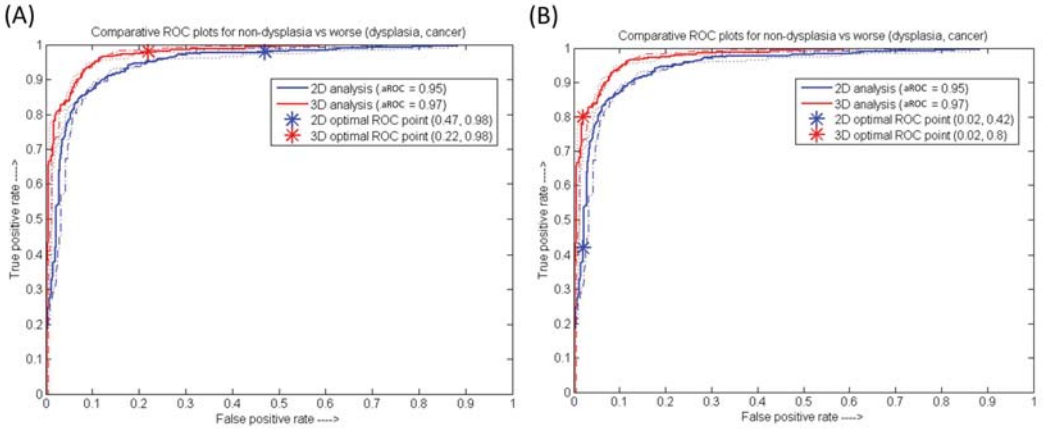

Figure 6.10: Comparative ROC plots illustrating the dependence of ROC operating point on the application at hand. (A) Diagnostic test that maximizes sensitivity, and (B) Screening test which maximizes specificity

6.4 Conclusions and discussion

Quantitative three-dimensional morphological biosignatures have been defined for an *in vitro* model of esophageal adenocarcinoma progression using automated 3D image analysis of single-cell optical CT images. The results of 3D morphological analysis were compared against equivalent 2D analysis of 2D images which simulated the depth of field of a 10x objective lens. The results of this study demonstrated that 3D morphological analysis of optical CT images provides significantly more sensitive and specific signatures to distinguish cell types than equivalent 2D analysis. Statistical analysis of the

morphological data revealed that 3D analysis required fewer morphological descriptors to achieve a similar or better performance than the equivalent 2D analysis. The ROC plots for 3D analysis also consistently exhibited relatively tighter 95% confidence intervals, indicating low variance in performance. Errors in 2D classification did not always correspond to errors with 3D classification i.e. the set of cells misclassified by 3D classifiers was not always a subset of misclassifications produced by 2D classifiers. Further, sample size estimation analysis revealed that fairly high classification performance required 175 to 200 cells per type and it saturated at that sample size.

This study revealed numerous novel insights about the morphologies of normal squamous, metaplastic, dysplastic, and cancer cells and their variation with progression. Marked intra-population morphological heterogeneity was observed in all cell types, and it increased from normal to cancer. Cancer cells could be clearly distinguished from normal squamous, metaplastic, and dysplastic cells. 3D analysis also revealed the higher relevance of intra-nuclear texture and the total volume of high density intra-nuclear regions instead of parameters that are conventionally known to have diagnostic relevance (for ex. nuclear size, NC ratio, nuclear polarity). In contrast to previous reports based on 2D analysis, nuclear pleomorphism was not found to be significantly altered in progression. However, the nucleus to cytoplasm volume ratio was found to be a powerful feature to distinguish metaplasia from dysplasia and cancer. The occasional difference between the most powerful 2D and 3D features for a given comparison is another result worthy of translational relevance. The relatively weak classification performance between normal squamous and metaplastic Barrett's cells could be attributed to the immortalization process. This postulate has been strengthened by recent observations in dissociated patient biopsies where normal squamous and metaplastic Barrett's cells show clearly different morphologies.

Although the results of this study were derived from immortalized cell lines, the marked distinctions in 3D morphometrics between precancerous stages (metaplasia, dysplasia) and adenocarcinoma and their improvements over equivalent 2D analysis show

promise for the application of optical CT imaging in clinical settings. For this study, an analytical approximation to 2D images acquired using a 10x objective lens in widefield mode was synthesized from the reconstructed 3D images. The good performance observed with 2D analysis could be attributed to the fact that the approximated 2D image is still better than a conventional 2D image acquired with a 10x lens. 3D analysis predominantly required a relatively fewer number of cells and features than 2D analysis to achieve an equivalent or better classification performance.

The experimental results demonstrate the potential of optical CT imaging and quantitative 3D image analysis for precise 3D cytopathological evaluation of cells. This approach could potentially serve as an adjunctive tool for 3D cytopathological assessment of cells for diagnosis or screening purposes. Multiple prior studies have been performed to compare the utility of 2D automated cytology relative to standard clinical cytopathology practice [80, 81], and have reported an improvement in reproducibility, reduction in assessment times, and relatively better sensitivity. A similar study could be pursued using optical CT technology. VisionGate Inc. – the manufacturer of the Cell-CT^{TM} – is already in the process of evaluating the efficacy of automated 3D cytopathology relative to standard cytopathology using lung cancer samples. A similar study could be performed for the diagnosis of Barrett's dysplasia.

Chapter 7

DISSERTATION SUMMARY AND FUTURE DIRECTIONS

7.1 Dissertation summary

Accurate quantitative characterization of the three-dimensional structure of cells and their nuclei provides crucial supplementary information for accurate detection of malignancy. Optical CT is a breakthrough imaging technique that generates individual cell images with an isotropic spatial resolution of 350 nm by performing computed tomography with white light excitation. The isotropy in spatial resolution obtained using optical CT enables a precise quantitative characterization of morphology that cannot be achieved by conventional microscopy modalities that generate 3D images by serially stacking 2D optical sections. The Cell-CT^{TM} optical CT platform enables a precise characterization of cellular morphology and its heterogeneity. Optical CT images pose a challenge for automated analysis given the unique imaging physics underlying their formation. This dissertation provided a modular computational framework to automatically generate sensitive and specific 3D morphological biosignatures indicative of cell health by volumetric image analysis of hematoxylin-stained cells imaged with the Cell-CT^{TM}. The dissertation research additionally optimized sample preparation, image acquisition, and data archival, and thus established a comprehensive foundation to facilitate the discovery of morphological biosignatures for cancer.

7.2 Primary dissertation contributions

1. A modular, automated computational framework has been designed, implemented, and optimized to perform high-throughput, 3D morphological analysis of volumetric images of transmission-mode Cell-CT^{TM}images. The capabilities of the framework can be seamlessly extended without affecting the current setup. The framework also provides automatic archival of results for future reference. Thus, it facilitates ease of use by biologists or any other users who may not be familiar with software programming or technical details related to computer vision or statistical analysis.

2. Custom, fully-automated 3D image processing methods have been developed to accurately delineate the cell, the nucleus and ultra-high density clumps within the nucleus. Prior to segmentation, a combination of anisotropic diffusion and power law transformation was applied to enhance the VOIs in the image. Adaptive intensity thresholding schemes then sufficed for the cell and UHD clumps. A hybrid method composed of a contrast-based heuristic followed by automated 2D GVF snake propagation was proposed to segment the nucleus. The efficacy of proposed segmentation techniques has been validated relative to manually traced ground truth volumes and achieved an average accuracy of 94$\%$ relative to user-defined ground truth. Methods were also developed to compute 79 biologically relevant morphological and texture parameters from the segmented VOIs.

   The developed framework allows a direct comparison of 3D morphological data with its 2D equivalent by computing the same 79 morphological features on a 2D slice that simulates a digital image acquired with a 10x objective lens.

3. The proposed image processing sequence was applied in conjunction with data mining techniques to investigate the existence and relative merit of a robust 3D morphological biosignature over known 2D signatures to distinguish normal squamous, metaplastic, dysplastic, and esophageal adenocarcinoma epithelial cells from each other. The data mining suites contains methods for supervised and

unsupervised learning, feature selection, and numerous metrics such as sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), ROC plots, and area under the ROC curve (AUC) to evaluate the performance of classifier models. The results from this research indicated that 3D biosignatures indeed exist to distinguish every combination of cell type in the progression model. As evidenced from the ROC plots and their tight 95% confidence intervals, adenocarcinoma cells could be distinctly distinguished from normal squamous, metaplastic and dysplastic cells. 3D morphological analysis consistently provided better classifier performance than the equivalent 2D analysis of the same cells. Further, 3D analysis required fewer features to yield a better classification performance. The relatively weak classification performance between normal squamous and metaplastic Barrett's cells and the distinct variations between the two cell types in patient samples reinforced the limitations of immortalized cell lines.

4. 3D analysis of optical CT images revealed several novel insights on phenotypic heterogeneity and progression in Barrett's Esophagus. Markovian texture features were found to have markedly high discriminative power. The most powerful 3D features did not always correspond to the most powerful 2D counterparts. While some conventional diagnostic parameters such as nucleus to cytoplasm ratio were found to be effective to distinguish metaplastic from dysplastic cells, others such as nuclear pleomorphism were not found to be significantly altered with progression in Barrett's Esophagus.

7.3 Potential avenues for future research

The groundwork provided by this dissertation research enables numerous interesting extensions to enhance its utility for automated biosignature discovery. A subset of future possibilities are discussed according to their thrust (technology or biology).

Potential technological advancements

1. Assessment of the effect of the point spread function (PSF) on the accuracy of feature measurement: The smallest measurable morphological feature in a micrograph directly depends on the size of the optical system's PSF. Morphological features smaller than the PSF would be blurred to approximately the size of the PSF by the process of convolution. Thus, the size and shape of the PSF dictate the measurement accuracy of a morphological feature. Tomographic reconstruction from projections produces a spherical (3D Gaussian) PSF. However, the PSF stemming from the projection image acquisition also needs to be factored in along with the reconstruction PSF. Conventional volumetric microscopies possess a non-spherical PSF which is elongated along the z-axis relative to the x- and the y- axes. Consequently, an orientation bias or dependence is introduced in the feature measurement, especially if the feature has a prominent z-axis component. Imaging techniques with a spherical PSF (as in the case of optical CT) would preclude any orientation dependence and thus result in more accurate measurement.

   Hence, a useful future study would be to compare the effects of spherical and non-spherical PSFs on the measurement accuracy of a feature as a function of its size and shape. Experiments using sub-micron order phantoms would further help validate the utility of optical CT imaging.

2. Comparison between morphological analysis using optical CT data and virtual 3D microscopy: In addition to 2D whole slide imaging used in digital pathology platforms, virtual 3D widefield microscopy [82] is another avenue that is being explored to leverage three-dimensional information about specimens used for pathological

inspection. Virtual 3D microscopy involves the digital acquisition followed by manual inspection of a select number of 2D widefield focal planes over the specimen volume. The process captures the pathologists' practice of focusing through the specimen to get a better understanding of morphological variations. Although virtual 3D microscopy of cells may provide more information than what is available in a single focal plane, it is still a pseudo-3D method (since the 3D image is synthesized by the person), and is thus far less informative than the corresponding optical CT image. An interesting study to validate this assertion would be to quantitatively compare the performance of 3D morphological analysis on the same cells by using optical CT data and virtual 3D microscopy data. Image data simulating virtual 3D microscopy can be easily acquired by the Cell-CT^{TM}since it provides the ability to acquire digital widefield 2D focal planes through the cellular volume.

3. Improvements to image analysis capabilities: Image analysis toolkit development for quantitative characterization of cellular phenotype is a subject of intensive investigation in the biomedical research community. Our computational framework currently contains the classical constituents necessary for volumetric image analysis. Its capabilities could be augmented by the addition of innovative analytical methods germane to morphological biosignature discovery. Alternative segmentation methods based on fast evolving 3D deformable models could be explored to segment the nucleus and ultra-high density clumps within the nucleus. Although this avenue is promising, it requires careful tailoring to be effective for optical CT data. A few other exemplar options are implicit methods [83] to detect the width of the nuclear envelope, spherical harmonic analysis for nuclear shape characterization [84], transport theory to quantify alterations in distribution of chromatin [85], and methods based on stability selection [86] to select morphological features with best discriminatory power. A pilot study on quantifying 3D nuclear shape [87] by spherical harmonic analysis has shown promising results. A short-term future effort will be to integrate this methodology into the computational framework reported here.

A related future work would be to extend current image processing capabilities to analyze images generated by fluorescence-mode optical CT. Fluorescence image data could pose new challenges due to their low signal to noise.

4. Streamlined and user-friendly automation software: A graphical user interface (GUI)-driven and user-friendly software tool would greatly facilitate the utility of the developed computational framework among non-engineers. A first step in this direction is to integrate the computational tools developed in this dissertation with data visualization tools being developed in our laboratory.

Interesting future biological investigations

1. Evaluation of consistency in biosignatures between immortalized cell lines and patient samples: This dissertation provided one exemplar application of 3D biosignature discovery using immortalized cell lines representative of progression in esophageal adenocarcinoma. An important research question commonly encountered in cancer biology is the real-world relevance of quantitative models derived from immortalized cells grown in monolayer. A related ongoing effort in our laboratory is the culture of primary esophageal epithelial cells derived from patient samples. The primary cells span the same progression spectrum as our immortalized model and would thus ideally facilitate objective investigations on the subject of consistency between immortalized cells and patient-derived primary cells.

2. Organ specific 3D morphological biosignatures: An interesting future project would be to leverage the proposed computational framework to investigate organ-specificity in 3D morphological biosignatures. A related pursuit is to probe the morphological differences between cancer cells in the primary tumor site and cells that have metastasized to another location. A third line of inquiry would be to identify the morphological correlates of a cancer's molecular classification. The outcomes of this research could ultimately have a far-reaching impact on the accuracy of diagnostic

pathology. Biological data in the form of normal-cancer cell line pairs for breast, pancreas, and colon epithelium is available to initiate this investigation.

3. Quantifying 3D cellular architecture in tissue: The current version of the Cell-CT^{TM} is tailored to perform optical CT on individual cells. A useful future avenue would be modify the technology to facilitate 3D imaging of thicker specimen such as clusters of cells or small pieces of tissue. This additional capability could enable isotropically resolved, micron scale volumetric imaging of entities that possess high biological and translational relevance. Tissue architecture is an important criterion for cancer diagnosis. Understanding the three-dimensional organization of cells in micrometastases or circulating tumor cells (CTCs) may provide vital clues about their function. Additionally, organoid models grown in 3D culture [88] have been shown to mimic *in vivo* conditions much better than immortalized cell lines grown in monolayer.

REFERENCES

[1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2012," *CA: a cancer journal for clinicians*, 2012.

[2] A. C. Society, "Cancer: Facts & figures 2012," 2012.

[3] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell*, vol. 100, no. 1, pp. 57–70, Jan 7 2000.

[4] D. Hanahan and R. Weinberg, "Hallmarks of cancer: the next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011.

[5] W. G. Feero, A. E. Guttmacher, W. G. Feero, A. E. Guttmacher, and F. S. Collins, "Genomic medicine:an updated primer," *New England Journal of Medicine*, vol. 362, no. 21, pp. 2001–2011, 2010.

[6] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler, "Cancer genome landscapes," *Science*, vol. 339, no. 6127, pp. 1546–1558, 2013.

[7] D. A. Haber, N. S. Gray, and J. Baselga, "The evolving war on cancer," *Cell*, vol. 145, no. 1, pp. 19–24, 2011.

[8] T. A. Yap and P. Workman, "Exploiting the cancer genome: strategies for the discovery and clinical development of targeted molecular therapeutics," *Annual review of pharmacology and toxicology*, vol. 52, pp. 549–573, 2012.

[9] B. Tran, J. E. Dancey, S. Kamel-Reid, J. D. McPherson, P. L. Bedard, A. M. Brown, T. Zhang, P. Shaw, N. Onetto, L. Stein, *et al.*, "Cancer genomics: technology, discovery, and translation," *Journal of Clinical Oncology*, vol. 30, no. 6, pp. 647–660, 2012.

[10] T. Fojo and D. R. Parkinson, "Biologically targeted cancer therapy and marginal benefits: are we making too much of too little or are we achieving too little by giving too much?" *Clinical Cancer Research*, vol. 16, no. 24, pp. 5972–5980, 2010.

[11] R. Bernards, "A missing link in genotype-directed cancer therapy," *Cell*, vol. 151, no. 3, pp. 465–468, 2012.

[12] P. M. LoRusso, R. Canetta, J. A. Wagner, E. P. Balogh, S. J. Nass, S. A. Boerner, and J. Hohneker, "Accelerating cancer therapy development: the importance of combination strategies and collaboration. summary of an institute of medicine workshop," *Clinical Cancer Research*, vol. 18, no. 22, pp. 6101–6109, 2012.

[13] M. Vanneman and G. Dranoff, “Combining immunotherapy and targeted therapies in cancer treatment,” *Nature Reviews Cancer*, vol. 12, no. 4, pp. 237–251, 2012.

[14] I. Bozic, B. Allen, and M. A. Nowak, “Dynamics of targeted cancer therapy,” *Trends in molecular medicine*, 2012.

[15] J. De Bono and A. Ashworth, “Translating cancer research into targeted therapeutics,” *Nature*, vol. 467, no. 7315, pp. 543–549, 2010.

[16] L. Chin, J. N. Andersen, and P. A. Futreal, “Cancer genomics: from discovery science to personalized medicine,” *Nature medicine*, vol. 17, no. 3, pp. 297–303, 2011.

[17] D. G. de Castro, P. A. Clarke, B. Al-Lazikani, and P. Workman, “Personalized cancer medicine: Molecular diagnostics, predictive biomarkers, and drug resistance,” *Clinical Pharmacology & Therapeutics*, vol. 93, no. 3, pp. 252–259, 2012.

[18] J. Nickerson, “Nuclear dreams: the malignant alteration of nuclear architecture,” *Journal of cellular biochemistry*, vol. 70, no. 2, pp. 172–180, 1998.

[19] D. Zink, A. Fischer, and J. Nickerson, “Nuclear structure in cancer cells,” *Nature Reviews Cancer*, vol. 4, pp. 677–687, 2004.

[20] L. True and C. Jordan, “The cancer nuclear microenvironment: interface between light microscopic cytology and molecular phenotype,” *J Cell Biochem*, vol. 104, no. 6, pp. 1994–2003, Aug 15 2008, true, Lawrence D Jordan, C Diana Review United States Journal of cellular biochemistry J Cell Biochem. 2008 Aug 15;104(6):1994-2003.

[21] I. Damjanov and F. Fan, *Cancer grading manual*. Springer, 2006.

[22] W. H. Organization, “Cancer pathology and genetics,” http://www.iarc.fr/en/publications/pdfs-online/pat-gen/, 2013.

[23] M. F. Kircher, H. Hricak, and S. M. Larson, “Molecular imaging for personalized cancer care,” *Molecular Oncology*, 2012.

[24] L. V. Wang and S. Hu, “Photoacoustic tomography: in vivo imaging from organelles to organs,” *Science*, vol. 335, no. 6075, pp. 1458–1462, 2012.

[25] H. Subramanian, P. Pradhan, Y. Liu, I. R. Capoglu, X. Li, J. D. Rogers, A. Heifetz, D. Kunte, H. K. Roy, A. Taflove, *et al.*, “Optical methodology for detecting histologically unapparent nanoscale consequences of genetic alterations in biological cells,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 51, pp. 20 118–20 123, 2008.

[26] L. Fass, “Imaging and cancer: a review,” *Molecular oncology*, vol. 2, no. 2, pp. 115–152, 2008.

[27] K. Eliceiri, M. Berthold, I. Goldberg, L. Ibáñez, B. Manjunath, M. Martone, R. Murphy, H. Peng, A. Plant, B. Roysam, *et al.*, “Biological imaging software tools,” *Nature Methods*, vol. 9, no. 7, pp. 697–710, 2012.

[28] E. Kotter and M. Langer, “Computer aided detection and diagnosis in radiology,” *European radiology*, vol. 21, no. 3, pp. 590–592, 2011.

[29] N. V. Orlov, W. W. Chen, D. M. Eckley, T. J. Macura, L. Shamir, E. S. Jaffe, and I. G. Goldberg, “Automatic classification of lymphoma images with transform-based global features,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 4, pp. 1003–1013, 2010.

[30] N. V. Orlov, A. T. Weeraratna, S. M. Hewitt, C. E. Coletta, J. D. Delaney, D. Mark Eckley, L. Shamir, and I. G. Goldberg, “Automatic detection of melanoma progression by histological analysis of secondary sites,” *Cytometry Part A*, vol. 81, no. 5, pp. 364–373, 2012.

[31] H. Chang, G. V. Fontenay, J. Han, G. Cong, F. L. Baehner, J. W. Gray, P. T. Spellman, and B. Parvin, “Morphometic analysis of tcga glioblastoma multiforme,” *BMC bioinformatics*, vol. 12, no. 1, p. 484, 2011.

[32] M. Muthu Rama Krishnan, C. Chakraborty, R. R. Paul, and A. K. Ray, “Hybrid segmentation, characterization and classification of basal cell nuclei from histopathological images of normal oral mucosa and oral submucous fibrosis,” *Expert Systems with Applications*, vol. 39, no. 1, pp. 1062–1077, 2012.

[33] W. Wang, J. A. Ozolek, and G. K. Rohde, “Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images,” *Cytometry Part A*, vol. 77, no. 5, pp. 485–494, 2010.

[34] E. Sabo, A. H. Beck, E. A. Montgomery, B. Bhattacharya, P. Meitner, J. Y. Wang, and M. B. Resnick, “Computerized morphometry as an aid in determining the grade of dysplasia and progression to adenocarcinoma in barrett’s esophagus,” *Laboratory investigation*, vol. 86, no. 12, pp. 1261–1271, 2006.

[35] C.-W. Wang and C.-P. Yu, “Automated morphological classification of lung cancer subtypes using h&e tissue images,” *Machine Vision and Applications*, pp. 1–9, 2012.

[36] M. M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel, and M. N. Gurcan, “Computerized classification of intraductal breast lesions using histopathological

images," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 7, pp. 1977–1984, 2011.

[37] C. Atupelage, H. Nagahashi, M. Yamaguchi, T. Abe, A. Hashiguchi, and M. Sakamoto, "Computational grading of hepatocellular carcinoma using multifractal feature description," *Computerized Medical Imaging and Graphics*, 2012.

[38] Q. Chaudry, S. H. Raza, A. N. Young, and M. D. Wang, "Automated renal cell carcinoma subtype classification using morphological, textural and wavelets based features," *Journal of Signal Processing Systems*, vol. 55, no. 1-3, pp. 15–23, 2009.

[39] E. Ozdemir, C. Sokmensuer, and C. Gunduz-Demir, "A resampling-based markovian model for automated colon cancer diagnosis," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 1, pp. 281–289, 2012.

[40] A. E. Rizzardi, A. T. Johnson, R. I. Vogel, S. E. Pambuccian, J. Henriksen, A. P. Skubitz, G. J. Metzger, and S. C. Schmechel, "Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring," *Diagnostic Pathology*, vol. 7, no. 1, p. 42, 2012.

[41] S. Doyle, M. D. Feldman, N. Shih, J. Tomaszewski, and A. Madabhushi, "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC bioinformatics*, vol. 13, no. 1, p. 282, 2012.

[42] S. Y. Park, D. Sargent, R. Lieberman, and U. Gustafsson, "Domain-specific image analysis for cervical neoplasia detection based on conditional random fields," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 3, pp. 867–878, 2011.

[43] S. Al-Janabi, A. Huisman, and P. J. Van Diest, "Digital pathology: current status and future perspectives," *Histopathology*, vol. 61, no. 1, pp. 1–9, 2012.

[44] F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital imaging in pathology: Whole-slide imaging and beyond," *Annual Review of Pathology: Mechanisms of Disease*, vol. 8, pp. 331–359, 2013.

[45] A. Huisman, L. S. Ploeger, H. F. J. Dullens, T. N. Jonges, J. A. M. Belien, G. A. Meijer, N. Poulin, W. E. Grizzle, and P. J. van Diest, "Discrimination between benign and malignant prostate tissue using chromatin texture analysis in 3-d by confocal laser scanning microscopy," *The Prostate*, vol. 67, no. 3, pp. 248–254, 2007.

[46] T. Planchon, L. Gao, D. Milkie, M. Davidson, J. Galbraith, C. Galbraith, and E. Betzig, "Rapid three-dimensional isotropic imaging of living cells using bessel beam plane illumination," *Nature Methods*, vol. 8, no. 5, pp. 417–423, 2011.

[47] J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sørensen, R. Baldock, and D. Davidson, "Optical projection tomography as a tool for 3d microscopy and gene expression studies," *Science*, vol. 296, no. 5567, pp. 541–545, 2002.

[48] M. Le Gros, G. McDermott, and C. Larabell, "X-ray tomography of whole cells," *Current opinion in structural biology*, vol. 15, no. 5, pp. 593–600, 2005.

[49] M. Fauver, E. Seibel, J. Rahn, M. Meyer, F. Patten, T. Neumann, and A. Nelson, "Three-dimensional imaging of single isolated cell nuclei using optical projection tomography," *Optics Express*, vol. 13, no. 11, pp. 4210–4223, 2005, times Cited: 10.

[50] Q. Miao, J. Rahn, A. Tourovskaia, M. Meyer, T. Neumann, A. Nelson, and E. Seibel, "Dual-modal three-dimensional imaging of single cells with isometric high resolution using an optical projection tomography microscope," *Journal of biomedical optics*, vol. 14, no. 6, pp. 064 035–064 035, 2009.

[51] M. Meyer, M. Fauver, J. Rahn, T. Neumann, F. Patten, E. Seibel, and A. Nelson, "Automated cell analysis in 2d and 3d: A comparative study," *Pattern Recognition*, vol. 42, no. 1, pp. 141–146, 2009, times Cited: 0.

[52] Q. Miao, A. Reeves, F. Patten, and E. Seibel, "Multimodal 3d imaging of cells and tissue, bridging the gap between clinical and research microscopy," *Annals of biomedical engineering*, pp. 1–14, 2012.

[53] G. Li, T. Liu, A. Tarokh, J. Nie, L. Guo, A. Mara, S. Holley, and S. Wong, "3d cell nuclei segmentation based on gradient flow tracking," *BMC cell biology*, vol. 8, no. 1, p. 40, 2007.

[54] C. Zanella, M. Campana, B. Rizzi, C. Melani, G. Sanguinetti, P. Bourgine, K. Mikula, N. Peyriéras, and A. Sarti, "Cells segmentation from 3-d confocal images of early zebrafish embryogenesis," *Image Processing, IEEE Transactions on*, vol. 19, no. 3, pp. 770–781, 2010.

[55] R. Chinta and M. Wasser, "Three-dimensional segmentation of nuclei and mitotic chromosomes for the study of cell divisions in live drosophila embryos," *Cytometry Part A*, 2012.

[56] K. Nandy, P. Gudla, R. Amundsen, K. Meaburn, T. Misteli, and S. Lockett, "Automatic segmentation and supervised learning-based selection of nuclei in cancer tissue images," *Cytometry Part A*, 2012.

[57] A. Bartesaghi, G. Sapiro, and S. Subramaniam, “An energy-based three-dimensional segmentation approach for the quantitative interpretation of electron tomograms,” *Image Processing, IEEE Transactions on*, vol. 14, no. 9, pp. 1314–1323, 2005.

[58] H. Nguyen and Q. Ji, “Shape-driven three-dimensional watersnake segmentation of biological membranes in electron tomography,” *Medical Imaging, IEEE Transactions on*, vol. 27, no. 5, pp. 616–628, 2008.

[59] F. Moussavi, G. Heitz, F. Amat, L. Comolli, D. Koller, and M. Horowitz, “3d segmentation of cell boundaries from whole cell cryogenic electron tomography volumes,” *Journal of structural biology*, vol. 170, no. 1, pp. 134–145, 2010.

[60] A. Rigort, D. Günther, R. Hegerl, D. Baum, B. Weber, S. Prohaska, O. Medalia, W. Baumeister, and H. Hege, “Automated segmentation of electron tomograms for a quantitative description of actin filament networks,” *Journal of structural biology*, vol. 177, no. 1, pp. 135–144, 2012.

[61] M. Uchida, Y. Sun, G. McDermott, C. Knoechel, M. Le Gros, D. Parkinson, D. Drubin, and C. Larabell, “Quantitative analysis of yeast internal architecture using soft x-ray tomography,” *Yeast*, vol. 28, no. 3, pp. 227–236, 2011.

[62] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 7, pp. 629–639, 1990.

[63] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, “Nonlinear anisotropic filtering of mri data,” *Medical Imaging, IEEE Transactions on*, vol. 11, no. 2, pp. 221–232, 1992.

[64] P. Rosin, “Unimodal thresholding,” *Pattern Recognition*, vol. 34, no. 11, pp. 2083–2096, 2001.

[65] N. Otsu, “A thresholding selection method from gray-level histogram,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[66] P.-S. Liao, T.-S. Chen, and P.-C. Chung, “A fast algorithm for multilevel thresholding,” *Journal of information science and engineering*, vol. 17, no. 5, pp. 713–728, 2001.

[67] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[68] C. Xu and J. L. Prince, “Snakes, shapes, and gradient vector flow,” *Image Processing, IEEE Transactions on*, vol. 7, no. 3, pp. 359–369, 1998.

[69] D. W. Shattuck, G. Prasad, M. Mirza, K. L. Narr, and A. W. Toga, “Online resource for validation of brain segmentation methods,” *NeuroImage*, vol. 45, no. 2, p. 431, 2009.

[70] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, “Textural features for image classification,” *IEEE Transactions on systems, man and cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[71] Y. Saeys, I. Inza, and P. Larrañaga, “A review of feature selection techniques in bioinformatics,” *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[72] G. Edgren, H.-O. Adami, E. W. Vainio, and O. Nyrén, “A global assessment of the oesophageal adenocarcinoma epidemic,” *Gut*, 2012.

[73] B. J. Reid, X. Li, P. C. Galipeau, and T. L. Vaughan, “Barrett’s oesophagus and oesophageal adenocarcinoma: time for a new synthesis,” *Nature Reviews Cancer*, vol. 10, no. 2, pp. 87–101, 2010.

[74] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, “Global cancer statistics,” *CA: a cancer journal for clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[75] V. da Silva, J. Prolla, P. Sharma, R. Sampliner, D. Thompson, and P. Bartels, “Karyometry in barrett’s esophagus,” *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology*, vol. 23, no. 1, pp. 40–46, 2001.

[76] H. Harada, H. Nakagawa, K. Oyama, M. Takaoka, C. Andl, B. Jacobmeier, A. von Werder, G. Enders, O. Opitz, and A. Rustgi, “Telomerase induces immortalization of human esophageal keratinocytes without p16ink4a inactivation,” *Mol Cancer Res*, vol. 1, no. 10, pp. 729–38, 2003.

[77] M. C. Palanca-Wessels, M. T. Barrett, P. C. Galipeau, K. L. Rohrer, B. J. Reid, and P. S. Rabinovitch, “Genetic analysis of long-term barrett’s esophagus epithelial cultures exhibiting cytogenetic and ploidy abnormalities,” *Gastroenterology*, vol. 114, no. 2, pp. 295–304, 1998.

[78] J. J. Boonstra, R. van Marion, D. G. Beer, L. Lin, P. Chaves, C. Ribeiro, A. D. Pereira, L. Roque, S. J. Darnton, N. K. Altorki, *et al.*, “Verification and unmasking of widely used human esophageal adenocarcinoma cell lines,” *Journal of the National Cancer Institute*, vol. 102, no. 4, pp. 271–274, 2010.

[79] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.

[80] C. V. Biscotti, A. E. Dawson, B. Dziura, L. Galup, T. Darragh, A. Rahemtulla, and L. Wills-Frank, “Assisted primary screening using the automated thinprep imaging system,” *American journal of clinical pathology*, vol. 123, no. 2, pp. 281–287, 2005.

[81] D. Schledermann, T. Hyldebrandt, D. Ejersbo, and B. Hoelund, “Automated screening versus manual screening: A comparison of the thinprep® imaging system and manual screening in a time study,” *Diagnostic cytopathology*, vol. 35, no. 6, pp. 348–352, 2007.

[82] T. Kalinski, R. Zwönitzer, S. Sel, M. Evert, T. Guenther, H. Hofmann, J. Bernarding, and A. Roessner, “Virtual 3d microscopy using multiplane whole slide images in diagnostic pathology,” *American journal of clinical pathology*, vol. 130, no. 2, pp. 259–264, 2008.

[83] R. I. Saye and J. A. Sethian, “The voronoi implicit interface method for computing multiphase physics,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 49, pp. 19 498–19 503, 2011.

[84] C. Ducroz, J.-C. Olivo-Marin, and A. Dufour, “Characterization of cell shape and deformation in 3d using spherical harmonics,” in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE, 2012, pp. 848–851.

[85] W. Wang, J. A. Ozolek, D. Slepcev, A. B. Lee, C. Chen, and G. K. Rohde, “An optimal transportation approach for nuclear structure-based pathology,” *Medical Imaging, IEEE Transactions on*, vol. 30, no. 3, pp. 621–631, 2011.

[86] N. Meinshausen and P. Bühlmann, “Stability selection,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 4, pp. 417–473, 2010.

[87] V. Nandakumar, X. An, Y. Wang, R. Johnson, and D. Meldrum, “Conformal mapping of nuclei in 3d tomographic cell images to assess shape heterogeneity,” in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE, 2012, pp. 222–225.

[88] J. Debnath, S. K. Muthuswamy, and J. S. Brugge, “Morphogenesis and oncogenesis of mcf-10a mammary epithelial acini grown in three-dimensional basement membrane cultures,” *Methods*, vol. 30, no. 3, pp. 256–268, 2003.