

Spectral Image Cytometry for Circulating Tumor Cell Identification

by

Richard Ross Ang

B.A.Sc., The University of British Columbia, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Mechanical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2016

© Richard Ross Ang, 2016

Abstract

Circulating tumor cells (CTCs) are exceedingly rare cancer cells shed from tumors into the bloodstream, where they have the potential to invade other tissues to seed metastases. CTCs are difficult to isolate but their critical role in tumor metastasis, as well as their proven prognostic value has attracted tremendous interest in recent years. While many methods have been developed to isolate CTCs, a major bottleneck to their clinical application has been the precise identification and characterization of these cells, owing to their tremendous phenotypic heterogeneity. To address these formidable challenges, a number of microscopy techniques have been applied to gather large amounts of information about captured cells. However, these studies are currently limited by two major concerns: First, due to the phenotypic plasticity of tumor cells, there may be significant variability in the properties of CTCs as observed using microscopy. Second, if the CTCs are subjected to multi-parameter analysis, the high-content data may be too expansive to analyze with a reasonable amount of time and effort. In this thesis, I developed an efficient and customizable spectral image cytometry platform to collect multi-spectral data from immunofluorescence micrographs of cell samples enriched for CTCs in order to quickly and easily analyze this information to facilitate CTC identification and characterization. This work includes the development of software tools to convert microscopy data for processing, to segment the images into single cell images, to rank potential CTCs, and to provide a user interface for rapid augmented review. The performance of this software platform has been evaluated by analyzing multi-spectral fluorescence imaging data previously collected by our group from ten patients with castrate resistant prostate cancer, and then comparing the result to unassisted manual reviews performed by blinded reviewers. The final CTC identification counts closely matched manual analysis with a slight increase in verified CTC counts, which is likely a result of the comprehensive nature of the automated screening process. The average computation time is 4.5 minutes per sample, which is faster than the time required to acquire the imaging data, and thus allows operators to quickly review results between acquisitions.

Preface

I wrote the majority of this manuscript and conducted the majority of conceptualization and development work for the spectral image cytometry software tools described. Evaluation of the software as described in Section 5.2 were performed by Emily Park and Buffy Chen. Antibody selection and immunofluorescent staining protocol was optimized by Sunyoung Park. Microscopy setup and calibration were performed by myself and Sunyoung Park.

This thesis utilizes de-identified data from a human cancer patient study under the UBC Clinical Research Ethics Board with certificates H10-01243, H13-00870.

Finally, a version of this thesis is currently being prepared for publication. Assistance with the preparation of this thesis and corresponding manuscript has been provided by Drs. Hongshen Ma and Simon Duffy.

Table of Contents

Abstract.....	ii
Preface.....	iii
Table of Contents	iv
List of Tables	vii
List of Figures.....	viii
List of Abbreviations	xii
Acknowledgements	xiii
Dedication	xiv
Chapter 1: Introduction and Background.....	1
1.1 Introductions	1
1.2 CTC Identification Criteria	3
1.2.1 Fluorescence Labeling	3
1.2.2 Cell Morphology	6
1.3 Fluorescence Cytometry Methods	7
1.3.1 Fluorescence Microscopy	7
1.3.2 Flow Cytometry	8
1.3.3 Spectral Confocal Microscopy.....	9
1.4 Image Cytometry Platforms for CTC Identification.....	11
1.4.1 CellSearch	12
1.4.2 Amnis ImageStream.....	13
1.4.3 Epic Sciences	14

Chapter 2: Design of Software for Spectral Image Cytometry for CTC Identification.....	16
2.1 Image Processing Workflow	16
2.2 Ingest Processing	17
2.3 Image Analysis.....	19
2.3.1 False Color Composite.....	19
2.3.2 Spectral Pixel Analysis	21
2.4 Segmentation and Sub-Region Analysis.....	25
2.5 Automatic Ranking Algorithm	27
2.6 User Review Process.....	28
2.6.1 Simple User Review Process	29
2.6.2 Graphical User Interface for Human Review	30
Chapter 3: Experimental Process for CTC Identification	33
3.1 Quality Control and Calibration.....	33
3.2 Workflow for Manual CTC Identification from Spectral Images	35
3.2.1 Manual Review of Spectral Images	35
3.2.2 Manual Review of Standard IF Images	37
3.3 Workflow for Semi-Automated CTC Identification from Spectral Images	40
Chapter 4: Results and Evaluation	42
4.1 Sensitivity vs. Specificity.....	42
4.1.1 EpCAM vs. CD45 Ratio RoC	43
4.1.2 Cell Rank RoC	44
4.2 CTC Identification Results	45
4.3 Performance and Ease of Use	46
Chapter 5: Conclusion.....	49

5.1	Summary	49
5.2	Statement of Impact	49
5.3	Future Work	50
Bibliography		51
Appendices.....		56
Appendix A Evaluation of Alternative Spectral Analysis Software Packages.....		56
A.1	Hyperspectral Image Analysis Toolbox.....	57
A.2	PoissonNMF	58
A.3	Multispec.....	58
A.4	Gerbil	59
A.5	Scyven.....	59
A.6	ZEN.....	60
A.7	Summary of Software Tested.....	60

List of Tables

Table 2.1 List of output file types and their purpose.	29
Table A.1 Summary of the tested results. See the software comparison for additional details on why a software failed or did not run	61

List of Figures

Figure 1.1 Metastatic dissemination of tumor cells, exemplified by prostate cancer. Cells develop in the primary tumor, shown in blue, and acquire a motile phenotype to disseminate as CTCs in the bloodstream. These cells also develop invasive and stem-like characteristics that allow them to initiate a tumor at a secondary site.....	1
Figure 1.2 Idealized emission spectra for marker based detection of CTC/WBC samples generated using Excel. C16H15N5 (DAPI) dye stains nuclear DNA directly, CK antibody bound to Alexa 488 (A488) dye stains cytokeratin common to CTCs, CD45 bound to Allophycocyanin (APC), and EpCam bound to Alexa 594 (A594)	5
Figure 1.3 Emission Spectra with idealized filters overlaid to capture peak signals. In practice filters are not ideal and overlap between signals is increased.	8
Figure 1.4 High level Overview of a Flow Cytometer. Coherent laser sources illuminate a single cell which is acquired by single pixel sensors with filter plates to capture spectral slices.....	9
Figure 1.5 Splitting light from a scanned confocal system into a continuous spectrum allows for a high spectral content image to be captured. PMT units on the ends can be configured to capture a fixed slice of the spectrum from either end.	10
Figure 1.6 Visual representation of the imaging setup with a 7680x7680x26 spectral image cube. Peak channels are listed as well as the quality control bright field channel which is obtained from the transmitted light through the sample while all color channels are fluorescent emissions from the sample.	11
Figure 1.7 Cell Search images displayed are a color composite of DAPI/CK with an image series of CK, DAPI, CD45, and a spare channel [55](© 2014 Ignatiadis et al. adapted under, CC BY 2.0). Noticeable oversaturation can be seen as well as automatic exposure control which can lead to variations in perceived intensities.	12

Figure 1.8 Example cell images using image stream flow cytometer with decreased cellular definition. Image retrieved from [58] (© 2016 Marques et al., adapted under CC BY 4.0). Analysis is performed semi automatically via an image panel or 2 parameter scatter plot.	13
Figure 1.9 Epic Sciences Full Channel Analysis image retrieved from [31] (© 2015 Werner et al., adapted under CC BY 3.0). Images are displayed in an image panel to users for review with a semi-automatic selection process.....	15
Figure 2.1 Overall workflow process from input file to output report.	17
Figure 2.2 Ingest workflow to extra image data and image metadata.	17
Figure 2.3 Metadata extraction tool interface displaying current file thumbnail and a snippet of the raw metadata. Used to determine the actual experimental setup from recorded metadata within sample files.	18
Figure 2.4 Split tasks for generating human readable false color composite images as well as processing raw image data into a more usable set of selection masks for downstream sub-region analysis.....	19
Figure 2.5 Compressing hyperspectral stack into a single image for easier review and management.	20
Figure 2.6 Comparison between microscopy software ZEN and custom false composite with negative pixel highlighting (white = negative pixel) and control pixel prioritization (blue nucleus enlarged).	21
Figure 2.7 Spectral pixel testing done across the entire gigapixel image, small subset shown with pixel types listed. (26x26 pixel close up of cell from ZEN lambda view)	22
Figure 2.8 Idealized representation of the global per pixel pass/fail spectral thresholding process. In practice control points can be reduced to critical markers shown as blue dots at the precise known peak for the individual dyes. The noise floor is determined through calibration studies as well as the relative intensity specification of 60%. The upper exclusion area is dynamically scaled per pixel to test for CD45 contamination in comparison to the EpCAM peak channel.	23

Figure 2.9 Any of the spectral channels can be classified on a per pixel basis as being a possible positive, negative or nucleus pixel. By creating a composite binary image based on nucleus pixels with negative/positive pixels nearby allows for the detection and sorting of cells while rejecting the background and debris (positive/negative).	25
Figure 2.10 Application specific scoring function used to sort results seen in Figure 2.11	27
Figure 2.11 Summary graphic overview with false color composite mapped to cell regions with the entire sample represented in one simple graphic. Used as the navigation overview map for review.....	28
Figure 2.12 Review software with experiment specific markup overlaid added to highlight marker positions as well as major areas of the software.	31
Figure 3.1 Calibration of microscope settings using control samples and initial trial run real world samples leading to a finalized configuration with main experiment verified by injecting blinded negative healthy samples randomly into the patient sample stream to verify workflow process. 34	
Figure 3.2 Illustration of sample confluence with multiple layers of cells overlaid on each other.	35
Figure 3.3 Manual review requires user to up-close manually click and drag on every potential cell. The number of tiles a user has to review is dependent on their screen resolution and monitor being used which adds additional variables in the review process.	37
Figure 3.4 Idealized spectral graph showing how clear positive cells exhibit clearly identifiable peaks while filter based processing shows intensities in all channels with the peak location unknown.....	39
Figure 3.5 Abnormal negative cell with all signals present and a weak CD45 signal. Spectral scanning shows this bump clearly while filter based processing is far more unclear with both the EpCAM and CD45 signal being mixed together.	39
Figure 3.6 Idealized graphs of the greatest difference between filter based imaging and full spectral imaging with the use of nearby dyes. Spectral imaging can support the addition of a fifth dye in the 550-600nm range (yellow) without issue while in the filter based system this would cause additional bleed over.....	40

Figure 3.7 Experimental workflow from patient sample input to final data output.	41
Figure 4.1 Cell ranking RoC plot illustrating reduction in false positive rate with increasingly stringent ranking cutoff.....	43
Figure 4.2 Cell ranking RoC plot illustrating reduction in false positive rate with increasingly stringent ranking cutoff.....	44
Figure 4.3 Comparison between the three methods of analysis. H010 is the negative control injected into the review process with only the manual IF method failing to exclude the negative control sample.	45
Figure 4.4 processing performance of the automated algorithm split by image file conversion, image processing, and report generation. A majority of the processing time is spent encoding video output report files. In addition the report generation is sensitive to the number of cells being processed.....	47

List of Abbreviations

AVI	Audio Video Interleave	(A common video container format)
CK	Cytokeratins	(Group of cell structure proteins)
CTC	Circulating Tumor Cell	(Rare Cancer Cell)
CZI	Carl Zeiss Image	(Proprietary XML image data format)
DAPI	4', 6-diamidino-2-phenylindole	(A common nuclear fluorescent stain)
DNR	Did not run	(Process failed to complete run)
EpCam	Epithelial cell adhesion molecule	(Cell adhesion molecule)
EMT	epithelial-to-mesenchymal	(Transition for mutating cells)
FCC	False Color Composite	(False coloration method)
FPR	False Positive Rate	(Rate of false positives within a sample)
HRM	Human Readable Medium	(Text, common standard image/video)
IF	Immunofluorescence	(Fluorescent markers bound to antibodies)
MJPEG	Motion JPEG Video	(Video sequence of JPEG images)
RBC	Red Blood Cell	(Primary cellular component of blood)
RoC	Receiver Operator Curve	(Plot of sensitivity vs. specificity)
ROI	Region of interest	(Specific area of analysis)
SD	Standard deviation	(Statistical variation)
TB	Terabyte	(Unit of digital data ~1400 CDs)
TPR	True Positive Rate	(Rate of true positives within a sample)
UV	Ultraviolet	(Wavelengths shorter than blue visible light)
WBC	White Blood Cell	(Immune System Cell)
XML	Extensible Markup Language	(A general purpose markup language)

Acknowledgements

This project was made possible with the support of my friends, coworkers, and family.

In particular, I owe a debt of gratitude to

Dr. Hong Ma, for his patience and support throughout the research process and for exposing me to many new ideas and projects.

Emily, Chao, Simon, Aline, and the rest of the Multiscale Design Lab for their support and company in the lab.

My parents and brothers, whose patience and support saw me through my endeavors.

Thank you

Dedication

My parents and brothers

1,3,4,5,6-Pentahydroxy-2-hexanone

Sodium 2-aminopentanedioate

3,7-dimethyl-1H-purine-2,6-dione

Chapter 1: Introduction and Background

1.1 Introductions

Circulating tumor cells (CTCs) have attracted a tremendous amount of attention because of the potential value of these cells in cancer research and treatment. In clinically advanced cancer, tumor cells may be shed from a localized tissue into the bloodstream, where they have the potential to metastasize to seed new tumors at secondary sites (Figure 1.1). These cancer cells in the bloodstream are collectively known as CTCs and they potentially represent cells derived from the primary tumor, metastatic tumor, or tumor cells that occupy the transitional state between primary and metastatic tumors [1]. Access to CTCs is important because metastasis is associated with 90% of cancer deaths [2] and the enumeration of CTCs has a well-established prognostic value. Specifically, the number of CTCs in peripheral blood has been found to correlate with reduced progression-free survival and reduced overall survival rate for patients with metastatic breast cancer [2], castration-resistant prostate cancer (CRPC) [3], as well as ovarian [4], lung, colon, and pancreatic cancers [5]. CTC isolation is particularly relevant in cancers where metastatic tissue is inaccessible. For example, prostate tumors typically metastasize to bone marrow [6] where tissue biopsy is difficult and painful. In contrast, CTCs enriched from a non-invasive blood sample can be used for patient diagnosis and stratification, to guide and monitor therapy, and can provide key insights into the genetic events that collectively contribute to tumor metastasis.

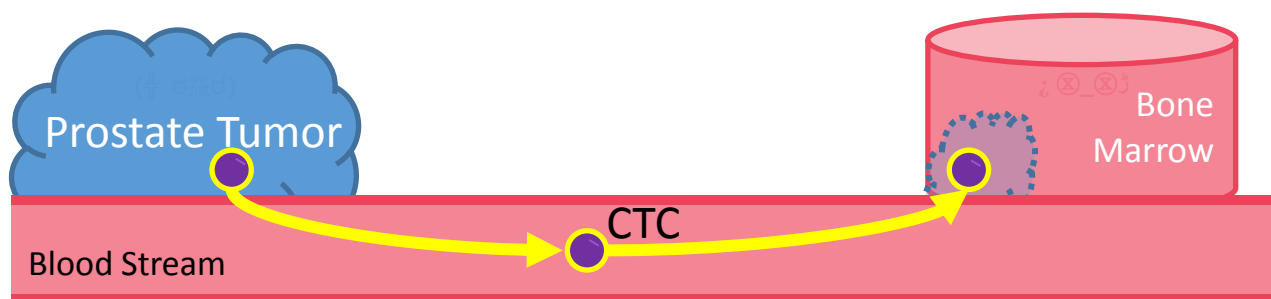


Figure 1.1 Metastatic dissemination of tumor cells, exemplified by prostate cancer. Cells develop in the primary tumor, shown in blue, and acquire a motile phenotype to disseminate as CTCs in the bloodstream. These cells also develop invasive and stem-like characteristics that allow them to initiate a tumor at a secondary site.

Despite their recognized clinical value, the identification and characterization of CTCs is significantly challenged because of the extreme rarity of these cells. Efforts to perform image cytometry on unprocessed blood must discriminate individual CTCs from among 5 million nucleated cells and 5 billion red blood cells [7], [8], [9]. Immunological and biophysical enrichment methods have aimed to reduce the complexity of CTC identification. However, these methods still invariably generate an impure sample. The CellSearch™ system, developed by Janssen Diagnostics, is the gold standard for CTC enrichment using immunoaffinity capture, but this system has been shown to provide limited yield and purity [10]. While methods for tumor cell enrichment generate isolates at high purity, these methods are associated with unacceptable loss of these highly relevant cells [10]. The impurity of most CTC isolates, is further compounded by an inherent heterogeneity among CTCs, where these cells may initially exhibit an epithelial phenotype, but progressively acquire characteristics of motile mesenchymal cells [11]–[14], proliferative stem cells [15]–[17] and invasive metastatic cells [1]. As a consequence of both the rarity and complexity of CTCs, there is a critical need for defined and objective criteria to identify these cells.

Current criteria for identifying CTCs relies on both cellular morphology and immunophenotyping. The traditional criteria for identifying CTCs included an intact nucleus, expression of epithelial cytokeratins (CKs), and absence of the CD45 leukocyte marker [18]. However, this definition has been challenged since CK expression may be reduced in tumor cells that undergo epithelial-to-mesenchymal transition (EMT) [14], [19]. These antigenic biomarkers may therefore be supplemented or replaced by disease-specific markers, such as PSMA for prostate cancer [20], or by inclusion of EMT (e.g. vimentin, N-cadherin) [11], [15] or stem cell markers (e.g. ALDH, CD44) [13], [17]. Similarly, CTCs are likely to be larger than leukocytes because of their epithelial origin. However, smaller CTCs have recently been associated with progressive disease [8], [21]. CTCs also typically have an irregular morphology [18], [21] but evidence of nuclear fragmentation or cytoplasmic blebbing indicate cell apoptosis and can serve as an important indicator of therapeutic efficacy [22]. With the expanding list of criteria for defining and categorizing CTCs, there is a need for improved multiparameter of these cells.

Currently, the CellSearch™ system is the only FDA-cleared system for the detection and enumeration of CTCs. This system first enrich for CTCs using EpCAM immunoaffinity, and then

identify them using immunofluorescence staining for EpCAM, CK, and CD45. While this method has undergone significant refinement, the inherent subjectivity of CTC detection has contributed to significant inter-lab variability [23], [24]. This thesis presents a system intended to standardize and simplify the detection of fluorescence-stained CTCs using multi-spectral analysis. This goal will be accomplished by presenting a workflow for the collection, processing, and presentation of multi-spectral image cytometry data.

The specific goals of this thesis are to develop an efficient and flexible spectral image cytometry software platform for semi-automated image cytometry to identify CTCs from an enriched sample. The following steps will be taken to achieve this goal:

1. Develop tools to acquire proprietary microscope multi-spectral imaging data into a format suitable for further analysis.
2. Develop tools to segment the composite microscopy image into single cell images.
3. Develop automated image processing software to rank individual cells based on their likelihood of being a CTC.
4. Develop a user interface for rapid augmented review of potential CTCs.
5. Evaluate the spectral image cytometry platform by comparing results from augmented user review against full manual review.

The remaining sections of this chapter describes the background of this work. Specifically, Section 1.2 describes immunofluorescence-based CTC labeling and morphological criteria for CTC identification. Section 1.3 describes existing cytometry methods for data acquisition. Finally, Section 1.4 reviews three existing CTC image cytometry workflows.

1.2 CTC Identification Criteria

1.2.1 Fluorescence Labeling

Identifying CTCs using fluorescence labeling first involves distinguishing CTCs from contaminating erythrocytes based on the presence of a nucleus, which stains positively for DAPI (4', 6-diamidino-2-phenylindole). To discriminate CTCs from contaminating leukocytes, fluorophore-conjugated monoclonal antibodies have been widely employed in the defining CTCs

based on their antigenic profile. CTCs have been identified based on the presence of the epithelial cell adhesion molecule (EpCAM) [25] and cytokeratins (CKs) 8, 18 and 19 [26]. Since EpCAM has become the primary criterion for immunocapture of CTCs, CK immunostaining is typically used for positive selection of these cells. However, CK expression may be inherently variable or may change due to an epithelial-to-mesenchymal (EMT) [14]. Consequently, patient specimens are also stained for the pan-leukocyte marker CD45, which should be exclusively expressed by hematopoietic cells. While an ideal scenario would involve the simultaneous staining of all relevant antigens, there are significant limitations that arise when staining for multiple biomarkers.

The primary limitation to simultaneous staining for multiple fluorescence biomarkers derives from overlap between fluorophore emission spectra. Identifying cells using immunofluorescence (IF) microscopy involves illuminating cells with an excitation light, which is typically broadband source, but could be restricted to a specific color. Specific antibodies bound to fluorophores will react with and emit a unique emission spectra which is typically a longer wavelength [27], [28]. In order to provide optimal response to experiments, users typically select the peak of the excitation and emissions. An example fluorescence emission spectra for DAPI, CK, EpCAM, and CD45 is illustrated in Figure 1.2 with DAPI being a nucleic acid stain and Alexa 488 (A488), Alexa 594 (A594), Allophycocyanin (APC) being respectively used as secondary antibodies bound to their respective primary antibody markers. Idealized spectral graphs derived from industry standard data are used throughout this thesis to demonstrate theoretical differences without the added noise and complexity of experimentally determined spectra. A complication of immunofluorescence is the limited optical spectrum available to both excite and detect emission responses. Fluorophores can easily overlap if not carefully chosen with sufficient spectral separation.

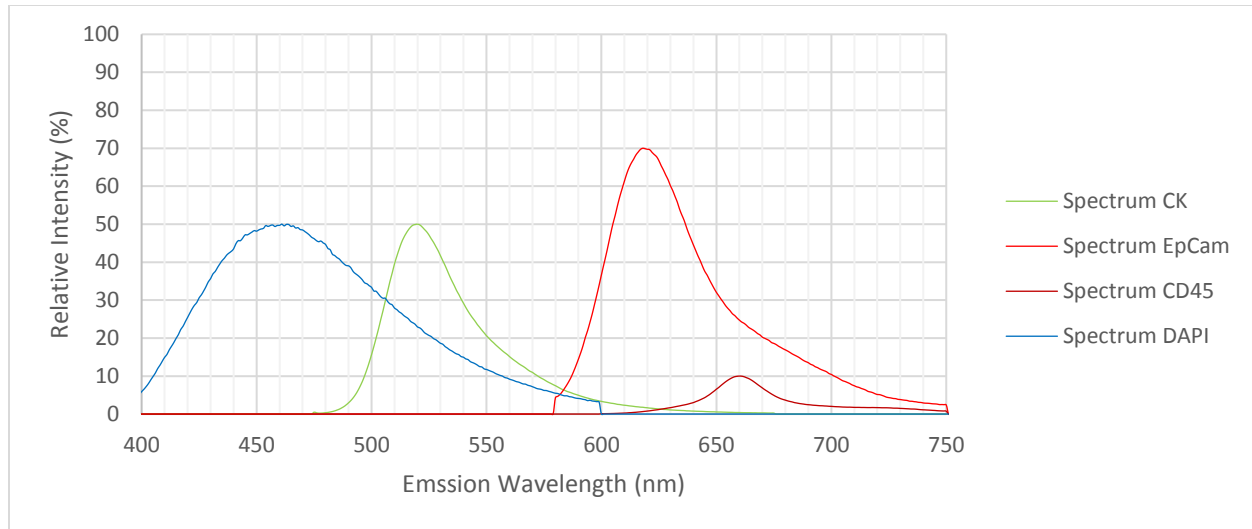


Figure 1.2 Idealized emission spectra for marker based detection of CTC/WBC samples generated using Excel. C16H15N5 (DAPI) dye stains nuclear DNA directly, CK antibody bound to Alexa 488 (A488) dye stains cytokeratin common to CTCs, CD45 bound to Allophycocyanin (APC), and EpCam bound to Alexa 594 (A594)

Another issue with immunofluorescence based profiling of CTCs is the variability in marker expression levels [12], [14], [29]. Additionally, WBCs have been observed to express CTC markers described previously due to WBCs consuming CTC fragments or attacking CTCs in circulation. By analyzing multiple CTC markers their highly variable expression can be more reliably detected.

Many CTC enrichment technologies also utilize one or more antigenic markers to enrich samples prior to image cytometry which can result in a lost subset of the CTC population as well the loss of even semi-quantitative analysis of the critical CTC markers being used to enrich the sample. An alternative to this approach is to develop imaging methods capable of processing impure specimens [30], [31]. The advantage of this approach is the loss-less discrimination of CTC sub-populations based on a range of criteria, including multiple protein biomarkers and morphological criteria. However, this approach generates a tremendous volume of complex data and a significant advance in image analysis would be required to interpret these data.

1.2.2 Cell Morphology

Cell morphology, including size, shape, and internal structure, can be further used to identify and analyze CTCs [32]. The intracellular space is the main volume within a cell which can contain structures such as the cytokeratin which form the cytoskeleton [33], [34]. Among the membrane bound organelles is the nucleus that contains DNA chromatin. The shape of the nucleus is typically spherical, owing to the rigid structural proteins of the nuclear lamina, but the nucleus can exhibit ‘blebbing’ and other abnormal morphologies, as a consequence of apoptosis [35]. Both cellular and nuclear morphology are valuable for identification of CTCs because cellular morphology can aid to discriminate these cells from leukocytes and nuclear morphology can discriminate viable CTCs from non-viable apoptotic artefacts.

CTCs have been discriminated from other circulating cells on the basis of cell size and nuclear morphology. In the study of CTCs from CRPC patients, some studies have shown significant heterogeneity in CTC size [7], [8] while other studies have shown more consistent CTC size [21]. In general, an average CTC size of 8 μm has generally been reported [36]–[38]. A significant limitation of size based discrimination is that hematological cells may exhibit significant overlap in size with CTCs. Additionally, cell shape in CTCs has been known to exhibit significant pleomorphism with shapes ranging from spherical to eccentric shapes [7], [8], [38], which may confound size-based sorting methods. In contrast to cell morphology, nuclear morphology has been underutilized as a selection criterion for CTCs. The nucleus forms a regular and rigid sphere, whereas CTCs may display irregular nuclear shape due to apoptotic stress and have greater nuclear cytoplasmic ratio known as (N/C), which is the ratio of the visible nuclear area to the visible cell area. In CTCs studies high N/C ratios have been correlated to poor disease outcomes along with high intra-patient variability [21], [39]. Since differences in N/C is likely to result in differences in cell deformability, our group’s CTC separation efforts have focused on the development of technologies to enrich for CTCs based on cell deformability [40], [41].

Coupling the morphology of a cell with an immunofluorescence response allows for more detailed analysis of cell populations and can provide information about the state and condition of a cell being observed. The location and spatial distribution of features is the purpose of morphological

analysis, which can potentially help to identify specific phenotypes without additional fluorescence markers.

1.3 Fluorescence Cytometry Methods

This section reviews three existing approaches for fluorescence cytometry. These methods include:

1) Standard fluorescence microscopy, which captures image data from specific fluorescence emission bands; 2) Flow cytometry, which captures fluorescence emission from single cells or from images of single cells in a calibrated flow stream; and 3) Spectral confocal microscopy, which captures fluorescence images from a scanned image using a scanned laser excitation.

1.3.1 Fluorescence Microscopy

Fluorescence Microscopy is a standard microscopy technique for cell phenotyping [42], [43]. Band pass optical filters are commonly used in wide field microscopes to image specific emission wavelengths allowing a standard grayscale camera to capture a wide variety of color channels sequentially. An idealized spectrum based on industry data [44] is shown in Figure 1.3 with markers used in CTC identification and arbitrarily narrow filter bands. In practice, filter selection is limited to industry standard filters and narrow band pass filters result in a loss of light capture efficiency. The limitations of standard fluorescence microscopy using filters is that they have limited spectral resolution.

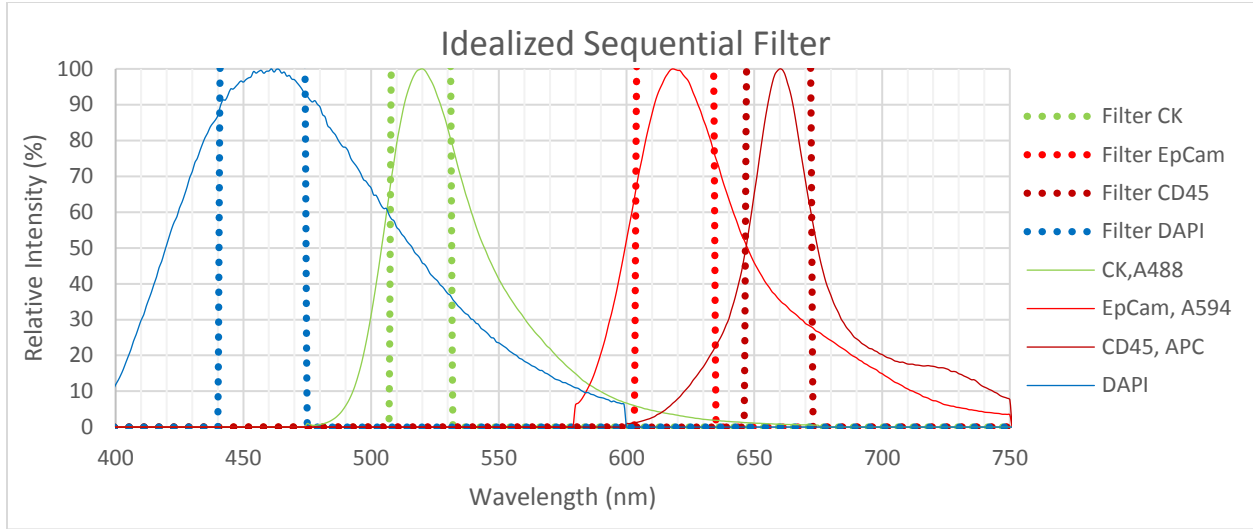


Figure 1.3 Emission Spectra with idealized filters overlaid to capture peak signals. In practice filters are not ideal and overlap between signals is increased.

To increase the spectral resolution many temporally multiplexed band pass filters can be used to capture an overall spectrum [43]. The sequential capture of images using filters multiplies the capture time by the number of bands desired. Alternatively, there are spatially multiplexed systems that allow for multiple filter bands to be captured simultaneously using micro lenses and other advanced optical techniques which sacrifice sensor resolution for additional color bands. A downside of wide field spectral microscopy is that filters must be switched for each emission band which multiplies capture time or the spatial resolution of the sensor has to be dedicated to an emission band dividing the resolution of the final image [45].

1.3.2 Flow Cytometry

Flow cytometry captures bright-field and fluorescence data from cells in fluid suspension as they are flowed individually in front of optical detectors [46]. Typical flow cytometers, shown in Figure 1.4, only captures a single pixel of information for each individual cell, which does not allow for morphology based discrimination. Additionally, many flow cytometers can also sort cells using electrostatic cell sorting for downstream analysis.

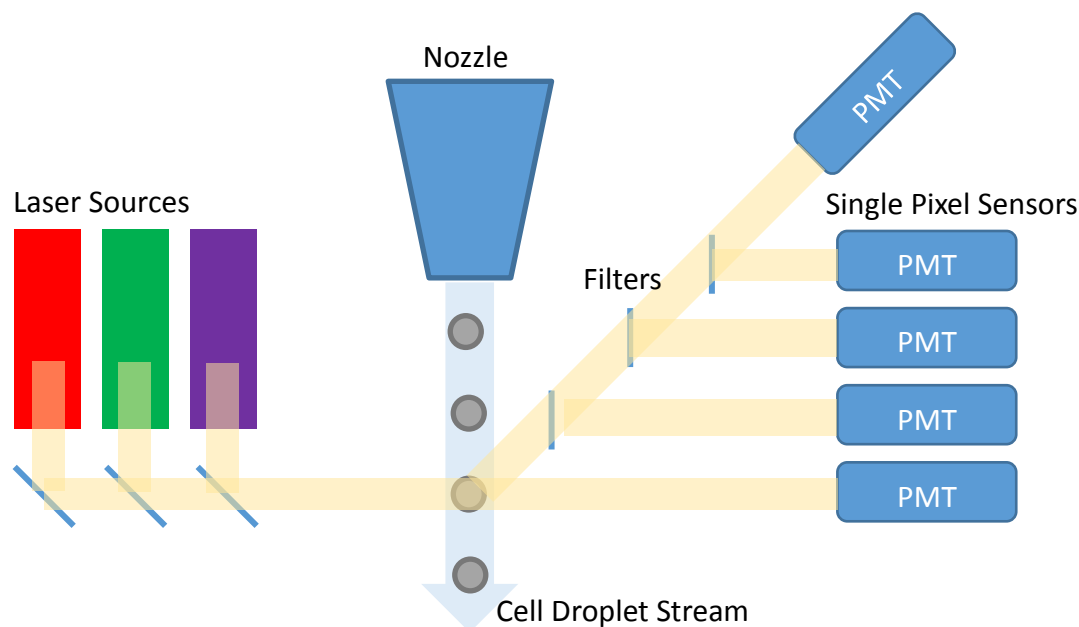


Figure 1.4 High level Overview of a Flow Cytometer. Coherent laser sources illuminate a single cell which is acquired by single pixel sensors with filter plates to capture spectral slices.

To address the lack of morphological image information of conventional flow cytometers, some image flow cytometers have been developed to incorporate the ability to imaging of the cells as they are measured by the detector [47]. As compelling as this approach may be, this system lacks the ability to localize or sort the selected cells after detection has occurred. The constraints of the microfluidic structure and processing used in existing state of the art imaging flow cytometers precludes this type of sorting [48], [49]. Furthermore, flow cytometers are an active instrument requiring single cell dispersions limited by the fact that there is always a dead volume of sample that the instrument cannot effectively process [50], [51] which can cause significant cell loss. These factors add to the general requirement for flow cytometers to process large quantities of cells in order to operate [52] with statistical confidence.

1.3.3 Spectral Confocal Microscopy

Confocal microscopes use a pinhole to control the exact focal plane received by a single pixel sensors in order to provide the ability to image samples in 3D. This pinhole enables the selection of the z-depth returned light allows for the suppression out of plane fluorescent emissions. On a

traditional microscope, out-of-plane emission creates unwanted background illumination. In order to build up an image of the cell, the point sensors receive light scanned rapidly by moving mirrors that raster scan excitation lasers across the field of view which responds with spectral emissions reflected back to form the spectral image stack. Because the system captures light as a single pixel at a time, capturing high spectral density images without increasing the capture time is easily possible by using a high channel count sensor [45] shown in Figure 1.5.

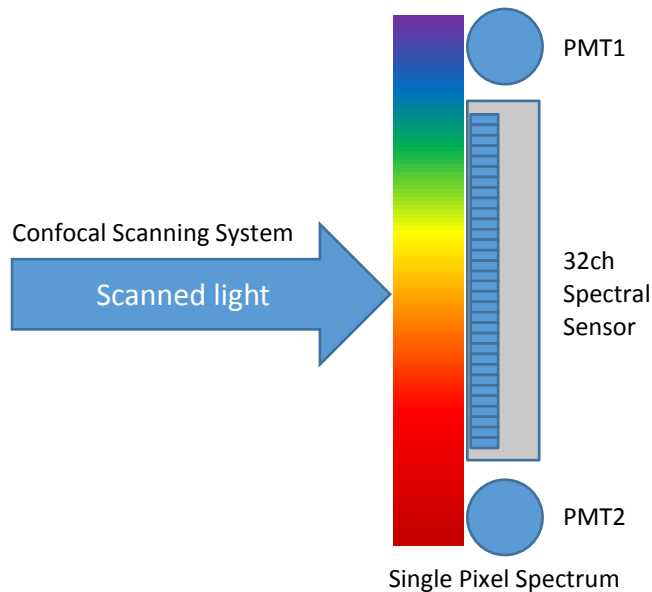


Figure 1.5 Splitting light from a scanned confocal system into a continuous spectrum allows for a high spectral content image to be captured. PMT units on the ends can be configured to capture a fixed slice of the spectrum from either end.

Diffraction gratings can be used to separate the emission light into a continuous spectrum that can be distributed over multiple sensors. This form of spectral decomposition is possible because only a single point of light is being received at a time. Another advantage of the multispectral confocal system is that it permits imaging through challenging conditions, such as the presence of auto-fluorescence. Background removal can be performed using spectral profiling on the background signal which can then be removed by isolating channels of high background signal and by tuning excitation energies to avoid exciting the background material. Transmitted light through the sample can be captured by a transmitted light sensor to obtain a bright-field like image. Unlike other spectral capture systems, for spectral confocal microscopy no filter movement or sequential

imaging is required and this capability greatly reduces capture time by enabling simultaneous spectral capture [53].

The work described in this thesis uses a specific spectral confocal microscope, Zeiss LSM 780, with an image tiling process that produces a gigapixel spectral image cube in approximately 5-10 minutes per image. This image cube shown in Figure 1.6 has sufficient spectral information to easily resolve 5 emission bands with an additional channel for brightfield imaging. The additional color information between the bands allows for the 5 emission bands to overlap without compromising the ability to correctly identify the desired spectral signature.

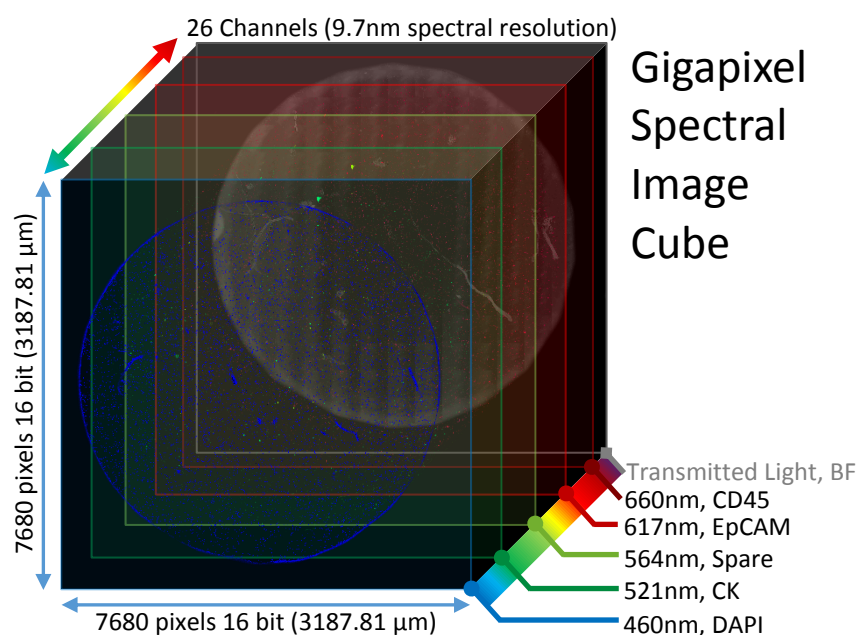


Figure 1.6 Visual representation of the imaging setup with a 7680x7680x26 spectral image cube. Peak channels are listed as well as the quality control bright field channel which is obtained from the transmitted light through the sample while all color channels are fluorescent emissions from the sample.

1.4 Image Cytometry Platforms for CTC Identification

Image cytometry is the measurement and characterization of cells performed by optical microscopy. There are a number of image cytometry platforms for CTC identification. Three prominent platforms will be reviewed here. First, the Cellsearch™ system developed by Janssen Diagnostics, which is currently the only FDA-approved CTC identification platform. Second, the

Amnis ImageStream™ imaging cytometer for CTC identification with a defining feature for its ability to process massive quantities of cells per sample. Finally, Epic Sciences, has a proprietary in-house process that claims to process samples solely through image cytometry.

1.4.1 CellSearch

The CellSearch system uses standard immunofluorescence processing to identify CTCs based on CK vs. CD45 expression after physical EpCAM enrichment [54]. Analysis is performed semi-automatically, with an automated image workflow capturing images from a tiled standard IF microscope scan. This scan presents sub-regions with automatic exposure to be reviewed by a human operator who judges if a region contains a CTC, as illustrated in Figure 1.7. The process has four channels for analysis with one channel open for other markers.

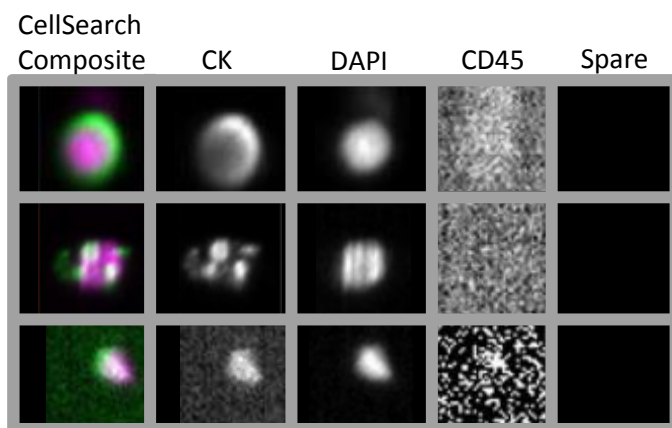


Figure 1.7 Cell Search images displayed are a color composite of DAPI/CK with an image series of CK, DAPI, CD45, and a spare channel [55](© 2014 Ignatiadis et al. adapted under, CC BY 2.0). Noticeable oversaturation can be seen as well as automatic exposure control which can lead to variations in perceived intensities.

While CellSearch offers a complete workflow for CTC enrichment and isolation, this platform has some important limitations associated with characterization of these cells. Firstly, CellSearch employs the positive selection of CTCs based on EpCAM immunomagnetic affinity capture, which represents a potential concern since it is possible that EpCAM low-expressing cells escape the capture process and a subpopulation of CTCs could be lost. This is a major criticism of this system, in light of evidence for a phenotypic shift in CTCs that corresponds with reduced EpCAM expression [34]. The loss of EpCAM expression is expected since the downregulation of EpCAM

is required for tumor cells to escape the tissue and enter into circulation through epithelial-to-mesenchymal transition (EMT) [56]. A second issue with the CellSearch process is that imaging is performed in a specialized micro pillar array. CTCs mechanically entrapped within this array cannot be easily extracted for downstream characterization and methods used to extract these cells typically sacrifice the positional information of the CTCs.

1.4.2 Amnis ImageStream

Amnis ImageStream, produced by MilliporeSigma, is an imaging flow cytometer that captures an image of each cell in flow in order to give users the ability to classify cells based on imaging criteria. The system offers a form of spectral imaging which is comprised of a total of 12 imaging bands utilizing a spatially multiplexed method including non-spectral bright field and dark field images. However, sorting and recovery of the cells post-process becomes impossible due to the dynamic method of measurement.

This approach has been used to identify CTCs using aspects of flow cytometry coupled with standard microscopy. A recent study demonstrated the ability for this imaging flow cytometer to identify a broader definition of CTCs, defined as any cell without CD45 [57]. Compared to CellSearch, however, the images obtained using this system has a lower quality with little to no intracellular detail visible within the cell, which thus makes morphological analysis difficult. Figure 1.8 shows an example image produced using the ImageStream with the variability in image quality.

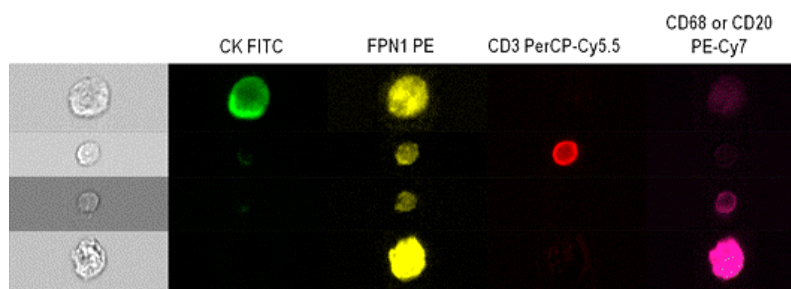


Figure 1.8 Example cell images using image stream flow cytometer with decreased cellular definition. Image retrieved from [58] (© 2016 Marques et al., adapted under CC BY 4.0). Analysis is performed semi automatically via an image panel or 2 parameter scatter plot.

Another issue with the ImageStream system is the loss of cells resulting from the need to flow each cell in front of an imager. A comparative study of CTC enumeration using the CellSearch and ImageStream systems showed the latter to be less efficient for rare cell detection due to low yield of cells [59]. This result is expected as flow cytometers typically require thousands of target cells for reliable detection. In fact, cell loss in other CTC studies are observed to be as high as 90% for standard flow cytometers with improved performance coming from the inclusion of imaging during flow cytometry [60]. Consequently, flow cytometry based CTC enumeration methods are typically limited to patients with high CTC counts or would require significant modification from the standard process.

1.4.3 Epic Sciences

Epic Sciences has a proprietary centralized lab operation that offers in-house CTC enumeration. The advantage of this system is that its use of large-scale immunofluorescence imaging to ensure the sample can be analyzed following only red blood cell (RBC) depletion, in the absence of CTC enrichment or WBC depletion step. This results greatly reduces the likelihood of target cell loss due to sample processing. The cells are spread on a custom blood smear plate, which has been validated through in house tests on numerous patient samples [31]. However, a drawback of Epic sciences system is the lack of technical detail on their software implementation, which is based off previous research for HD-CTCs [30]. Another major concern is that this method is only offered as an in-house service, requiring samples to be submitted to the dedicated laboratory for processing. Consequently, there is limited information available on this system, but available literature suggest that CTC identification is performed using three-channel standard fluorescence imaging (Figure 1.9) with an optional fourth channel. An additional feature of the Epic Science system is the ability to extract single cells using a process similar to micropipette aspiration.

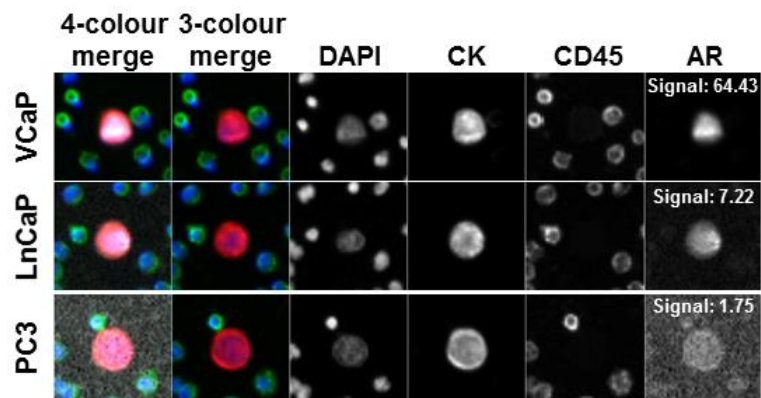


Figure 1.9 Epic Sciences Full Channel Analysis image retrieved from [31] (© 2015 Werner et al., adapted under CC BY 3.0). Images are displayed in an image panel to users for review with a semi-automatic selection process.

Chapter 2: Design of Software for Spectral Image Cytometry for CTC Identification

To address the need for rapid CTC identification using spectral imaging, we developed a flexible software platform for spectral image cytometry to augment the work of a human reviewer. This goal is accomplished by automatically eliminating many clearly negative cells and produce subset of potentially positive cells for downstream human review. This software platform also outputs physical coordinate information with guidance maps for individual cell identification and retrieval if desired in follow on physical processing for tests such as single cell genome sequencing.

This chapter describe the overall software workflow from input file conversion to the user interface for final human review. Section 2.1 describes the high-level workflow in three major sections. Section 2.2 describes the input file conversion and metadata processing. Sections 2.3 describes the pixel-based image processing component with the spectral basic analysis producing selection masks for use in sub-region analysis. Section 2.4 describes sub-region analysis which utilizes the mask outputs from the previous section to group pixels together for analysis as cells. Section 2.5 describes the spectral aware ranking process. Finally, Section 2.6 describes the user-interface software for augmented human review of the ranked candidates in order to positively identify CTCs.

2.1 Image Processing Workflow

The developed spectral image cytometry software has three major components which starts with ingest processing which is required for adapting multiple propriety data formats for image analysis. This process converts the propriety formats into raw image data and textual metadata associated with the file. Following this, image processing is required to convert the raw image data into usable information to produce result reports. This image processing utilizes pixel-based processing to generate masks used to enclose selection regions that may contain cells. The report generation uses the processed image data and sub-region lists to create ordered reports suitable for rapid human review. These three major steps represent a high-level overview of the software and is shown in figure 2.1.

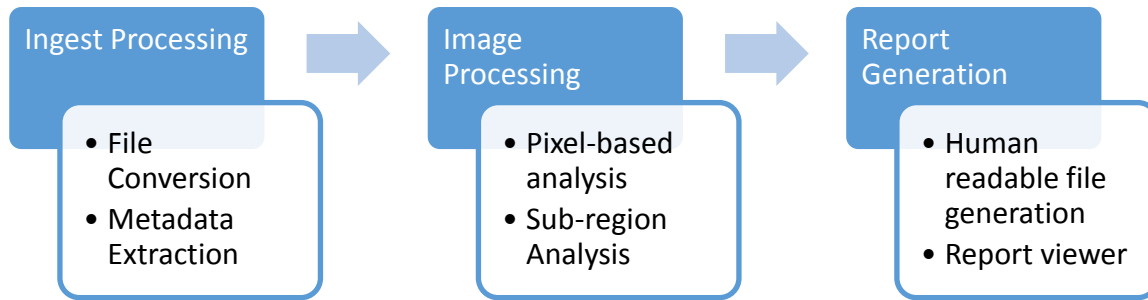


Figure 2.1 Overall workflow process from input file to output report.

2.2 Ingest Processing

A significant challenge in multispectral analysis is that imaging platforms typically encode spectral data in proprietary formats that must be decoded before further processing. A further challenge is that metadata on the microscope imaging parameters encoded within these formats are often lost during the image conversion process performed by existing commercial tools. Computation tasks shown in Figure 2.2 are utilized to address these two issues. Specifically, the image ingest operation is performed using an open source bio-formats image conversion processor which provides support for a wide variety of proprietary formats and forms part of a larger open source processing package called OMERO [61]. Output from the open source library is then converted into a simple 3D matrix containing the spectral image cube.

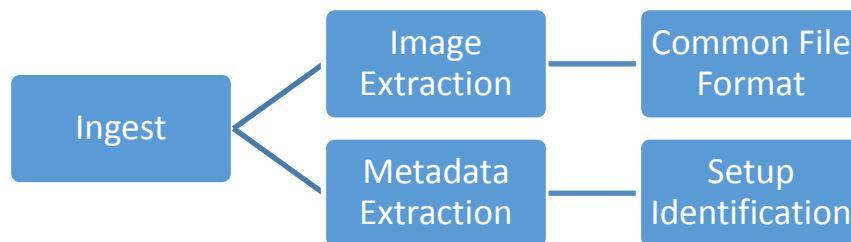


Figure 2.2 Ingest workflow to extra image data and image metadata.

Microscopy images acquired by the Zeiss LSM 780 system are written in a proprietary Zeiss image container format, known as a Carl Zeiss Image (CZI), which can pack multi spectral/spatial/temporal image series. To allow the automated image cytometry method to be a more general purpose system, an open source image library is used to convert the CZI into a MATLAB native binary format which serves as a common file format for image data stored as a

raw bitmap. A potential complication of using a general purpose data format is that company specific metadata, such as microscope setup information, is lost during conversion. To address this issue, another tool was developed using Zeiss CZI API to extract from company specific Extensible Markup Language (XML) specific imaging parameters used to setup the experiment.

Because multiple experiments relating to CTCs have been captured using this software tool a potential risk of file mishandling exists. Because only experiment data from prostate and bladder cancer patients compose of the test dataset used in this thesis a method to verify a file's origin was required. As these specific prostate and bladder cancer images contain specific metadata parameters they can be detected by reviewing the microscope setup data contained within the CZI files. To enable the review of such experimental metadata to ensure its consistency to the particular setup and calibration, a tool was developed, seen in Figure 2.3, to extract the image metadata of the large quantity of image files produced in the course of research.

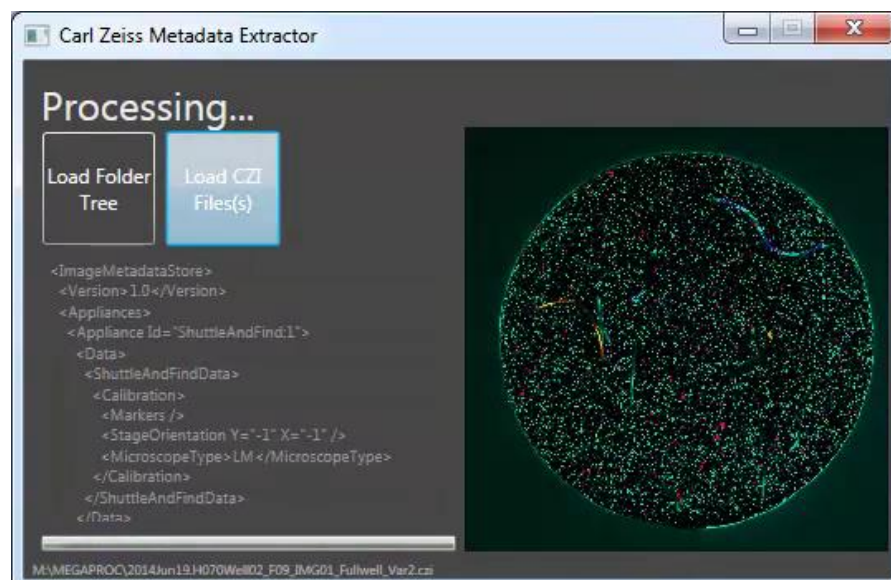


Figure 2.3 Metadata extraction tool interface displaying current file thumbnail and a snippet of the raw metadata. Used to determine the actual experimental setup from recorded metadata within sample files.

More specifically to identify a particular experiment type, the metadata is analyzed to compare the image size, the excitation laser beam settings, the laser splitters used, and the optical setup of the microscope. With the management and review of the collected meta-data information across hundreds of samples also enables the supervision of experimental conditions. This metadata signature tracking and review process can detect the use of improper configurations and suspect

sample images. Overall the careful management and monitoring of captured metadata is a critical aspect of an imaging workflow as during two years of research over 8 TB of image data has been produced. The embedded metadata also enables blinded manual processing by scrambling human readable identifiers, which enables the recovery of the original embedded metadata after blinding for human review.

2.3 Image Analysis

The raw spectral data represents more color information than a human can naturally process with 26 bands being simultaneously captured and the false color composite generates a customizable three color image that is easily viewed by human reviewers. With over a gigapixel worth of image data a rapid and efficient data analysis method is required to reduce the big data set into a more manageable fixed set of selection masks. Spectral pixel analysis performs a simplified spectral selection process on a per pixel basis to generate selection masks used in sub-region generation. Sub region analysis is required to convert the selection masks into pixel groups for reporting and ranking processes. This sub region processing also performs minor filtering processes to remove undesired targets. An overview of these steps are shown in figure 2.4.

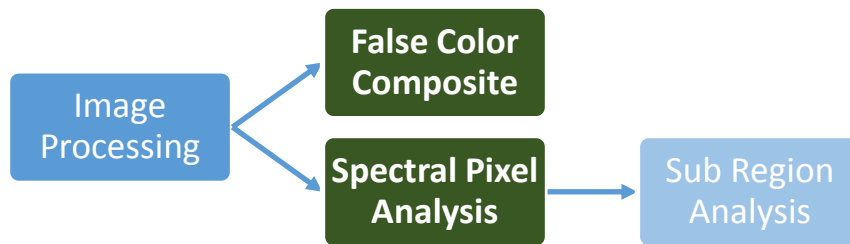


Figure 2.4 Split tasks for generating human readable false color composite images as well as processing raw image data into a more usable set of selection masks for downstream sub-region analysis.

2.3.1 False Color Composite

The false color composite (FCC) tool allows the users to generate a customizable composite image shown in Figure 2.5 enabling rapid screening for potential CTCs by allowing the user to prioritize specific channel of interest using an easily identifiable colors. Another function of the FCC tool is to perform histogram compensation on each channel, which is used to account for the different

expected intensities of each dye. The Zeiss ZEN Black software provided with the LSM 780 microscope can generate its own FCC. However, ZEN Black does not give the user control over the color mapping process. Consequently, channels for EpCAM and CD45 are colored almost identically, which makes it more difficult to distinguish non-CTCs from leukocytes. Furthermore, ZEN Black selects the color of each pixel base color based on the peak channel, which makes it possible for a bright channel to completely obscure the presence of a dimmer channel.

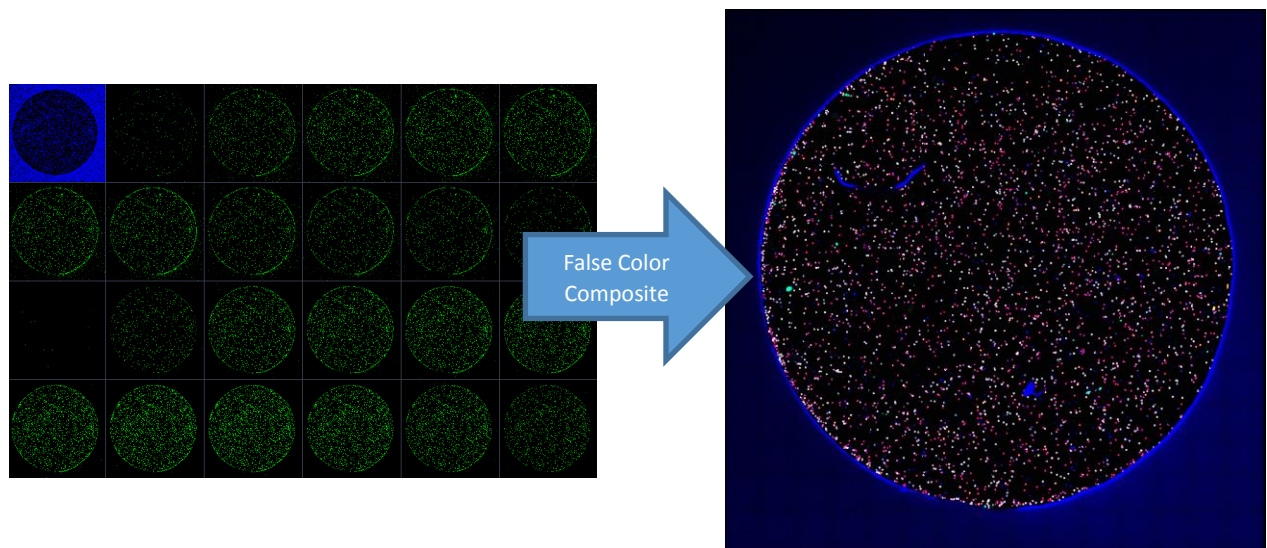


Figure 2.5 Compressing hyperspectral stack into a single image for easier review and management.

Our FCC tool allows for custom prioritization of spectral channels and customized color mappings. The difference between false color composites generated using ZEN Black and the improved color mapping method is shown in Figure 2.6. The false color composite image has been optimized for clear discrimination of relevant cell features that define CTCs. For example, while the CD45–positive cells are represented the Zen Black image, a false color composite (FCC) image can incorporate color re-mapping to highlight this antigen profile in a high contrast color (e.g. white), which greatly simplifies the identification of non-target cells. Furthermore, while the weaker DAPI signal is typically obscured within the ZEN Black image, intensity customization allows the DAPI signal to be clearly visible through the FCC. This feature provides an important advantage for identifying CTCs because nuclear-to-cytoplasmic ratio is an important defining characteristic of CTCs and because nuclear blebbing and other morphological characteristics may discriminate

apoptotic CTCs from more aggressive subpopulations [21], [39]. Distinguishing between clearly negative and positive peak channels is made extremely obvious even if only a few pixels contain the peak channel. This priority system allows for a per channel customization of how pixel coloration is selected and graded in order to generate the false color composite. This prioritization system decreases the threshold so that even a non-peak priority signal will be used in priority to other channels.

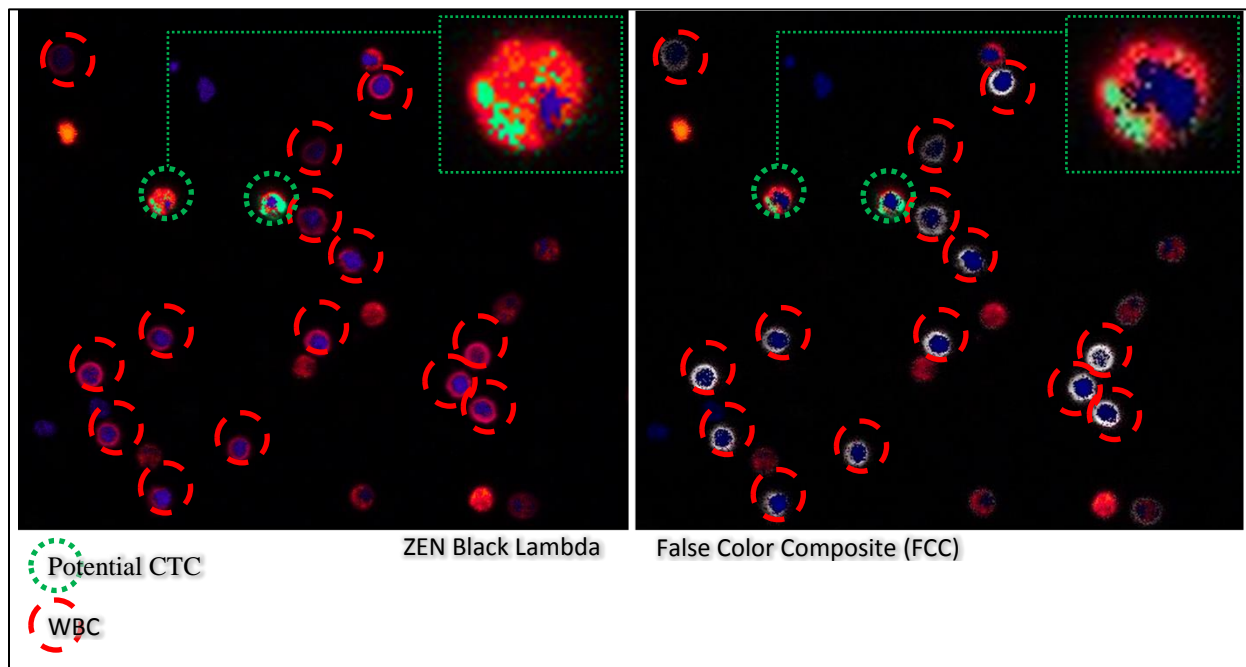


Figure 2.6 Comparison between microscopy software ZEN and custom false composite with negative pixel highlighting (white = negative pixel) and control pixel prioritization (blue nucleus enlarged).

2.3.2 Spectral Pixel Analysis

Pixel-based processing involves analyzing the entire gigapixel image on a per pixel basis which results in billions of computations requiring significant amounts of computation and memory. As a result, pixel-based processing requires extensive optimization and cannot use highly iterative or complex algorithms due to performance and memory limitations. A simplified selection algorithm that efficiently utilizes the additional spectral data to improve selection performance was needed.

The LSM 780 acquires intensity data from 26 wavelengths on each pixel. Our spectral image cytometry software first analyzes each pixel to generate the masks required to segment the image into separate single cell images. This analysis is performed on every pixel without determining whether the pixel is part of a cell. A key advantage of this approach that memory use is fixed to the number of independent test groups with a single result selection mask, regardless of the number of cells. Follow on analysis of these single cell images are described in the next section.

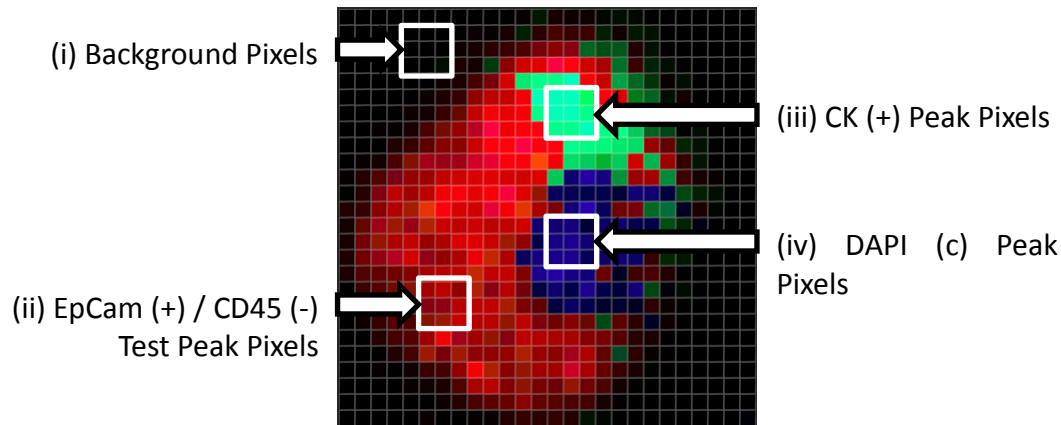


Figure 2.7 Spectral pixel testing done across the entire gigapixel image, small subset shown with pixel types listed. (26x26 pixel close up of cell from ZEN lambda view)

Figure 2.7 shows an example image of a potential CTC showing various types of pixels including background pixels (i) with a very low signal, EpCAM positive pixels (ii) with a red coloration, CK positive pixels (iii) with a green coloration, and DAPI positive pixels (iv) with a blue coloration. Cells are identified as CTCs when they stain positively for CK and EpCAM, and negatively for CD45. Background noise may vary depending on the wavelength due to the auto-fluorescent properties of the imaging media or sample holder which can be suppressed with per channel threshold limits to define a per channel noise floor.

The linear un-mixing is process that is commonly used to unmixed spectral information and assumes that for every wavelength the signal intensity is a linear sum of its components. This process then relies on reference spectral data for each expected component and attempts to solve the percentage of each component for each wavelength on a per pixel basis. This function exists in ZEN Black, as well as other software packages, can be used to discriminate between channels. However, a significant practical limitation of this approach is that the fluorescence signal is quite often non-linear which confounds linear unmixing algorithms which depend on a linear

decomposition of spectral signals. A critical issue with linear unmixing is that the reference must be exactly correct on a per sample basis requiring highly controlled samples. Several linear unmixing software tools, as well as more sophisticated algorithms, has been tested. None of these methods proved to be adequate. The results of our tests are shown in Appendix A.

To alleviate the inability to handle non-linear samples and inability to re-calibrate on a per sample basis a simplified spectral pixel analysis method is used to identify potential pixels that may contain CTC markers. Initially customizable thresholding is used to eliminate background noise. Following this all pixels are passed through a customizable relative selection algorithm that compares spectral channels for defined relations. In this experiment cells are tested by the relative signal strength between the peak of the EpCAM-positive selection peak channel and the CD45-negative selection peak channel. This approach was taken to ensure that the EpCAM signal is significantly stronger than the negative signal by an order of magnitude as shown in Figure 2.8. The process separates WBCs from potential CTCs on a per pixel basis.

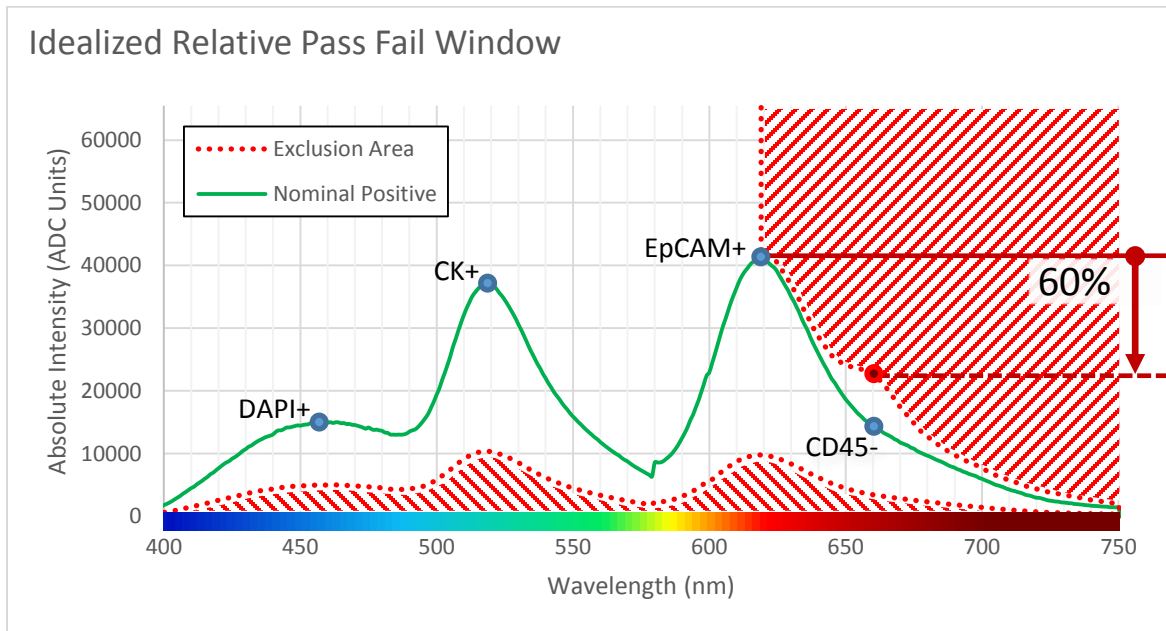


Figure 2.8 Idealized representation of the global per pixel pass/fail spectral thresholding process. In practice control points can be reduced to critical markers shown as blue dots at the precise known peak for the individual dyes. The noise floor is determined through calibration studies as well as the relative intensity specification of 60%. The upper exclusion area is dynamically scaled per pixel to test for CD45 contamination in comparison to the EpCAM peak channel.

After testing positive pixels retain their spectral information while negative pixels are zeroed by the whole image relative intensity test. The positive pixels are then merged into a single image with a blurred DAPI image as a filter to smooth noise while all critical positive channels are summed to improve sensitivity. The merged image is then thresholded using experimentally determined limits, shown above, and a binary dilation is performed to ensure the mask fully encloses the cell volume. This merged image becomes the CTC filter which represents the selection mask used by the following section for CTC identification. To detect all cells, the relative intensity test is skipped and all channels positive and negative are used to identify non-positive cells.

Because this simplified algorithm is highly customizable create a proper calibration profile before use is critical for proper operation. To setup the system the acquisition of images from empirical control testing and startup calibration experiments is required to model the expected background noise and to define the thresholds for each channel. This calibration experiment process is described in Chapter 4. The results of this generalized calibration enables the formal experiment to perform well in most patient samples while allowing for the nonlinearities that may occur in real world patient samples. In real world samples oversaturation, can occur due to variability in patient cells resulting in abnormally high expression vs. most patient cells. However, the developed relative selection process degrades gracefully by becoming increasingly pessimistic rather than failing outright in such oversaturation. This graceful degradation is due to the ratio based test becoming more limiting the more oversaturation occurs with extremely oversaturation almost entirely excluded. This behavior is desirable in accommodating a wider diversity in patient samples while not preferring oversaturated results. Traditional linear un-mixing methods do not work with such non-linear samples, as the most common nonlinearity is oversaturation of the imaging sensor. In typical microscopy oversaturation is controlled by altering settings on a per sample basis but in real world patient samples performing a per sample calibration would be impossible as it is unknown if a sample contains target CTCs.

2.4 Segmentation and Sub-Region Analysis

Segmentation is required to convert raw selection masks into sub-regions for reporting and statistical analysis. Individual cells are analyzed by first segmenting the composite spectral image into multiple regions of interest (ROIs) each containing a single cell, and in some cases, cell clumps and debris. By individually processing selected ROIs, each ROI can be ranked on its likelihood of being a CTC.

Segmentation of the composite image is performed by generating ROI containing individual cells. Our algorithm uses a combination of the cell morphology and spectral data as certain spectral channels such as the cell nucleus can be used to generate mask images to guide separation of cells using spectral data and to extract sub regions rapidly in a highly customizable and resilient manner. Specifically, this task is performed by detecting three classes of pixels that bin any number of spectral channels into a set of binary images for CTC negative pixels (CD45+), positive pixels (EpCAM+,CK+), and nucleus pixels (DAPI) illustrated in Figure 2.9. A nucleus pixel is a signal that is expected in both positive and negative cells and can be used to locate all cells of interest. The segmentation process is also designed to also ignore debris and contaminated antibody clumps by identifying regions such as a cell nucleus stained with a nuclear dye. The positive channels (CK and EpCAM) are the desired target markers and negative channel (CD45) to be markers that are unwanted in positive selections.

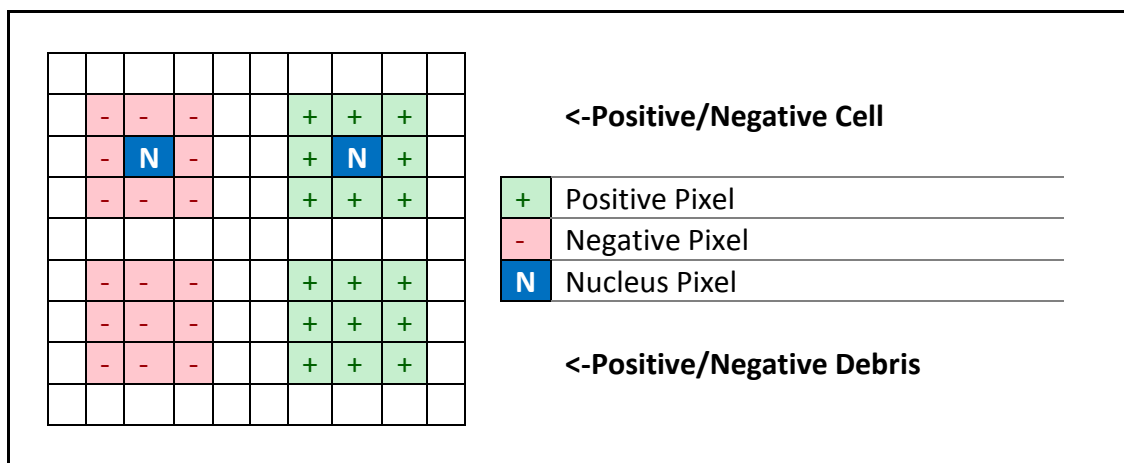


Figure 2.9 Any of the spectral channels can be classified on a per pixel basis as being a possible positive, negative or nucleus pixel. By creating a composite binary image based on nucleus pixels with negative/positive pixels nearby allows for the detection and sorting of cells while rejecting the background and debris (positive/negative).

Use of the nuclear DAPI marker as the nucleus channel, allows downstream segmentation to take advantage of a physically segmented signal, greatly enhances debris and background rejection. As a result of the physically segmented signal, even when cells are physically nearby their nucleuses they will be separated. Segmented regions that lack nucleus signals can be quickly rejected as debris. The use of internal markers such as nuclear stains, along with the low cell density, removes the need for watershed separation algorithms, which often over-segment irregular shaped cells. An additional problem with watershed separation methods is over segmentation with multi-tile images stitched internally on the microscope. The even slight misalignment of the edges of such tiles would always be split by a proper watershed operation. This over-segmentation is undesirable because of the minor random offsets that occur in the tiling process can be ignored if no watershed separation is performed.

The described simplified segmentation method using spectral pixel analysis mask results has several key advantages over traditional analytical strategies. Firstly, the memory footprint required for spectral image clustering and downstream sub region processing is greatly reduced from 3.1GB to 120MB just by classifying pixels and collapsing the image to a fixed set of binary images. Typical un-mixing programs are general purpose and lack such optimizations for biological cell processing. Second, ratiometric data from different spectral channels are used to exclude false positive cells (cells that have undesirably high negative peak channel signals relative to positive signals) compensates for spectral overlap in a manner that remains memory efficient, by processing all pixels within an ROI as a single group. Finally, the process is tolerant of overexposed images which is highly desirable in samples with highly variable biological responses. Due to the high variability in patient samples, compared to cultured cancer cell lines, the property of exposure tolerance is required. The grouped ROI spectral testing is used to determine a cells rank described in section 3.5.

To minimize selection errors due to abnormally shaped cells and debris, morphological filtering is performed in order to exclude extremely irregular non-cell type features (such as debris or fibers). Each segmented objective is analyzed for its circularity. If the tested object is extremely eccentric it is considered to be debris. This filter removes segmented sub-regions that exceed relaxed geometric standards for what would be considered a cell by analyzing the major and minor axis of an assumed ellipse shapes major axis enclosing the longest distance and the minor axis the shortest

distance. The major axis is limited by the experimentally determined norm to remove extremely large cells. The ratio between the major and minor axis is limited to exclude highly excentric ellipses. Other forms of sub region property filtering include minimum and maximum area definitions. This filter removes selections smaller or larger based upon the expected CTC size.

2.5 Automatic Ranking Algorithm

The automatic ranking of segmented cells based on their likelihood of being a CTC is accomplished through a series of distinct phases. In the first phase, outputs from the image processing algorithm are averaged in groups of pixels from the segmented ROIs to produce review graphics and images. Both negative and positive selection targets are averaged across all enclosed pixels within the segmented region, which produces an average spectral response curve. An oversaturation warning counter counts the number of pixels contained within the ROI that exceeds a customizable upper limit per channel. A second phase follows the calculation of average intensities, wherein ROIs are ranked based on their overall intensity within the possible positive and negative cell groups. This ranking process, shown in Figure 2.10, is highly customizable with a number of parameters defined through iterative control testing described in Section 4.1.

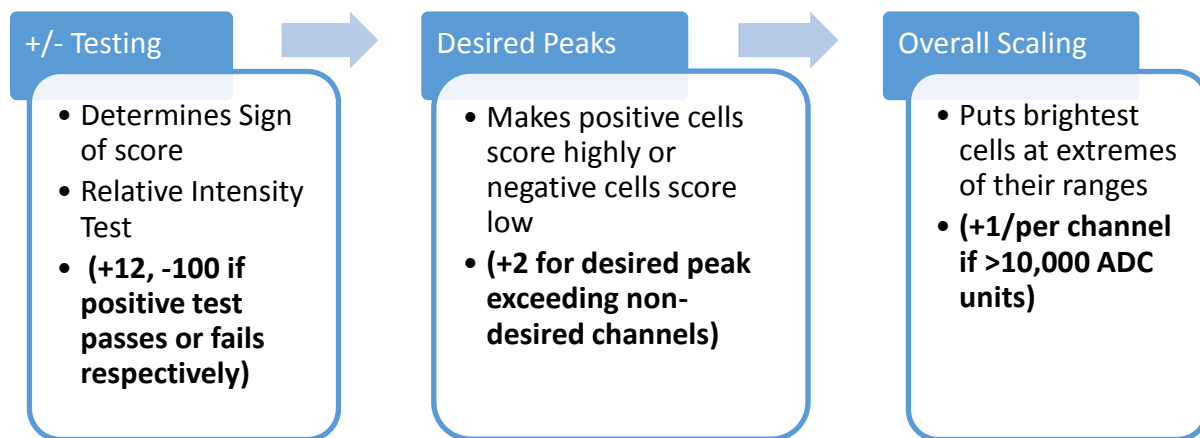


Figure 2.10 Application specific scoring function used to sort results seen in Figure 2.11

The two ordered lists of negative and positive cells are then joined into one result dataset. This arrangement allows for faster manual review and places all strong candidate cells at the start of the dataset. In the final phase, single-cell image outputs and multiple overview thumbnails are generated to ease a manual review of the automated output information. The average spectral

graphs for every cell are converted into two false color intensity graded plots. These plots represent a heat map of the channel intensity per cell as seen in Figure 2.11. This heat map uses the same false color composite algorithm as the color composite image using the averaged per cell spectral information to create an easier to view horizontally spread plot of the cell average intensity per channel and color. Thumbnails of the multiple whole sample layers are also produced for easier identification in sample reports. A pure intensity plot is also possible with no false coloration occurring and can be seen in Figure 2.12. Reviewers would review the sorted results and select cells that they view as truly positive dependent on the specific sample being tested.

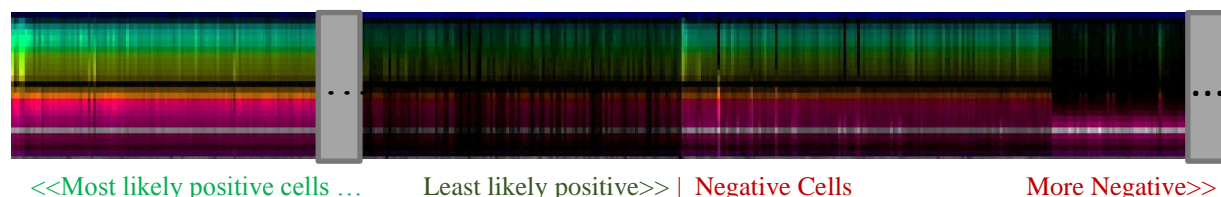


Figure 2.11 Summary graphic overview with false color composite mapped to cell regions with the entire sample represented in one simple graphic. Used as the navigation overview map for review.

Images from the color composite, bright field, and potential CTC filters, images are extracted per cell and placed above a graph containing the spectral average values and pixel count of oversaturated cells. The outline of the selection volume is drawn on the bright field layer for verification of the automatic selection process. These images are then placed into a video file as a series of images in order to package the sequence in a viewable format.

2.6 User Review Process

Although the automated screening process provides an idea of the sample state, a human review of the automated results provides the same quality as a fully manual review process. However, the automated process also offers greatly increased throughput by excluding thousands of undesired cells automatically. The automated process brings the most likely positive cells to the front ranked sorted list greatly expediting the review process. Two modes of review are available for the automated results: a software agnostic video file and custom review software. Both modes of review use the same human readable file collection which can be contained inside a .zip folder.

2.6.1 Simple User Review Process

Table 3.1 lists the output files and common viewers that can be used to review the automated process outputs. By utilizing commonly available formats, the review process does not require any specialized software. Reviewers can use video player controls to sequentially view a file, or they can open image files in any common image viewer/editor to markup the images as required. The text based output file can be opened in any spreadsheet/text editor to access the numerical data from the automated output. The output can also be linked to other software processes in order to provide the physical coordinate information. These common formats allow the output report files to be easily shared and reviewed without the use of proprietary tools. Additional image information is included with a thumbnail of the sample as well as image preview charts that facilitate easier navigation.

Name	Type	Review Purpose
videorpt	MJPEG	Image Sequence of all positive and negative visual results
celldata	CSV	Data file containing all text data on positive and negative results
j/gelly	JPEG	summary heat map image mapped to intensity or FCC
FR	JPEG	full resolution compressed markup images

Table 2.1 List of output file types and their purpose.

Reviewers open the video report file in any video player that supports the common motion JPEG format and then manually record the cell index for target in a frame by frame manner. Video files also contain header frames to ensure the sample information and report metadata are tightly coupled. Frame rates are set to be slow to ensure playback is consistent with one frame per second so each sample is easily viewed even if frame by frame playback is not used. After marking down the cell indexes, reviewers can edit the cell data text file to record their selections. They can also obtain the numerical position and perform statistics on their selections. In addition, common photo editing software can be used to markup the heatmap images and full resolution images in order to generate visual reporting documents. These processes can be streamlined using custom designed review software that is described in the following section.

2.6.2 Graphical User Interface for Human Review

The results files can be easily refined using a simplified report viewer that takes the report container and displays all the relevant information in user friendly manner. This software takes all common file outputs and integrates the video/image/text results. Interactivity of the integrated results allows the reviewer to use a mouse or keyboard to rapidly navigate. Another important aspect, is that the review software is generalized allowing for the MATLAB image processing configuration to be changed without altering the review software. The overall interface is shown in Figure 2.12 with experiment specific markup overlaid. The index/position/area describes the cell index within the ranked sample result list as well as the XY pixel coordinate position in the actual image. The area is listed as the pixels enclosed within the sub region and is outlined in green over the bright field (BF) image. Images of each sub region are captured per result with the false color composite (FCC) used to quickly assess the cell and the bright field (BF) used to identify nearby debris or anomalies. The mask image is the output result of the spectral pixel processing algorithm with bright pixels representing more positive pixels. The spectral graph below the images of the sub region provides the average spectral response of the cell which can be used to positively identify the type of cell as well as see anomalous spectral responses. Starting from the left of the spectral graph the nuclear marker DAPI's signal strength is listed followed by the two CK/EpCAM positive peaks. Finally, the CD45 peak is marked which is where the relative selection tests against any possible residual CD45 expression. The oversaturation of the image data is listed above the spectral graph and allows users to see if a cell has abnormal marker expression. The Gel View on the top right provides the primary user navigation as well as a rapid overview of the entire sample and selects cells based on the mouse position over the gel view. The coloration of this graph can be switched between an intensity only colorization or use the FCC composition method. This gel view has the DAPI channel at the top followed by the CK/EpCAM/Cd45 channel peaks marked by the arrows in Figure 2.12. This overview gel view provides all the spectral result curves for all samples in an easy to review method without the need for large image panels. When a cell is selected, it is marked with a magenta marker at the bottom and the currently selected cell is highlighted with a thin magenta line. Full resolution markup image provides positional information on selected cells which can be useful in single cell capture applications in that it provides a physical map relative to the sample holder. Users can also select and delete already marked cells from this

view. It can also be switched between the FCC/BF/Mask views for the entire sample well. The text summary on the side lists the textual results of the selected cells and is used to save an amended result table with manually screened results. At the bottom right is a visual feedback for numerical navigation where users can enter a number into the program by typing a number anywhere if the program is in focus to quickly jump to a cell index position. This can be useful if another reviewer asks another to look at a cell within a sample by its index number. These navigation methods are synchronized so that if a user navigates using any of the gel view/full resolution markup/text summary/numerical input all the graphical elements will update to match the selection. The user can also reduce the program window horizontally with the window snapping to exclude the text summary first then full resolution image and gel view and finally just the spectral graph. The expansion and contraction of the software allows users with larger monitors to use more of the Gel View at the same time or shrink the program to use alongside other software.

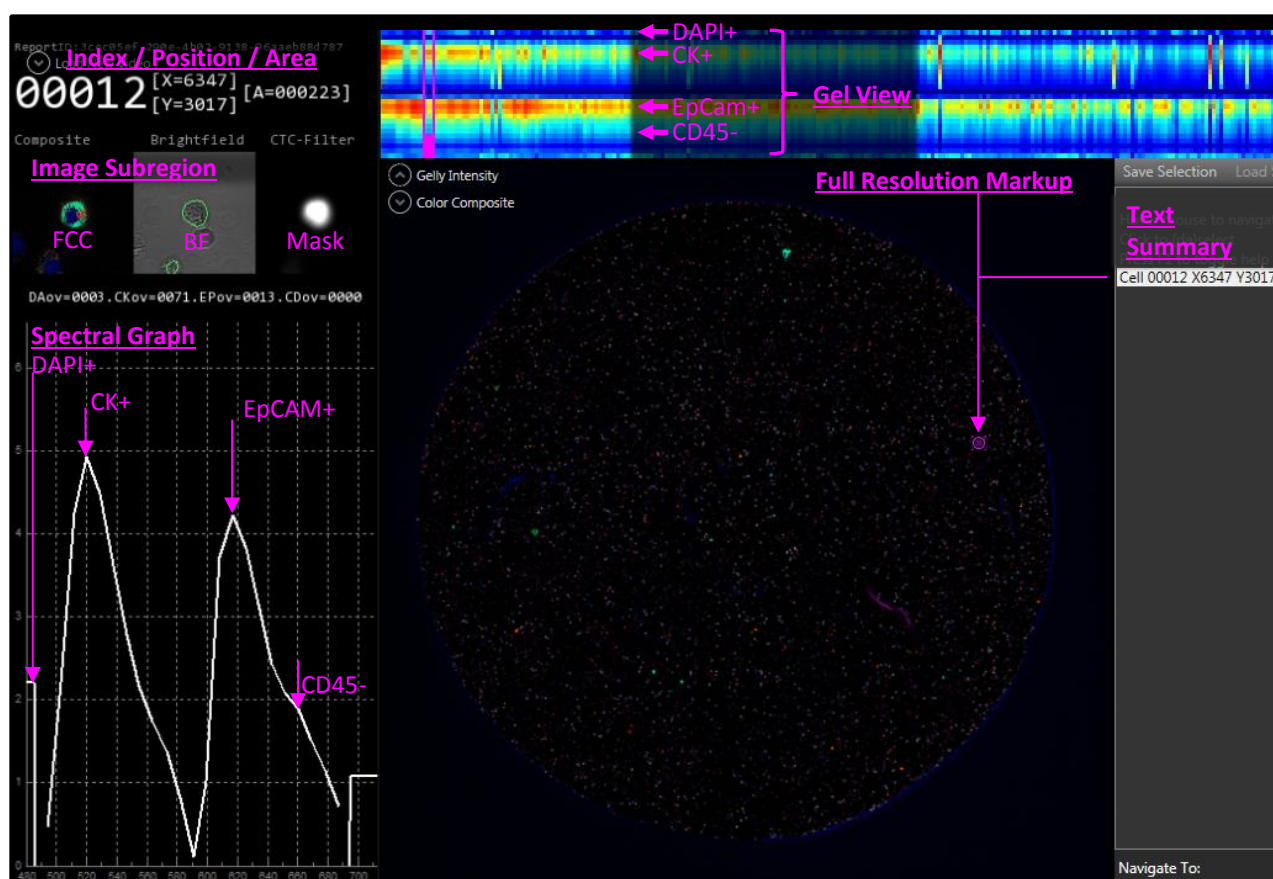


Figure 2.12 Review software with experiment specific markup overlaid added to highlight marker positions as well as major areas of the software.

The review software displays a spectral graph for every cell and highlights its physical position on the thumbnail overview image of the sample. In addition, the intensity graph images provide a rapid navigation tool to enable scrubbing through the report with simple mouse over and keyboard navigation. Users can select cells and generate markings on the intensity plot and the overview image. The selection allows the user to view a reference of the cells already selected. This method is a significant improvement over ZEN (ZEN does not support selecting and rendering a significant number of potential cells on a single image with spectral information enabled). If a reviewer knows a particular cell index they can also type the index in to jump to a particular cell.

After a user has selected the cells, a text summary list can be reviewed quickly before exporting an annotated text file. The annotated text file contains the list of cells that were positively identified during review. Compared to previous manual reviews with the original microscope software, time to select and identify positive cells is greatly improved. In addition, full resolution markup images are produced allowing for users to directly navigate to positively identified cells on the imaged sample plate.

Chapter 3: Experimental Process for CTC Identification

This chapter describes the development of the experimental process for CTC identification using spectral image cytometry. Section 3.1 describes the imaging quality control and calibration process. Section 3.2 describes two manual review processes that serve as the baseline case for CTC identification. Section 3.3 describes the overall experimental workflow for spectral image cytometry software platform developed in this thesis.

3.1 Quality Control and Calibration

Blood samples from healthy donors are used as a negative control to calibrate the microscope system and identify the noise floor of the fluorescence markers for positive selection. Cancer cells from LNCaP and UC13 cell lines are used as positive controls to establish the staining and imaging conditions for CTCs. These healthy donors and cultured cancer cells are used to generate initial calibration data shown in figure 3.1. A potential challenge of using cultured cancer cells as positive control is that patient CTCs can vary wildly from lab cultured homogenous cancer cell lines. Furthermore, blood samples are sourced from terminal cancer patients undergoing experimental treatments that can dramatically alter the composition and status of their blood cells. As a result, some patient samples were excluded from the formal enumeration process because the imaging conditions for those samples were not consistent. Additional initial patient samples were processed in order to aid in an iterative calibration of microscope settings and tuning the preparation of the samples for imaging. These images were not used for counting until the process was optimized to produce stable manual identification results as shown in Figure 3.1 under the iterative process to determine final settings. The formal process used these final settings and processed patient samples with injected blinded negative control samples to verify the process during the experiment.

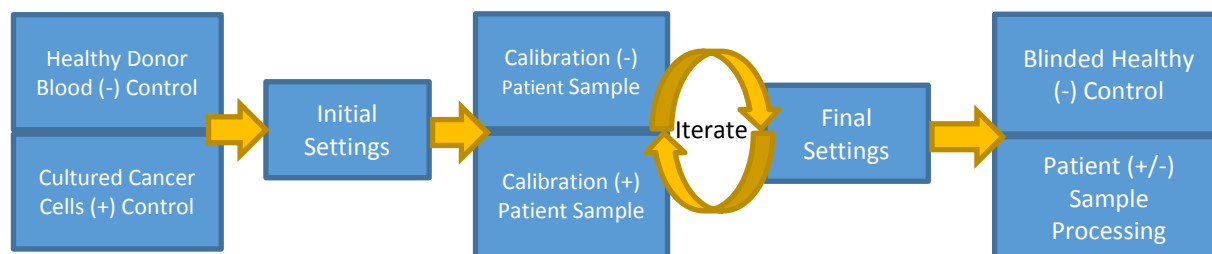


Figure 3.1 Calibration of microscope settings using control samples and initial trial run real world samples leading to a finalized configuration with main experiment verified by injecting blinded negative healthy samples randomly into the patient sample stream to verify workflow process.

Quality control is a critical component of imaging studies with highly variable samples. A number of sample criteria were developed to identify substandard samples. Specifically, since the LSM 780 confocal system captures a single focal plan, it is important to ensure a stable focus is maintained over the entire sample scan and that the cell sample is maintained in a single monolayer. It is further required that the cells do not form confluent layers where many cells are in contact with each other as this risks undesired absorption of emission light from the IF process by non-target cells occluding the optical return path. To ensure a stable focus is maintained we use a confocal sectioning to scan several distinct depth layers present in a sample. This is made possible with the confocal system's pinhole which can create precise slices of a sample while excluding out of focus light and is as shown in Figure 3.2. An occasional problem is the presence of red blood cells, which are not stained by any of the fluorophores. However, an excessive number of RBCs can prevent the formation of a single monolayer of cells to corrupt the fluorescence imaging process which can be detected using the depth scan described.

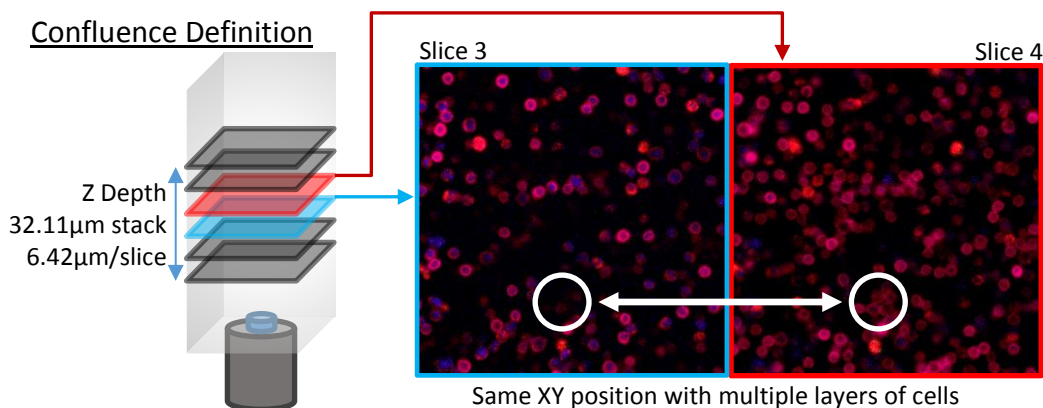


Figure 3.2 Illustration of sample confluence with multiple layers of cells overlaid on each other.

Additional criteria for quality control include that cells should also be free of fluorochrome contamination, which can be detected by extremely intense points of light. Samples are excluded for having microbial contamination which appears in brightfield imaging as smaller than cell organisms. When viewed under live microscopy, microbial contaminations move actively in the stained media where all cells should be dead. These contaminated samples are excluded as the activity of microbial contaminants can alter the expression of IF markers or even consume the target cells. Additionally, negative WBCs must be present in the sample for positive cells to be accurately identified. If all cells appear positive, then a staining failure or sample anomaly is likely and a sample is rejected.

3.2 Workflow for Manual CTC Identification from Spectral Images

3.2.1 Manual Review of Spectral Images

Manual image processing of the multispectral images involves human reviewers selecting and reviewing thousands of cells individually. This task is performed by blinded reviewers using the ZEN Black microscope software using the trained user performing spectral signal processing and morphological evaluation. Due to software performance limitations of ZEN, only a few hundred individually selected cells could be stored at a time. The primary challenge for the completely manual process is its low throughput with the process typically requiring hours to complete.

Depending on sample density, some samples having tens of thousands of cells and it can be difficult for a manual reviewer to accurately screen an entire sample without missing some cells. Because some samples require processing multiple sample wells, due to cell density, manual counting times can increase dramatically; thus reducing the processing ability of lab staff. The review process involves an initial manual pre-screening process to determine which cells might be potential CTCs this involves identifying clearly negative cells as leukocytes by their distinctive shape and coloration. Because this pre-screening does not actually select cells to be evaluated by their spectral content review artifacts can occur between reviewers as initial pre-screening process is performed by the reviewers individual judgement. The time required a single well in a 384 imaging well plate can easily exceed 30 minutes with an average selection speed of approximately 30 cells selected per minute with actual numbers depending on the particular reviewer. Potential cells are often times greater than 1000 cells per well. The following Figure 3.3 illustrates the manual selection process and how users must manually scan across the image selecting potential cells in a number of review tiles. Overall this means the manual review process has a time performance limitation as well as a potential to undercount CTCs. However, the manual process produces the highest quality data and can accurately consider all morphological and marker based criteria serving as a good baseline to compare against.

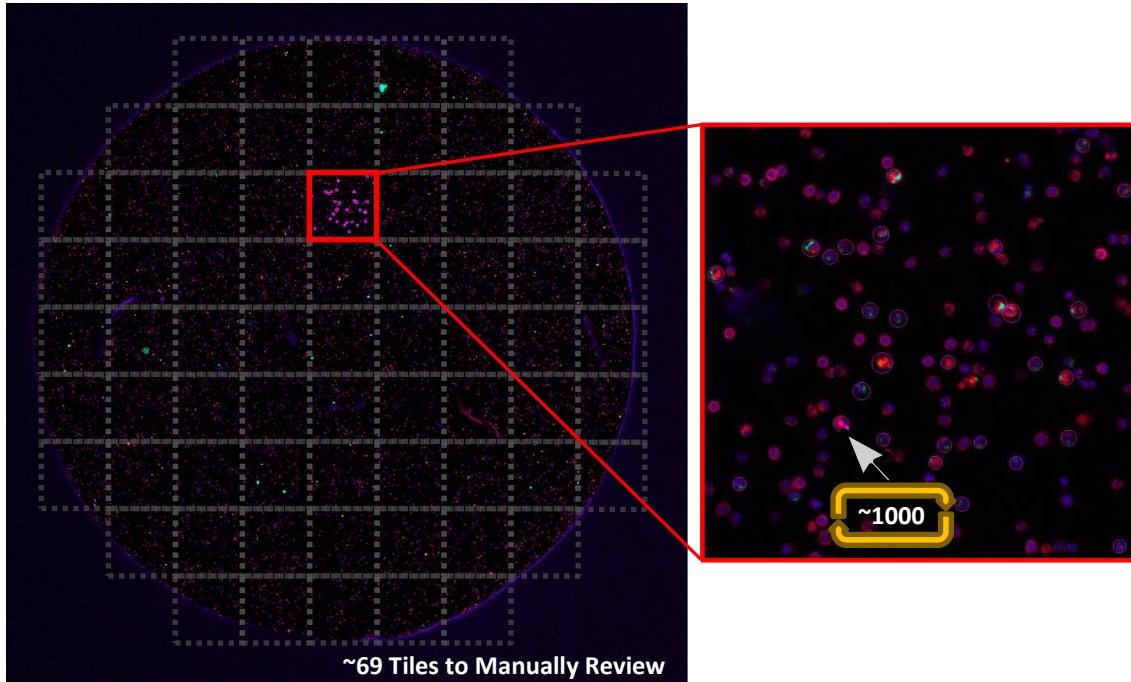


Figure 3.3 Manual review requires user to up-close manually click and drag on every potential cell. The number of tiles a user has to review is dependent on their screen resolution and monitor being used which adds additional variables in the review process.

3.2.2 Manual Review of Standard IF Images

Initial setup testing involved the use of standard IF imaging but using the same experimental setup this process was not successful. To verify that multi-spectral imaging improves performance over standard IF imaging a simulation using the same data was performed to verify the expected improvement in performance. Because multispectral acquisition provides additional information compared to standard IF acquisition down sampling and emulation of Standard immunofluorescence (IF) data is possible by extracting emulated single channel images from a spectral stack. An advantage of standard IF images is that they can be reviewed rapidly. However due to the close placement of fluorescence markers, bleed over from nearby wavelengths may result in significant number of false positive. Standard IF images were generated by flattening spectral image cube into 4 channels based on common real world filters used in standard inverted microscopes.

Idealized graphs below (Figure 3.4 to 3.6) use reference spectra to illustrate the difference between standard IF filter and a smooth spectral output. The filter based intensity plots provide significantly less information about the actual sample response when compared with a spectral capture that can positively identify the presence of negative fluorophores. In actual experiments variations between and within samples causes additional variations for which spectral capture can detect and compensate. Instances have been observed in practice whereby unexpected spectral responses occur with no corresponding dye. A filter based system would create the false appearance of two or more dyes near the unexpected fluorescence emission. In the spectral system unexpected emissions are also excluded.

In the positive case, shown in Figure 3.4, the spectral scan provides clear peaks within the expected locations. On the other hand, filter based analysis shows significant signals in all channels. Another benefit to the spectral scanning system is additional dyes can be added without causing significant disruption to the workflow. Typically, with filter based systems for every additional dye scanning time increases multiplicatively. Furthermore, imaging requires multiple exposures with various filter elements being switched in and out. Conversely, by spectral scanning with a single exposure, many wavelength bins can be captured with sufficient resolution to show the actual peak wavelengths as well as contaminants. In actual experiments, due to non-ideal conditions, additional selection criteria can be added. An example is testing for peak broadening which is a sign of extreme overexposure or physical contamination. In addition to increased discrimination, a spectral scanning system allows for the capture of all spectral channels simultaneously. A spectral scanning system also allows for increased sensitivity as all received light is captured while in standard IF filters reject light from portions of the spectrum.

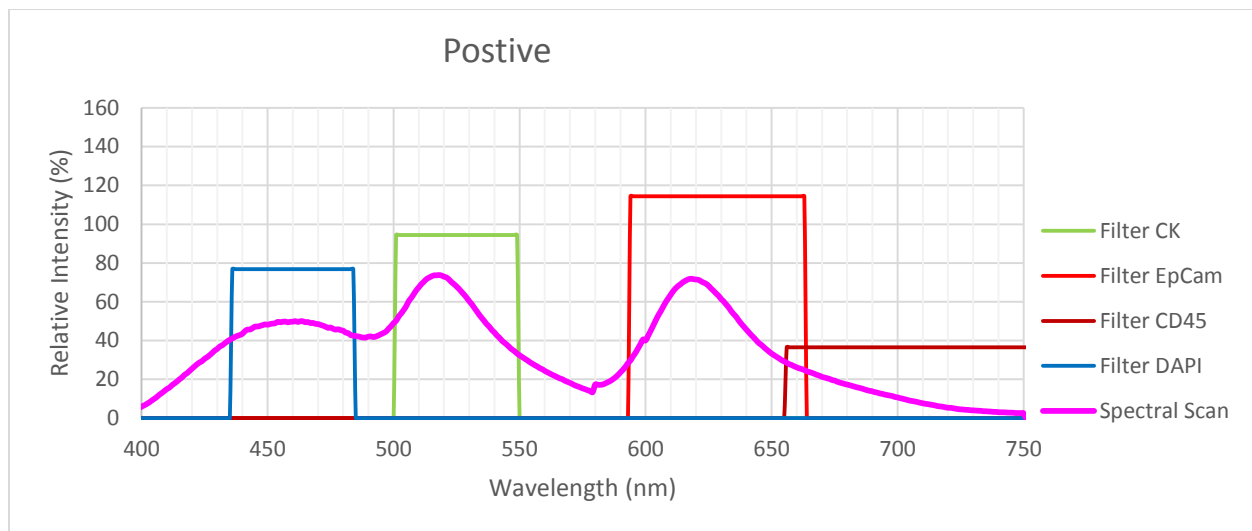


Figure 3.4 Idealized spectral graph showing how clear positive cells exhibit clearly identifiable peaks while filter based processing shows intensities in all channels with the peak location unknown.

In an abnormal false-positive case, spectral scanning also provides a clear indication of even slightly negative cells with positive signals mixed as shown in Figure 3.5. The emulated filter based system appears similar to the positive case with overall signals increased for both CD45 and EpCAM.

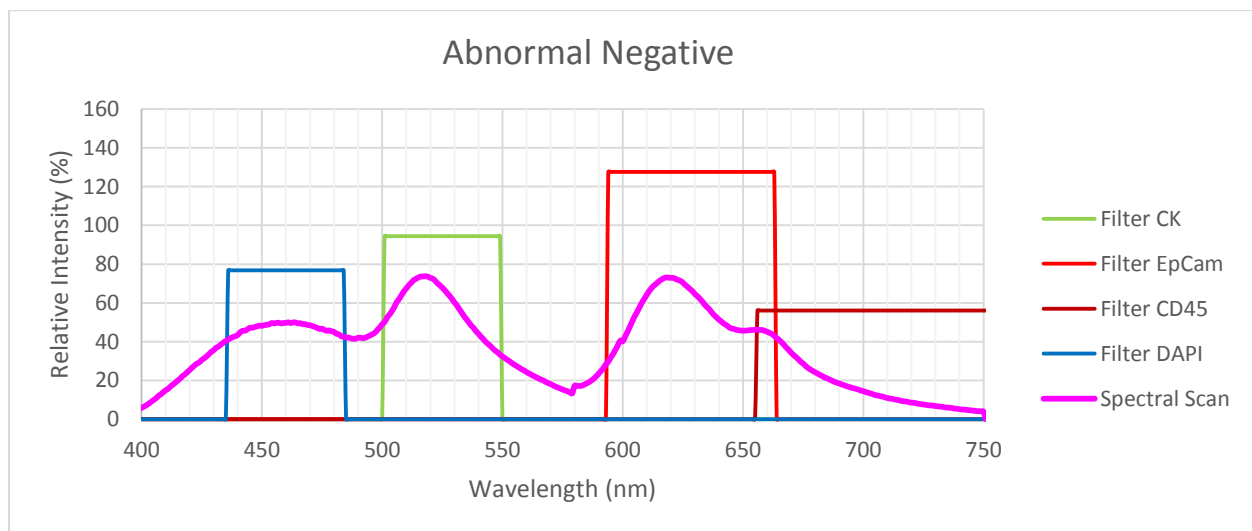


Figure 3.5 Abnormal negative cell with all signals present and a weak CD45 signal. Spectral scanning shows this bump clearly while filter based processing is far more unclear with both the EpCAM and CD45 signal being mixed together.

Finally, in the negative case shown in Figure 3.6, both methods are fairly clear. However, the filter based method still shows some signal in positive channels whilst the spectral data would show the peaks clearly. In practice, filter based systems are not able to create perfectly clean filter bands because there is always some additional roll off compared to the idealized plot shown.

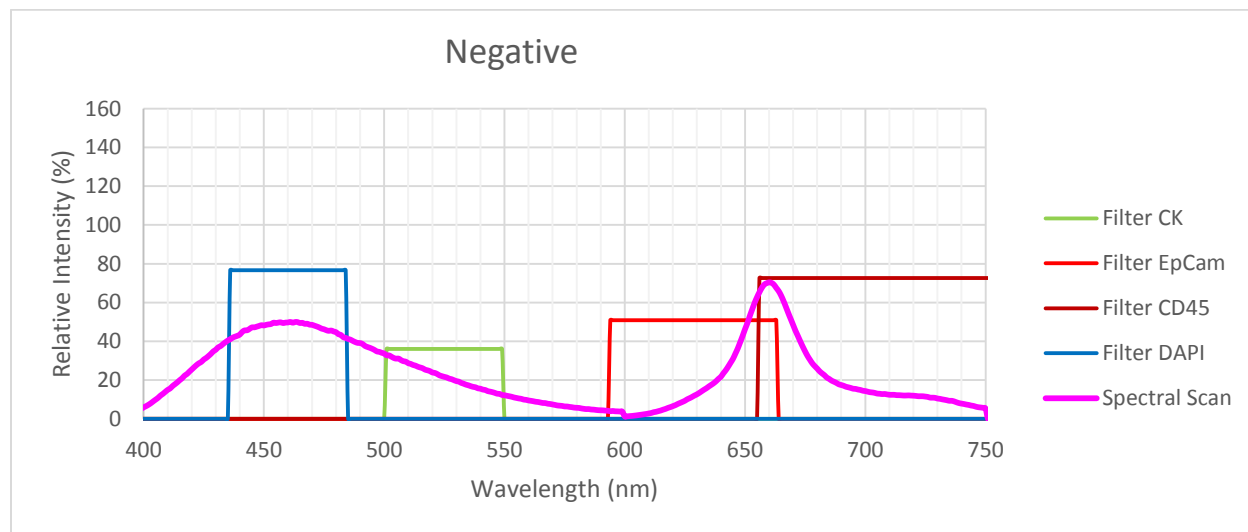


Figure 3.6 Idealized graphs of the greatest difference between filter based imaging and full spectral imaging with the use of nearby dyes. Spectral imaging can support the addition of a fifth dye in the 550-600nm range (yellow) without issue while in the filter based system this would cause additional bleed over.

Standard filter based imaging can be improved by using extremely tight emission and excitation filters in combination with illumination specific scanning. This requires multiple exposures and increased imaging time due to the loss of sensitivity from tight filter bands restricting photon capture. Standard non-spectral confocal systems can also capture arbitrarily tight wavelength bands but multiply the scan time and exposure requirements by the number of overlapping dye sets. By reducing the received light and requiring multiple exposures, the required light exposure increases dramatically. This can also negatively affect the sample quality and survival rate due to photo degradation of the dyes and cells being imaged. Another advantage of a spectral scanning system is its ability to emulate filter based capture by defining digital filter sets in post processes.

3.3 Workflow for Semi-Automated CTC Identification from Spectral Images

For validation of this analytical system, tumor cells were enriched using the microfluidic ratchet deformability-based sorting mechanism, developed in our laboratory [41]. Briefly, whole blood

obtained from patients was flowed through an array of microscale funnels, where the cells were vertically fractionated based on cell deformability. Cell fractions containing the more rigid leukocytes and tumor cells were obtained as a liquid suspension and were immunostained for CD45, CK and EpCAM and stained with the DAPI nucleic acid dye. Cells were stored in suspension at 4°C overnight, for processing on the next day under ideal conditions. Just prior to imaging, cells were seeded into optical glass bottomed 384 well plates with a diameter of approximately two millimeters and collected to the bottom of the plate by centrifugation. The wells were visually inspected and, if cell density was too high, the suspension was diluted and split to additional sample wells.

With the sample carrier plate loaded onto the microscope, the well was centered and calibrated to the automated stage to image multiple wells as tiled image sets. Microscope profiles specific to the experiment are selected and a large image was automatically captured. Initial quality control measures on sample images were performed at this point and if significant errors have occurred the operator had the opportunity to re-plate or recapture the image after correcting for any errors in the setup. The overall workflow is depicted in Figure 3.7, with this thesis work focused on the scanning process, automated analysis, and assisted review.

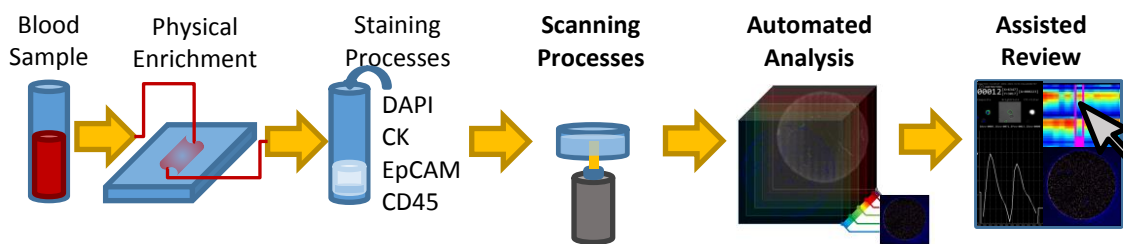


Figure 3.7 Experimental workflow from patient sample input to final data output.

After the images for all samples were captured they are saved to the local drive and network backup. These images are then processed using the automated image cytometry algorithm. This produces a report for manual verification. An approximate automated count is provided as an interim number. Assisted review is then performed to obtain the final CTC count.

Chapter 4: Results and Evaluation

This chapter describes the validation of the thesis and overall results. Section 4.1 describes the validation of selection parameters in a sensitivity vs. specificity test. Section 4.2 presents the results of a comparison between the gold standard manual analysis, emulated manual IF, and this thesis's developed semi-automated process. Finally, section 4.3 presents the processing performance and system requirements.

4.1 Sensitivity vs. Specificity

The Receiver Operator curve (RoC) illustrates the performance of a selection algorithm versus a random selection process. In order to plot a RoC the ground truth, known as the true positives, is required which is obtained from manual review of the results. False positives are any automatically labeled positive cell which are not part of the true positive list. Finally, selection parameters must be varied to plot the sensitivity versus specificity. Sensitivity is the ability for the algorithm to detect true positive cells. This is calculated using true positives selected through manual review relative to the automated algorithm's positive selections shown in Equation 1.1. Specificity is the ratio of false positives vs. automated algorithm selected negatives as shown in equation 1.2. The diagonal in a RoC curve indicates a perfectly random selection process with a perfect algorithm having a 100% true positive rate with a 0% false positive rate forming a corner shaped curve.

$$\text{Sensitivity} = \text{True Positive Rate (TPR)} = \frac{\Sigma \text{True Positive}}{\Sigma \text{Algorithm Positive}} \quad (1.1)$$

$$\text{Specificity} = \text{False Positive Rate (FPR)} = \frac{\Sigma \text{False Positive}}{\Sigma \text{Algorithm Negative}} \quad (1.2)$$

The RoC can also be used to demonstrate how the tuning of the selection parameters could improve selection performance. In addition the RoC demonstrates the typical tradeoff between capture all positive results with false positives included or having no false positive results but missing true positives. This results in an optimization process that can reduce the number of automatically selected potentially positive CTC cells without discarding any true positive cells. This is a critical focus as reducing true positive selection count would prevent any manual reviewer from detecting

cells pre-filtered by automated processes. This optimization target allows the automated method to augment the standard manual review process by removing many tedious cell selections and creating uniform selection criteria without removing true positive cells.

4.1.1 EpCAM vs. CD45 Ratio RoC

The proximity of the industry standard Alexa 594 dye bound to EpCAM and the APC dye bound to markers represents a major challenge to discriminating closely spaced markers. Because the study of multiple positive markers is desired only one channel can be used for negative selection. However properly detecting this negative result is critical. Both the EpCAM (Alexa 594) and CD45 (APC) markers have reference spectra described previously with peaks in specific wavelengths. The difference between the positive EpCAM peak and negative CD45 peak is one of the main ratio-based selection criteria and this ratio shows a strong correlation to selection efficiency and as seen in Figure 4.1. This ratio is defined as the absolute ADC units for EpCAM's expected peak over the absolute ADC units for CD45 peak also described earlier in pixel processing sub-section.

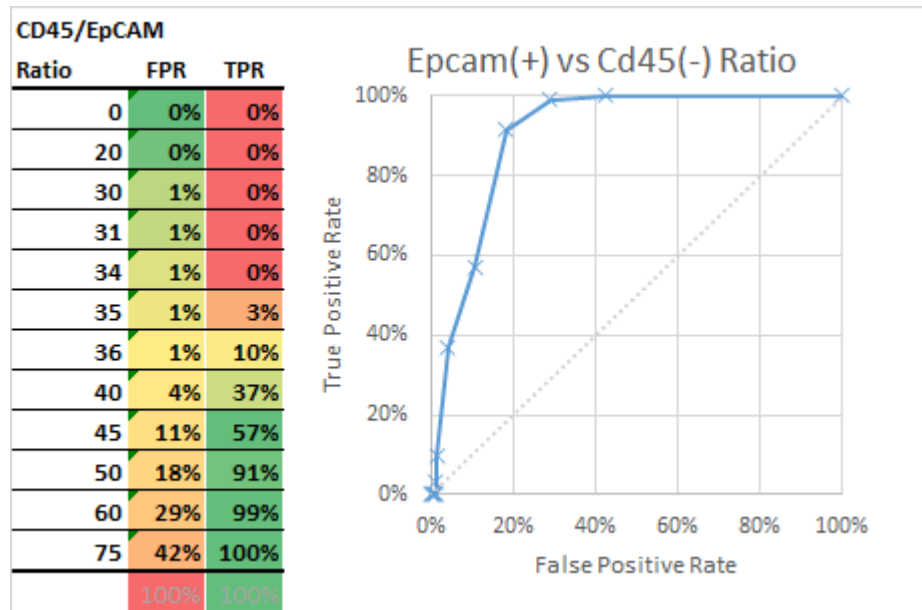


Figure 4.1 Cell ranking RoC plot illustrating reduction in false positive rate with increasingly stringent ranking cutoff.

Physically the dye the EpCAM marker is bound to (Alexa 594) has a theoretical lower limit on the ratio between its peak and where a CD45 (APC) peak would be. Anything below this limit would

not be considered EpCAM positive. This can provide another verification of the ratio-based analysis. The reference spectra of Alexa has a residual ratio of 35 which is illustrated in the actual data. There are no positive results found below this ratio as the experimental true positive rate is 0% below a ratio of 35. Because the idealized spectra does not account for background/noise or the presence of other dyes, the specific optical setup utilizing the ratio-based selections requires additional criteria to further improve performance.

4.1.2 Cell Rank RoC

The cell ranking selection system was intended to sort cells by their likelihood of being a true positive. The initial cutoff was selected to be extremely relaxed with only clearly negative cells being rejected. As seen in Figure 4.2 the RoC curve shows a significant improvement in false positive rates by increasing the score limit to 27. Because the negative population within samples is so large a reduction in the false positive rate greatly reduces the number of potentially positive cells displayed for final review.

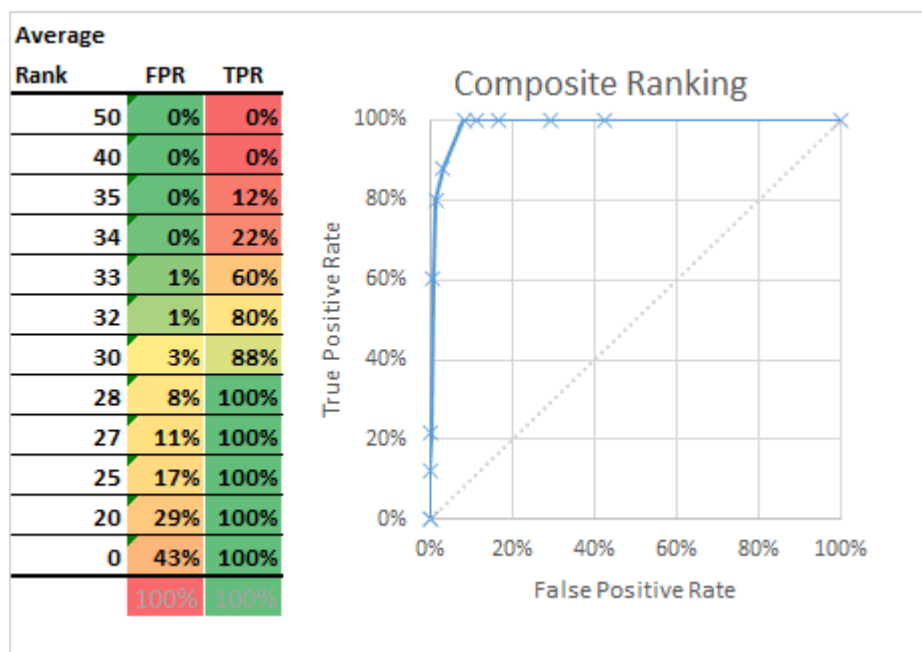


Figure 4.2 Cell ranking RoC plot illustrating reduction in false positive rate with increasingly stringent ranking cutoff.

This composite score, which is based off the spectral intensity cutoffs and relative selection algorithms, allows for greatly improved selection performance by eliminating weakly positive cells that reviewers do not consider as positive. Furthermore, using the composite score also allows for sorting of results more effectively based on multiple parameters.

4.2 CTC Identification Results

To compare the performance of the fully manual spectral analysis against the semi-automatic method the ten samples blindly selected from existing prostate and bladder cancer patient results. In all samples analyzed the semi-automatic method this thesis developed closely match or slightly exceeded the counts found in direct manual review as shown in Figure 4.3. Due to the semi-automated nature of the process the slight increases are likely due to the improved consistency in the selection process. This is achieved by automated analysis of every cell and sorting them before review. Moreover, review can be done concurrently while sample imaging is being processed.

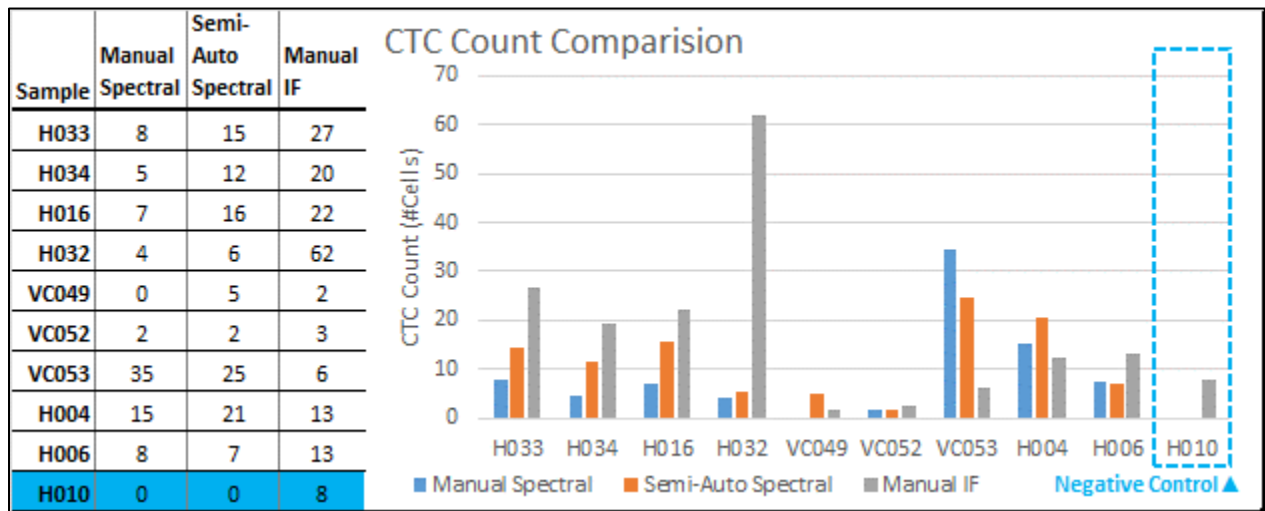


Figure 4.3 Comparison between the three methods of analysis. H010 is the negative control injected into the review process with only the manual IF method failing to exclude the negative control sample.

The semi-automated method passes the control sample without issue. Also this method provides important feedback on sample quality by showing side by side views of the fluorescence image. Visual quality control for imagery is taken from the bright field layer. Users can also verify that the selection of cells was correctly outlined by the automated method, by observing the actual cell

outline regardless of its fluorescence response. In contrast, the manual standard IF method does not successfully discriminate for the negative control sample H010. Furthermore, manual standard IF has high variability when compared to the semi-automated and manual review process.

In sample number H032 the manual IF method had irregular results, with one reviewer's result excluded as an outlier due to the count exceeding 500% relative to the other reviewer results. Sample VC049 has a zero count for the manual spectral review with a small count for both the semi-automatic and manual IF process. This is likely due to difficulty in testing every cell in manual spectral review which sometimes results in under selection of cells.

Overall results from the semi-automatic spectral method are comparable to the manual spectral review. An additional benefit is that every reviewer will review the same cells in the same order, unlike the manual method which does not guarantee all cells of interest are being reviewed.

4.3 Performance and Ease of Use

The overall run time results are broken down by: image conversion and metadata handling; automatic image cytometry process; and report generation all shown in Figure 4.4. The images initial conversion typically takes one minute per image with no further conversions needed. Because this process is completely automatic, no user interaction is needed. This process can occur on a server in order to batch convert incoming files. Additionally, the automatic image processing method accepts meta-data separated files into specific recognition configurations and produces output report files dependent on the total number of cells within a sample. Typical samples containing less than ten thousand cells and take about five minutes to process per sample. Atypical samples containing more than ten thousand cells can exceed the target processing time of 5 minutes per sample. This is due to the image processing producing outputs for all cells regardless if they are negative or positive. The output is an ordered list of most likely positive then most negative cells. Average processing times are within the microscope imaging time tolerances allowing for most samples to be processed in between acquisition steps, thereby increasing the operator's productivity.

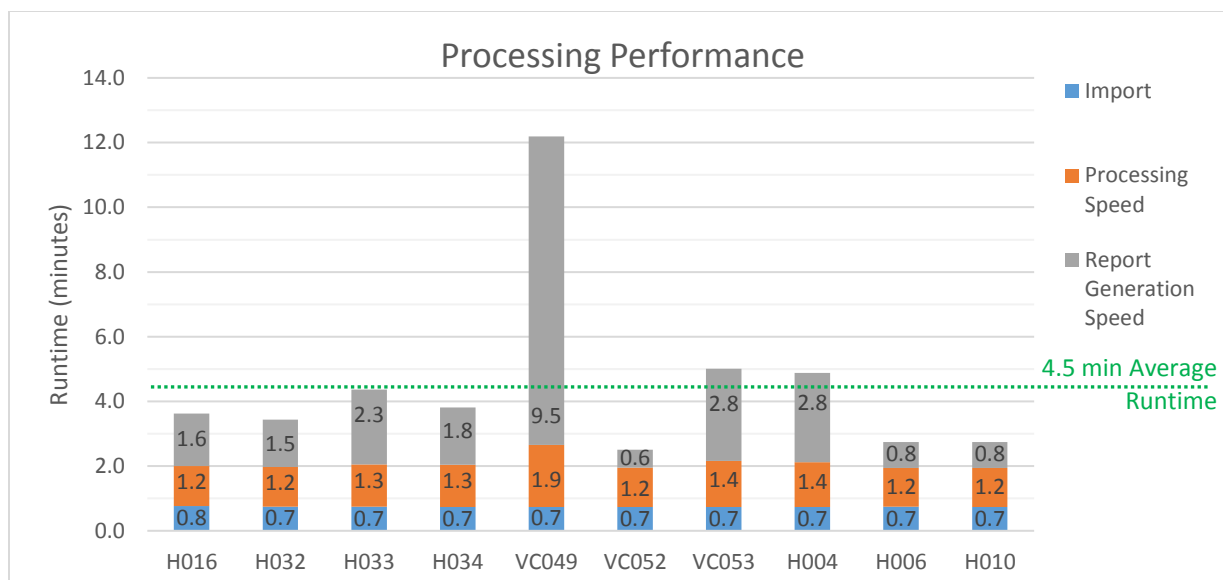


Figure 4.4 processing performance of the automated algorithm split by image file conversion, image processing, and report generation. A majority of the processing time is spent encoding video output report files. In addition the report generation is sensitive to the number of cells being processed.

Processing requirements are limited by the universal file import conversion provided by a pre-existing medical imaging library. The library written in java requires a continuous block of memory for initial conversion processes. As a result the software only uses native MATLAB data files and requires an intermediate step to decouple the conversion software from the image processing. This allows for a separate computer or server to process incoming files for fast loading processing by any analysis programs. The processing algorithm operates with a 26 channel ~60megapixel per layer image requiring approximately 8 gigabytes of system memory.

Because image processing and review is rapid enough to operate in parallel with imaging tasks, operators are able to promptly receive feedback on their experiment quality. Furthermore, preliminary results are produced while they are still capturing microscope image data.

By employing disk streaming, report viewing requires a fraction of the memory used for processing because no further processing is done on the output results within the review software. This is a more user-friendly approach when compared to the previous manual review process which requires users to open the full gigapixel spectral image within the proprietary ZEN microscope software.

Ease of use was one of the primary research objectives of the automated algorithm. This primary goal is achieved by creating easy to use software and simplifying review. Ease of use is validated by an open output format with simple user interface design. The final workflow improvement of the software is greatly reduced average analysis times. The streamlined workflow also reduces manual review to a selection of a few handful of pre-filtered images with no requirement to select or adjust the image algorithm prior to use.

The configurability of the algorithm allows for multiple experiments to use the same software workflow as well as for in process adjustments, subsequently improving overall results. This is verified through process parameter optimization and validation of increasing selection performance described previously in the RoC curves.

Chapter 5: Conclusion

5.1 Summary

The spectral image cytometry software platform described in this thesis identifies rare CTCs from background cells through a highly efficient and semi-automated process. This process is comprised of the image acquisition, automated pre-screening, and assisted manual review process. Overall processing times are greatly improved in the semi-automated workflow. Because the process leverages human reviewers for the critical final selection process, the quality of results is improved over purely manual selection processes. Purely manual selection processes tend to overwork reviewers with tedious and repetitive tasks which are wholly eliminated by this algorithm's automation.

5.2 Statement of Impact

CTC image cytometry is an important process that is challenged by biologically inherent variability and unknown properties. However, CTC image cytometry is enhanced by the use of semi-automatic pre-screening software processing.

This thesis developed a spectral image cytometry software that processes input data from proprietary microscope formats into usable common files for a wide applicability. Automated image processing algorithms successfully segment and identify potential CTCs from gigapixel large spectral image cubes. This data was successfully used to produce an automatic ranking and human readable report for review.

The overall process performed similarly to the previous gold standard fully manual review process while providing a possible improvement in selection performance. The software runtime speed required to generate human readable files was shorter on average than the instrument capture time which allows for an optimal workflow process whereby experiments can be processed and reviewed while acquisition is occurring.

5.3 Future Work

Potential future improvements to the process include, the configuration of the systems requires editing the process configuration files directly. Currently, without any guided setup configuration software, users must be trained before this process can be used in other experiment setups. A guided graphical user interface to the automated algorithm configuration process would greatly improve the ability of the software to be adapted to other projects reducing training requirements.

Currently, metadata from the proprietary microscope are not directly presented within the output report files. The metadata could be integrated to create a single report file for all review requirements. Existing manual review processes can result in experimental errors in file handling processes. Streamlining the manual review process can eliminate such errors.

Fixed processing time targets can also be adopted to allow for the subsampling of negative population selections. Processing all negative cells composes the bulk of report building processing time. This issue can be further optimized by intelligent negative cell skipping and other performance optimizations.

Bibliography

- [1] N. Aceto, M. Toner, S. Maheswaran, and D. A. Haber, “En Route to Metastasis: Circulating Tumor Cell Clusters and Epithelial-to-Mesenchymal Transition,” *Trends Cancer*, vol. 1, no. 1, pp. 44–52, Sep. 2015.
- [2] M. Cristofanilli *et al.*, “Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer,” *N. Engl. J. Med.*, vol. 351, no. 8, pp. 781–791, Aug. 2004.
- [3] D. C. Danila *et al.*, “Circulating tumor cell number and prognosis in progressive castration-resistant prostate cancer,” *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 13, no. 23, pp. 7053–7058, Dec. 2007.
- [4] A. Poveda *et al.*, “Circulating tumor cells predict progression free survival and overall survival in patients with relapsed/recurrent advanced ovarian cancer,” *Gynecol. Oncol.*, vol. 122, no. 3, pp. 567–572, Sep. 2011.
- [5] S. Nagrath *et al.*, “Isolation of rare circulating tumour cells in cancer patients by microchip technology,” *Nature*, vol. 450, no. 7173, pp. 1235–1239, Dec. 2007.
- [6] E. T. Keller and J. Brown, “Prostate cancer bone metastases promote both osteolytic and osteoblastic activity,” *J. Cell. Biochem.*, vol. 91, no. 4, pp. 718–729, Mar. 2004.
- [7] D. Marrinucci *et al.*, “Cytomorphology of circulating colorectal tumor cells: a small case series,” *J. Oncol.*, vol. 2010, p. 861341, 2010.
- [8] D. Marrinucci *et al.*, “Case study of the morphologic variation of circulating tumor cells,” *Hum. Pathol.*, vol. 38, no. 3, pp. 514–519, Mar. 2007.
- [9] M. Yu, S. Stott, M. Toner, S. Maheswaran, and D. A. Haber, “Circulating tumor cells: approaches to isolation and characterization,” *J. Cell Biol.*, vol. 192, no. 3, pp. 373–382, Feb. 2011.
- [10] B. Hong and Y. Zu, “Detecting circulating tumor cells: current challenges and new trends,” *Theranostics*, vol. 3, no. 6, pp. 377–394, 2013.
- [11] A. J. Armstrong *et al.*, “Circulating tumor cells from patients with advanced prostate and breast cancer display both epithelial and mesenchymal markers,” *Mol. Cancer Res. MCR*, vol. 9, no. 8, pp. 997–1007, Aug. 2011.
- [12] S. A. Joosse *et al.*, “Changes in keratin expression during metastatic progression of breast cancer: impact on the detection of circulating tumor cells,” *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 18, no. 4, pp. 993–1003, Feb. 2012.
- [13] C. Raimondi *et al.*, “Epithelial-mesenchymal transition and stemness features in circulating tumor cells from breast cancer patients,” *Breast Cancer Res. Treat.*, vol. 130, no. 2, pp. 449–455, Nov. 2011.
- [14] B. Willipinski-Stapelfeldt *et al.*, “Changes in Cytoskeletal Protein Composition Indicative of an Epithelial-Mesenchymal Transition in Human Micrometastatic and Primary Breast Carcinoma Cells,” *Am. Assoc. Cancer Res.*, vol. 11, no. 22, pp. 8006–8014, Nov. 2005.

- [15] B. Aktas, M. Tewes, T. Fehm, S. Hauch, R. Kimmig, and S. Kasimir-Bauer, "Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients," *Breast Cancer Res. BCR*, vol. 11, no. 4, p. R46, 2009.
- [16] A. Jaggupilli and E. Elkord, "Significance of CD44 and CD24 as cancer stem cell markers: an enduring ambiguity," *Clin. Dev. Immunol.*, vol. 2012, p. 708036, 2012.
- [17] V. Paradis *et al.*, "De novo expression of CD44 in prostate carcinoma is correlated with systemic dissemination of prostate cancer," *J. Clin. Pathol.*, vol. 51, no. 11, pp. 798–802, Nov. 1998.
- [18] E. Racila *et al.*, "Detection and characterization of carcinoma cells in the blood," *Proc. Natl. Acad. Sci.*, vol. 95, no. 8, pp. 4589–4594, Apr. 1998.
- [19] S. D. Mikolajczyk *et al.*, "Detection of EpCAM-Negative and Cytokeratin-Negative Circulating Tumor Cells in Peripheral Blood," *J. Oncol.*, vol. 2011, p. 252361, 2011.
- [20] B. J. Kirby *et al.*, "Functional characterization of circulating tumor cells with a prostate-cancer-specific microfluidic device," *PLoS One*, vol. 7, no. 4, p. e35976, 2012.
- [21] S. Park *et al.*, "Morphological Differences between Circulating Tumor Cells from Prostate Cancer Patients and Cultured Prostate Cancer Cells," *PLoS ONE*, vol. 9, no. 1, p. e85264, Jan. 2014.
- [22] D. Marrinucci *et al.*, "Fluid biopsy in patients with metastatic prostate, pancreatic and breast cancers," *Phys. Biol.*, vol. 9, no. 1, p. 016003, Feb. 2012.
- [23] J. Kraan *et al.*, "External quality assurance of circulating tumor cell enumeration using the CellSearch(®) system: a feasibility study," *Cytometry B Clin. Cytom.*, vol. 80, no. 2, pp. 112–118, Mar. 2011.
- [24] A. G. J. Tibbe, M. C. Miller, and L. W. M. M. Terstappen, "Statistical considerations for enumeration of circulating tumor cells," *Cytometry A*, vol. 71A, no. 3, pp. 154–162, Mar. 2007.
- [25] P. T. Went *et al.*, "Frequent EpCam protein expression in human carcinomas," *Hum. Pathol.*, vol. 35, no. 1, pp. 122–128, Jan. 2004.
- [26] T. E. Witzig *et al.*, "Detection of circulating cytokeratin-positive cells in the blood of breast cancer patients using immunomagnetic enrichment and digital microscopy," *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 8, no. 5, pp. 1085–1091, May 2002.
- [27] "Nikon MicroscopyU | Fluorescence Microscopy | Introduction." [Online]. Available: <http://www.microscopyu.com/articles/fluorescence/fluorescenceintro.html>. [Accessed: 20-Apr-2016].
- [28] D. Coling and B. Kachar, "Theory and application of fluorescence microscopy," *Curr. Protoc. Neurosci.*, pp. 2–1, 2001.
- [29] T. M. Gorges, I. Tinhofer, M. Drosch, L. Röse, T. M. Zollner, and T. Krahn, "Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition," *BMC Cancer*, vol. 12, p. 178, May 2012.

- [30] M. Wendel *et al.*, “Fluid biopsy for Circulating Tumor Cell identification in Patients with early and late stage Non-Small Cell Lung Cancer; a glimpse into lung cancer biology,” *Phys. Biol.*, vol. 9, no. 1, p. 016005, Feb. 2012.
- [31] S. L. Werner *et al.*, “Analytical Validation and Capabilities of the Epic CTC Platform: Enrichment-Free Circulating Tumour Cell Detection and Characterization,” *J. Circ. Biomark.*, p. 1, 2015.
- [32] S. Chen *et al.*, “Recent Advances in Morphological Cell Image Analysis, Recent Advances in Morphological Cell Image Analysis,” *Comput. Math. Methods Med. Comput. Math. Methods Med.*, vol. 2012, 2012, p. e101536, Jan. 2012.
- [33] P. Chu, E. Wu, and L. M. Weiss, “Cytokeratin 7 and Cytokeratin 20 Expression in Epithelial Neoplasms: A Survey of 435 Cases,” *Mod. Pathol.*, vol. 13, no. 9, pp. 962–972, Sep. 2000.
- [34] V. Barak, H. Goike, K. W. Panaretakis, and R. Einarsson, “Clinical utility of cytokeratins as tumor markers,” *Clin. Biochem.*, vol. 37, no. 7, pp. 529–540, Jul. 2004.
- [35] N. Atale, S. Gupta, U. C. S. Yadav, and V. Rani, “Cell-death assessment by fluorescent and nonfluorescent cytosolic and nuclear staining techniques,” *J. Microsc.*, vol. 255, no. 1, pp. 7–19, Jul. 2014.
- [36] H. K. Lin *et al.*, “Portable filter-based microdevice for detection and characterization of circulating tumor cells,” *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 16, no. 20, pp. 5011–5018, Oct. 2010.
- [37] S. Zheng *et al.*, “3D microfilter device for viable circulating tumor cell (CTC) enrichment from blood,” *Biomed. Microdevices*, vol. 13, no. 1, pp. 203–213, Feb. 2011.
- [38] F. A. W. Coumans, G. van Dalum, M. Beck, and L. W. M. M. Terstappen, “Filter characteristics influencing circulating tumor cell enrichment from whole blood,” *PloS One*, vol. 8, no. 4, p. e61770, 2013.
- [39] S. T. Ligthart *et al.*, “Circulating Tumor Cells Count and Morphological Features in Breast, Colorectal and Prostate Cancer,” *PloS One*, vol. 8, no. 6, p. e67148, 2013.
- [40] S. M. McFaul, B. K. Lin, and H. Ma, “Cell separation based on size and deformability using microfluidic funnel ratchets,” *Lab. Chip*, vol. 12, no. 13, pp. 2369–2376, Jul. 2012.
- [41] E. S. Park *et al.*, “Continuous Flow Deformability-Based Separation of Circulating Tumor Cells Using Microfluidic Ratchets,” *Small*, vol. 12, no. 14, pp. 1909–1919, Apr. 2016.
- [42] “Tools of Cell Biology - The Cell - NCBI Bookshelf.” [Online]. Available: <http://www.ncbi.nlm.nih.gov/books/NBK9941/>. [Accessed: 25-Apr-2016].
- [43] C. A. Combs, “Fluorescence Microscopy: A Concise Guide to Current Imaging Methods,” *Curr. Protoc. Neurosci. Editor. Board Jacqueline N Crawley Al*, vol. 0 2, p. Unit2.1, Jan. 2010.
- [44] “Fluorescence SpectraViewer.” [Online]. Available: <https://www.thermofisher.com/ca/en/home/life-science/cell-analysis/labeling-chemistry/fluorescence-spectraviewer.html>. [Accessed: 20-Apr-2016].

- [45] Y. Hiraoka, T. Shimi, and T. Haraguchi, “Multispectral Imaging Fluorescence Microscopy for Living Cells,” *Cell Struct. Funct.*, vol. 27, no. 5, pp. 367–374, 2002.
- [46] M. Brown and C. Wittwer, “Flow Cytometry: Principles and Clinical Applications in Hematology,” *Clin. Chem.*, vol. 46, no. 8, pp. 1221–1229, Aug. 2000.
- [47] D. A. Basiji, W. E. Ortyn, L. Liang, V. Venkatachalam, and P. Morrissey, “Cellular Image Analysis and Imaging by Flow Cytometry,” *Clin. Lab. Med.*, vol. 27, no. 3, p. 653–viii, Sep. 2007.
- [48] Y. Han and Y.-H. Lo, “Imaging Cells in Flow Cytometer Using Spatial-Temporal Transformation,” *Sci. Rep.*, vol. 5, p. 13267, Aug. 2015.
- [49] H. M. Davey and D. B. Kell, “Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses,” *Microbiol. Rev.*, vol. 60, no. 4, pp. 641–696, Dec. 1996.
- [50] H. Shapiro, “How Flow Cytometers Work — and Don’t Work,” in *In Living Color*, R. A. Diamond and S. D. M. B. (ASCP)CQ, Eds. Springer Berlin Heidelberg, 2000, pp. 39–56.
- [51] D. Marie, N. Simon, L. Guillou, F. Partensky, and D. Vaultot, “Flow Cytometry Analysis of Marine Picoplankton,” in *In Living Color*, R. A. Diamond and S. D. M. B. (ASCP)CQ, Eds. Springer Berlin Heidelberg, 2000, pp. 421–454.
- [52] A. L. Allan and M. Keeney, “Circulating tumor cell analysis: technical and statistical considerations for application to the clinic,” *J. Oncol.*, vol. 2010, p. 426218, 2010.
- [53] L. Bickford *et al.*, “Enhanced multi-spectral imaging of live breast cancer cells using immunotargeted gold nanoshells and two-photon excitation microscopy,” *Nanotechnology*, vol. 19, no. 31, p. 315102, 2008.
- [54] D. R. Shaffer *et al.*, “Circulating Tumor Cell Analysis in Patients with Progressive Castration-Resistant Prostate Cancer,” *Am. Assoc. Cancer Res.*, vol. 13, no. 7, pp. 2023–2029, Apr. 2007.
- [55] M. Ignatiadis *et al.*, “International study on inter-reader variability for circulating tumor cells in breast cancer,” *Breast Cancer Res.*, vol. 16, p. R43, 2014.
- [56] K. Polyak and R. A. Weinberg, “Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits,” *Nat. Rev. Cancer*, vol. 9, no. 4, pp. 265–273, Apr. 2009.
- [57] L. F. Ogle *et al.*, “Imagestream detection and characterisation of circulating tumour cells – A liquid biopsy for hepatocellular carcinoma?,” *J. Hepatol.*, vol. 65, no. 2, pp. 305–313, Aug. 2016.
- [58] O. Marques *et al.*, “Local iron homeostasis in the breast ductal carcinoma microenvironment,” *BMC Cancer*, vol. 16, p. 187, 2016.
- [59] N. López-Riquelme *et al.*, “Imaging cytometry for counting circulating tumor cells: comparative analysis of the CellSearch vs ImageStream systems,” *APMIS*, vol. 121, no. 12, pp. 1139–1143, Dec. 2013.

- [60] E. E. Reyes *et al.*, “Quantitative characterization of androgen receptor protein expression and cellular localization in circulating tumor cells from patients with metastatic castration-resistant prostate cancer,” *J. Transl. Med.*, vol. 12, no. 1, p. 1, 2014.
- [61] C. Allan *et al.*, “OMERO: flexible, model-driven data management for experimental biology,” *Nat. Methods*, vol. 9, no. 3, pp. 245–253, Mar. 2012.
- [62] A. R. Kherlopian *et al.*, “A review of imaging techniques for systems biology,” *BMC Syst. Biol.*, vol. 2, p. 74, 2008.
- [63] E. Arzuaga-Cruz *et al.*, “A MATLAB toolbox for hyperspectral image analysis,” in *Geoscience and Remote Sensing Symposium, 2004. IGARSS’04. Proceedings. 2004 IEEE International*, 2004, vol. 7, pp. 4839–4842.
- [64] R. A. Neher, M. Mitkovski, F. Kirchhoff, E. Neher, F. J. Theis, and A. Zeug, “Blind Source Separation Techniques for the Decomposition of Multiply Labeled Fluorescence Images,” *Biophys. J.*, vol. 96, no. 9, pp. 3791–3800, May 2009.
- [65] L. Biehl and D. Landgrebe, “MultiSpec—a tool for multispectral–hyperspectral image data analysis,” *Comput. Geosci.*, vol. 28, no. 10, pp. 1153–1159, Dec. 2002.
- [66] Jordan, Johannes and Angelopoulou, Elli, “Gerbil - A Novel Software Framework for Visualization and Analysis in the Multispectral Domain,” 2010.
- [67] N. Habili and J. Oorloff, “ScyllarusTM: From Research to Commercial Software,” 2015, pp. 119–122.
- [68] A. Orth, M. J. Tomaszewski, R. N. Ghosh, and E. Schonbrun, “Gigapixel multispectral microscopy,” *Optica*, vol. 2, no. 7, p. 654, Jul. 2015.

Appendices

Appendix A Evaluation of Alternative Spectral Analysis Software Packages

Most spectral analysis software are designed for spectral processing of satellite and aerial image data. A few specialized tools, for spectral analysis software, exist for biological unmixing applications. Another issue is that most spectral processing software demands control over the incoming sample. This is needed in order to produce useable results, which is a complication when working with samples with highly heterogeneous properties such as those from a terminal cancer patient. In this section six available tools for spectral processing are evaluated to test their suitability for the experimental process described previously.

The image data used is a 1.5Gigapixel 7680x7680x26 (X, Y, Lambda (colors)) image cube. The test criteria is whether the tools produce clear visual separated images with sufficient detail for a reviewer to evaluate the automatic results. Another issue with highly variable samples is that precise spectral calibration information is not possible on a per sample basis. This is due to the time overhead and possibility of not sufficient spectrally pure quality references within a random sample. Typical acquisition times for sample images are about 5-10 minutes each with hundreds of samples processed in the course of research. As a result processing has been defined to target a similar timescale of 5-10 minutes for processing, with an upper limit of 30 minutes. If the method exceeds the upper limit it is considered to be unsuitably slow. Prompt processing times are required because the image operator needs feedback to evaluate if their sample is ideal for time sensitive follow on processing such as: single cell extraction; sequencing; DNA storage; and many other processing techniques that demand fast turnaround times.

In addition to time constraints, running the software on local machines, which are ideally the workstations operating the microscope, is ideal for user workflow purposes. This is because operators can analyze capture data on the spot in order to decide what steps to take next. As a result the specifications for the target machine have been set to be a Quad Core Intel Haswell i7 with 32GB of memory. For testing purposes a more powerful machine was used with 64GB of memory and a Hexa Core Intel Haswell i7 processor. A tested software package is deemed to fail to process

if: it crashed or exceeded the 32GB of memory limit; took longer than 30 minutes to process a single image; or produced erroneous output images.

The following subsections describe various software packages tested against the actual image data captured by the microscope. These subsections also describe the software packages performance and if any failed to process the manner in which they failed. A review paper of medical imaging listed 16 common biological image analysis software packages with only one software package supporting unmixing via a plugin[62]. As a result the search scope was expanded as spectral unmixing appears to be far more commonly used in analysis of planetary surfaces in the geospatial studies field. Many packages are designed for these planetary surface applications without much support for biological imaging. Tests were carried out to validate software for use within the workflow. The general unsuitability of the tested software necessitated the development of custom algorithms.

A.1 Hyperspectral Image Analysis Toolbox

Hyperspectral Image Analysis Toolbox is a MATLAB based tool running its own custom user interface [63]. This software is designed to use a variety of methods including principle component analysis (PCA) to allow for detection of the predominant spectra even where exact calibration information is not available. The performance was not tested due to the ingest functionality of the package not working correctly. This software was also mainly designed and tested for reviewing ROIs from hyperspectral satellite imagery. Because the software could not open the gigapixel image cube no successful separation was completed.

Complications arose in attempting to load the extremely large image data into the plugin, resulting in the software exceeding system memory when it attempted to import the test image stack. Processing was considered a failure due to its inability to open the entire image stack. Furthermore, processing the file in chunks would result in lost cell information due to the exact cell positions being generally random.

A.2 PoissonNMF

PoissonNMF is an open-source spectral image processing plugin for ImageJ [64]. This software package uses automatic spectral detection to extract expected spectral components without the need for calibration data. However, high and low accuracy separation settings results in unacceptable runtime ranging from 20 minutes to over 60 minutes. Even with the lower accuracy automatic calibration, the separation performance was not satisfactory. There was poor unmixing in cells with mixed markers. However, the plugin was able to separate spectrally pure cells with only one marker per pixel. A common issue with samples is that most algorithms expect spectrally pure pixels to obtain the reference spectrum. However, biological samples markers such as CD45, EpCam, and CK can exist in the same pixel making obtaining a per sample reference spectrum automatically or manually challenging.

Sample images loaded into PoissonNMF were successfully loaded. However, the interface became sluggish once processing started and memory exceeded 32GB. The program was forced closed after 60 minutes of processing using default settings. Parameters were then adjusted to attempt to reduce the runtime to approximately 20 minutes. This resulted in poor spectral separation with only unmixing pure pixels.

A.3 Multispec

Multispec is a custom software package written primarily to process geospatial images with the ability to apply spectral unmixing functions to images [65]. The software supports automatic unmixing in the absence of a pre-calibration dataset. Although the software could load and view the image with no apparent slowdown, processing failed to start due to a built in memory estimator asking for 256GB of memory to complete the processing of the sample image file. Since this amount of memory is impractical for a workstation computer the software package exceeded target system requirements.

While pre-processing with Multispec used very little memory, Multispec does not appear to retain the image file in memory. Rather it appears to take a crudely subsampled thumbnail. Furthermore, processing times are extremely long with this program with frequent freezing. Initially setting it to lowest possible quality results in 30 minutes to complete the task. Furthermore there is no support

for 64bit operating systems or large address awareness, and as a result the program cannot effectively take advantage of systems with greater than 3GB of memory. All images were stored on PCIe SSD disks for maximum disk streaming speeds. Solid state disk speeds were in excess of 800MB/s to minimize the disk streaming effect. No separation was successfully completed using this software package.

A.4 Gerbil

Gerbil is a relatively new spectral unmixing software designed to provide an easy to use environment for processing spectral image results [66]. This software uses a number of modern techniques including neural network type processing. One caveat, however, is that the memory requirements are significantly higher than most software packages. During testing the Gerbil software suite exceeded system memory limits, freezing the software and operating system due to memory exhaustion. Due to these factors tuning the software was difficult as processing would destabilize the computer, requiring full restarts when attempting alternative settings. While sub-regions were successfully separated cleanly between positive and negative cells, no full region processing was successfully completed.

Loading a file into Gerbil and performing the separation was a simple task and the program worked properly after a short delay. However, a severe downside of Gerbil was that for the software to automatically crop images to a block of just 512x512 pixels requires 225 repetitions with an estimated sequential run time of 11.25 hours. To process the image in one shot the estimated memory required is 3490GB which is impractical for even most high performance servers. The result produced an image which appeared to separate possible cells for negative cells. Gerbil also provided additional feedback on each target. This was the only software tested that actually produced correct outputs for a number of samples tested.

A.5 Scyven

Scyven is a standalone spectral processing software package [67] that specializes in analysis of spectral images captured from real world scenes. The Scyven software suite has built in features

to compensate for spectral illumination content as well as reflection properties. Both of these functions appear to be integral to the program and cannot be disabled. Further still, these functions are not relevant to a flat confocal image with no background illuminant, as there is no reflection to compensate for in a fluorescence emission process. As with other software that exceeded system memory limits instabilities made tuning difficult. Scyven also supports manual, automatic, and assisted spectral identification. However, this feature was not successfully used on the gigapixel image cubes.

Moreover, Scyven exceeded the target runtime of 30 minutes and subsequently exhausted all system memory. The software could not process the files successfully and while it could open the gigapixel image the interface was sluggish to use after importing the file.

A.6 ZEN

Zeiss provides specialized software with their spectral microscopy platforms. The software suite includes a tool that allows for spectral unmixing using automatic or pre-defined calibration curves. Initial testing attempted to build a spectral library, but changes in sample conditions necessitated constant updates to a per sample calibration requirements. The constant updates were considered unfeasible due to the additional overhead in finding quality pure calibration debris to calibrate the unmixing algorithm. The ZEN microscope control software also includes a proprietary automatic unmixing algorithm called automatic component extraction (ACE) which uses less than 32GB of memory and typically completes analysis under 15 minutes. However the resulting images from a series of test samples were often either blank or did not correctly identify the desired spectral components. This algorithm is likely very similar to other PCA type automatic identification methods.

A.7 Summary of Software Tested

Of the six software packages none successfully processed the sample images in a reasonable timeframe. Furthermore, packages that may have processed sample images had requirements beyond realm of feasibility for use in a biological lab equipped with the latest workstations. The

following table 3.2 summarizes the sample testing performed on the various software packages and includes the results of this thesis's lightweight processing algorithm. The table demonstrates the comparative advantages of the lightweight processing algorithm which achieves a low memory footprint with fast processing times. Additional investigation into hardware accelerated unmixing systems such as dedicated DSP/FPGA/distributed GPU processing was considered to be outside the scope of this thesis. While these hardware accelerated unmixing systems offer significantly faster processing times, they suffer from a lack of available memory to store the input gigapixel image data.

Other software packages developed specifically for gigapixel multispectral images were found to use guided linear unmixing within a MATLAB based process. However the code is not public and the test data was performed using idealized plastic calibration beads instead of actual patient samples that exhibit greatly increased variations in signal response [68]. As a result, this software could not be tested and is not included in the table below. Given that the software uses linear unmixing with assisted calibration, it is a reasonable inference that this software would fail to process more realistic samples with non-linear and challenging imaging conditions (where per sample calibration is impractical).

Multispectral Software	Memory Requirement	Runtime	Result
Hyperspectral Image Analysis Toolbox	>32GB	NA	DNR
PoissonNMF	>32GB	>60	Incorrect Separation
Multispec	~256GB	NA	DNR
Gerbil	~3490GB	~675	Correct Separation
Scyven	>32GB	>30	DNR
ZEN	<32GB	~15	Incorrect Separation
ctcFastSelector (matlab) Thesis	<32GB	<5	Correct Separation

Table A.1 Summary of the tested results. See the software comparison for additional details on why a software failed or did not run

The test platform used a more powerful system to provide the best case scenario for the tested software. While actual use of the system was done under more reasonable specifications for practical purposes. The ability to run the software during image acquisition enables a number of highly desirable options such as: re-imaging suspect samples; reducing time delays between

processes; and eliminating post processing review (as analysis review could occur during sample acquisition for fast sample runs).