Automatic Detection and Classification of Tumor Infiltrating Lymphocytes

A thesis submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

by

Andrew Janowczyk
(Roll No. 08405201)

Under the guidance of
Professor Sharat Chandran

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY–BOMBAY
2013

Thesis Approval

The thesis entitled

Automatic Detection and Classification of Tumor Infiltrating Lymphocytes

by

Andrew Janowczyk
(Roll No. 08405201)

is approved for the degree of

Doctor of Philosophy

_____	_____
Examiner	Examiner
_____	_____
Guide	Chairman

Date: _____

Place: _____

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

(Signature)

_____

(Name of the student)

_____

(Roll no)

Date: _____

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY, INDIA

CERTIFICATE OF COURSE WORK

This is to certify that **Andrew Janowczyk** (Roll No. 08405201) was admitted to the candidacy of Ph.D. degree on April 2012, after successfully completing all the courses required for the Ph.D. programme. The details of the course work done are given below.

S. No	Course Code	Course Name	Credits
1	CSS801	Ph.D. Seminar	4
2	CS 717	Statistical Relational Learning	6
3	CS 722	Advanced Computer Graphics	6
		Total Credits	**16**

IIT Bombay

Date: Dy. Registrar (Academic)

Abstract

The latest global statistics from 2008 show that 12.7 million new cancer cases and 7.6 million cancer deaths occurred worldwide[1]. This accounts for 13% of all deaths for that year, making cancer a common threat to all families. As technology becomes more efficient, a trend towards computer aided diagnostic (CAD) tools for identification, prognosis prediction and reoccurrence likelihood is becoming a reality. A cornerstone of CAD systems is the field of digital histopathology, where analyzing large quantities of cellular images can leverage the research from the mature field of computer vision.

In this thesis we present two novel general theories, one focusing on homogeneity and one which quantifies heterogeneity. From these theoretical roots, we explain our two algorithms, Hierarchical Normalized Cuts (HNCuts) and Local Morphologic Scale (LMS), which when used together, create a system which can automatically segment (via homogeneity) and classify (via heterogeneity) lymphocytes from digital histology images as either tumoral or stromal. There is strong evidence that this laborious task is a valuable prognostic indicator, motivating the development of algorithms for its automation.

The first algorithm, HNCut, is responsible for segmenting the stained lymphocytes from the digital image. By using a hierarchical data structure with a novel frequency weighted mean shift for clustering, our approach is able to perform segmentation with very high throughput in a reproducible robust manner. We take our theory which contributes a general approach towards segmenting colors of interest defined by a user specified swatch and apply it to the aforementioned histology task, which previously took a pathologist a day to complete. Using our algorithm it can be completed in a matter of seconds.

The second algorithm, LMS, classifies the lymphocytes identified by HNCut as either tumoral or stromal based on their local morphology. While available theories tend to focus on quantifying homogeneity, our theory contributes a general approach towards quantifying

[1] World Heath Organization, GLOBOCAN 2008

heterogeneity. This departure from the typical paradigm affords us the opportunity to more aptly undertake typical biomedical problems such as classification. Since lymphocytes appear the same regardless of their class (with surrounding cells providing the classification), they represent a good opportunity to demonstrate our approach. By quantifying the heterogeneity of the local neighborhood into a feature set, we create a classifier which competes with the current state of the art. The notable benefits are a low dimensional representation, and the ability to easily parallelize the approach to take advantage of multiple cores.

Using real-world datasets we compare our approaches to industry standard approaches. In culmination, we show how when combined, the two algorithms create a robust high throughput system for the automatic detection and classification of lymphocytes, a notable contribution to the CAD paradigm.

Contents

List of Tables

List of Figures

Nomenclature

$\mathbb{R}$	Set of real numbers
$\mathbb{N}$	Set of natural numbers
G	The Gaussian function
$\mathcal{N}_\kappa$	κ-neighborhood around a pixel ($\kappa \in \{4, 8\}$)
$< \ldots >$	Ordered sequence of values
$\mathbf{X}$	Matrix
$\mathbf{X^i}$	i-th column of Matrix $\mathbf{X}$
$\mathbf{X_i}$	i-th row of Matrix $\mathbf{X}$
$\mathbf{x}$	Vector
$\|\mathbf{x}\|$	Cardinality operator of a vector **x**
$\|\|\mathbf{x}\|\|_p$	p-norm operator of a vector **x**
$\mathbf{x}^T$	Transpose operator of a vector **x**
x_i	*i*-th element of vector $\mathbf{x}$
x	Scalar value

Chapter 1

Introduction

The practice of medicine extends from the very early stages of the human society. It is known that the Egyptians and Babylonians both introduced the concepts of diagnosis, prognosis, and medical examination over 5,000 years ago. The Egyptians are credited with the oldest description of cancer (although under a different name) [1], in an Egyptian textbook which dates back to around 3000 B.C. It specifies that there is no treatment. The famous Indian physician Sushruta, in his 600 B.C text Sushruta Samhita, identifies cancer as inflammatory or non-inflammatory swelling and classifies tumors as either as 'Granthi' (minor neoplasm) or 'Arbuda' (major neoplasm) [1, 2]. His Ayurveda based treatments used various herbs. Hippocrates is credited with the origin of the word "cancer", and documented several kinds around 400 B.C.; his treatments were blood letting and obscure diets.

Yet, in 2008, 12.7 million new cancer cases and 7.6 million cancer deaths occurred worldwide[2]. This accounts for 13% of all deaths for that year, making it the leading cause of death world wide according to the World Health Organization. It is curious to see a problem so old still an ever present threat even in our current day.

One of the most notable attributes of the human race is the application of tools and technology to resolve problems. The question then becomes, how can we use advances of medical, mathematical and computational resources, defined as both knowledge and efficiency, to further advance our cause of a disease-free healthy life.

One field which has grown along these lines especially in recent years is that of computer aided diagnostics (CAD). While the field encompasses many disciplines, techniques and

[1] http://www.cancer.org/cancer/cancerbasics/thehistoryofcancer/the-history-of-cancer-what-is-cancer

[2] World Heath Organization, GLOBOCAN 2008

modalities, its sole purpose is to assist medical professionals in their decision making process by providing accurate information. This information can come from statistical analysis of past patients, or individual data points such as identifying anomalies in biopsies, MRI, or X-rays.

In this thesis we focus specifically on the field of histopathology, or the study and diagnosis of disease using cellular and tissue information, for CAD. A typical example is obtaining a biopsy from a suspected tumor and examining the sample under the microscope to both confirm a cancer diagnosis and stage the cancer. The exciting part of this approach is that with the advent of microscope slide scanners, we can create digital images of exactly what the pathologist is seeing under the microscope. Once in digital form, not only can the images be easily shared across institutions, but the application of previous research from the rich field of computer vision becomes possible.

There are a multitude of reasons driving these types of research. Firstly, one of the scarcest resources in medicine is the time of trained professionals. Simply by automating time consuming tasks, resources can be refocused on problems which are known to be difficult or impossible for computers to solve, raising the overall level of care. Secondly, by exactly quantifying attributes of diseases in a reproducible manner, we can eliminate observer variance and raise the level of precision of diagnostic tools. Additionally, with these algorithmic standards in place, cross institutional collaboration and knowledge sharing becomes easier as there is already a consensus in place on the values of properties being discussed. Lastly, and perhaps the most interesting for scientists, is the possibility of mining large amounts of both old and new data to identify patterns and prognostic indicators. For example, there are in existence large tissue repositories around the world which could contain valuable insights into the behavior, treatment and outcome of various diseases, yet the process of analyzing this data unaided is intractable. By developing high throughput algorithms, which can accurately quantify and inspect various correlations, we potentially have the ability to unearth powerful approaches to these age old diseases.

In the future, the gold standard of medical care will be personalized medicine. Personalized medicine is the idea that medical decisions and practices will be custom tailored to each individual patient's exact medical situation, resulting in the most direct treatment with the least amount of collateral damage. The key to reaching this goal is, simply stated, the creation, management and analyses of vast amounts of previous information. This is exactly where algorithms, and in a small part this thesis, fit into the future.

1.1 Problem Definition

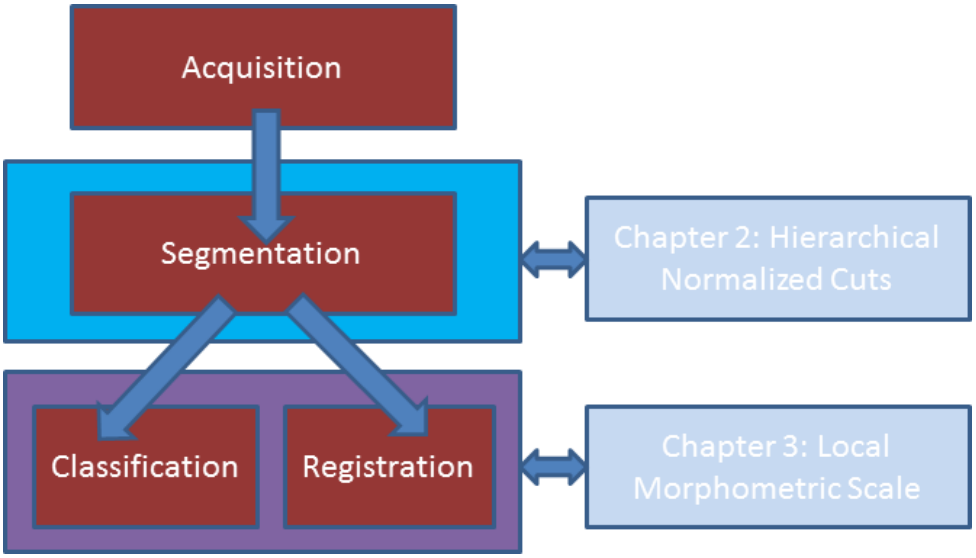


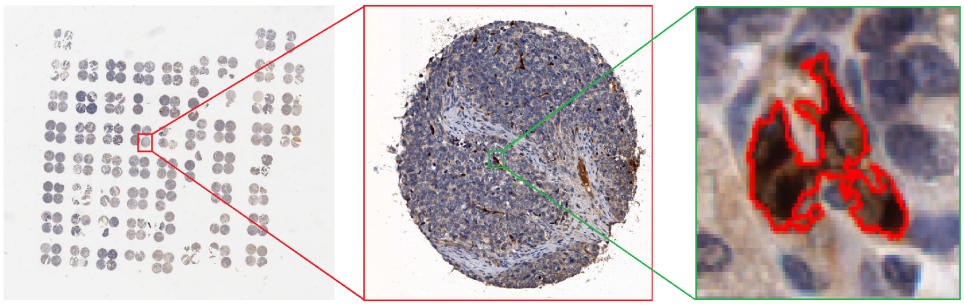
Figure 1.1: Three major tasks associated with computer aided diagnostic include acquisition of data, segmentation of regions of interest (ROI), and lastly the classification and registration of these ROI. The thesis makes notable contributions in segmentation (third chapter) and classification (fifth chapter), with a potential extension to registration .

Figure 1.1 gives a high level overview of the broad research fields used in CAD, and where this thesis fits relative to them. Naturally, all information must first be acquired from its respective data source. With public and private histology banks in existence, storing years of samples across thousands of patients, a vast amount of data is already in existence. Once obtained, the image data is pre-processed to extract biological information via segmentation; thus determining regions of interest (ROI). This information can include the size, location and chromatic properties of various cellular entities or organs. Once the region is identified, one of two courses of action is typically taken. The first possible course, classification, is illustrated by the case of cancer detection where the region is classified as either cancerous or non-cancerous. The second possible course, registration, is rooted in the desire to align the current region to an existing model, perhaps to identify anomalies in shape or location of critical organs.

This thesis develops techniques applicable to a vast number of areas in the broader computer vision domain including segmentation, classification, and registration via two novel state of the art methods. Here we focus on the important problem domain of histopathology and more specifically on determining the class of a lymphocyte as either tumor infiltrating (TIL) or non-TIL. We illustrate how our rapid automated detection system progresses by performing

segmentation which feeds a morphologically aware classification system. Each of the individual sub-problems and challenges are explained below.

1.1.1 Segmentation Using Hierarchical Normalized Cuts



(a) (b) (c)

Figure 1.2: (a) A tissue micro array and (b) a representative magnified tissue cylinder culled out from (a) with the extracted biomarker presented in (c) delineated in red. A typical microarray could contain over 500 individual cylinders, making the biomarker detection via traditional image analysis algorithms a challenge.

One of the common tasks in digital pathology is quantification of properties associated with extent of stain as a result of staining for identification of biomarkers. For example in the domain of Ovarian cancer (OCa), recent works [3, 4, 5, 6, 7] suggest that specific tumor vascular biomarkers (TVMs) may be identifiable on OCa tissue microarrays (TMA) that could have prognostic significance, helping to not only predict the survival rate but also help determine a more specific course of treatment. It has also been suggested that genes expressed uniquely by the vasculature of tumors may provide important therapeutic targets. Biomarkers are typically discovered by staining explicitly for TVMs of interest on OCa TMAs, essentially requiring a vast study for each biomarker of interest. Precise quantification of the extent and intensity of the stain could serve as a prognostic metric reflecting risk of disease recurrence and patient survival. However, due to the data size involved in each of the studies it is currently infeasible in terms of both time and effort for an expert pathologist to perform this segmentation manually.

Our specific problem definition in this sub-domain is then identified as taking a histology image and extracting the stained regions as demonstrated in Figure 1.2. These types of images are especially challenging because the visual appearance of specimens are inconsistent as they are affected by temperature, time, concentration of stains, scanning equipment and other en-

vironmental variables. The confidence necessary to provide an industry standard approach is non-trivial as these variances, along with a wide range of user defined input parameters, create a complex problem domain.

Challenges

While examining histology images are notoriously challenging, a few reasons which make our particular problem unique are:

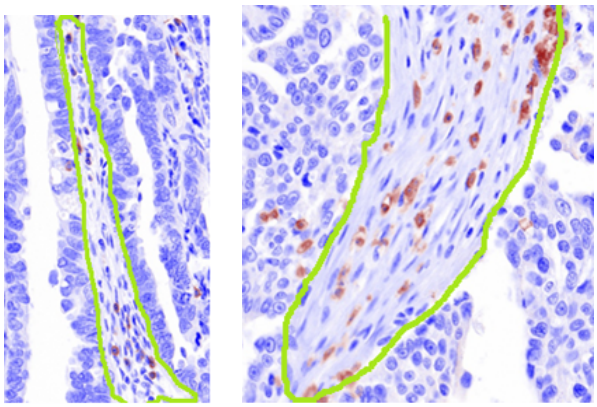
- Datasets are very large and thus require a highly efficient algorithm to make computation tractable. Additionally, a large number of these datasets already exist in tissue repositories waiting to be mined.

- Images are not consistent across the dataset due to lighting, staining, and human preparation variations. This anomaly becomes more significant as various institutions create samples at different times, essentially ensuring a large variance in visual appearance.

- Annotation of training data is laborious and time consuming, and thus limited supervised data is available. Additionally, each new stain would require an equal investment to re-annotate and thus re-train.

- Precise reproducibility based on a wide range of input parameters is necessary for confidence and data exchange between operators. For an algorithm to become useful, institutions need to witness that the output created is less variant than intra-expert variability.

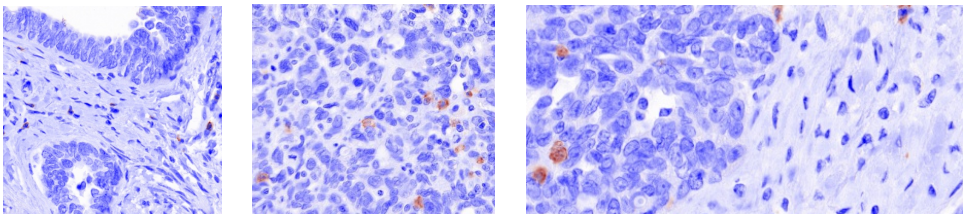
1.1.2 Local Region Classification



(a) (b)

Figure 1.3: Stroma region manually circled in green. Although they are both stromatic regions, notice the stark difference in size and shape between the two regions in (a) and (b). Selection of an appropriate window size or shape in a typical approach such as texture features is difficult.

In the domain of classification, we aim to provide a signature at the pixel level which can be used to successfully differentiate tumor regions from stromal regions. As shown in Figure 1.3, the chaotic nature of region size and shape prevents the selection of optimal operating parameters for standard industry algorithms, such as window size for texture features. The specific challenging task we have chosen is the classification of a region as either tumoral or stromal. One application of the resulting identification is to separate tumor infiltrating lymphocytes (TILs) from non-TILs, as shown in Figure 1.4. A lymphocyte is a type of white blood cell that is sent to the proximity of objects which the body considers foreign. Recent work [8, 9, 10, 11, 12, 13, 14, 15, 16, 17] has suggested that a valuable prognostic indicator is based on the extent to which the patient's own immune response, namely their lymphatic response, has attacked the cancer. While TILs and non-TILs are visually identical, their sole differentiating factor is their location in or around a tumor. This motivates the necessity for a solution to the broader problem of classifying a region as tumor or stroma. We note, however, that the technique developed here is not specific to only tumor and stroma classification.



(a) (b) (c)

Figure 1.4: Lymphocytes in all panels are stained in red. All of the lymphocytes in (a) the stroma region image are non-TILs, while all of the lymphocytes in (b) the tumor region image are TILs. In (c) we see both TILs (left half) and non-TILs (right half).

Challenges

Due to the chaotic nature of cancer cell growth, the associated cellular structure is unorganized and unpredictable making analysis and generalization difficult. In addition there are the following specific challenges:

- Domain specific approaches require information about individual cells, such as size and dispersion pattern relative to its peers. The segmentation of individual cells is difficult due to clumping as a result of a three dimensional tissue sample being scanned in two dimensions, and thus computationally expensive methods are needed for cell separation.
- Selection of an appropriate window size for standard approaches. We can see from Figure 1.3 that the pre-selection of a texture window would be challenging because of the varying sizes and shapes of such a window (a normal rectangle would not work).
- The stroma region is often nestled between areas of tumor, making not only its boundaries not clearly defined but also the size of the associated region difficult to pre-determine.

1.2 Contributions Made in This Thesis

The problem domain we have chosen is the identification of lymphocytes as either tumor infiltrating (TIL) or non-TIL. Our approach is to first use our Hierarchical Normalized Cuts (HNCut) to segment necessary information for our Local Morphologic Scale (LMS) algorithm to successfully classify pixels of interest as either tumor or stromal regions. This thesis contributes the necessary theory and implementation for the chain of these two novel algorithms.

- A high-throughput hierarchical segmentation scheme which not only operates on large (1.5 million pixels) images in under 10 seconds, but also is easily scalable to entire TMAs.
- Parameter insensitive segmentation for large images and the ability to discriminate between regions with similar color values.
- Layman initialization of the system is possible, obviating the need for detailed ground truth annotation from an expert that is required for more sophisticated supervised classifiers.
- A novel signature definition allowing for quantitatively characterizing local heterogeneity. This is especially relevant in the context of histopathology which consists of notoriously heterogeneous images.
- Our rotationally invariant, non-domain specific, quantitative signature at the pixel level which can be used for region classification, segmentation, and registration.
- This signature is accurate across a range of window sizes, overcoming common downfalls of texture and template matching based classifiers.
- While our approach is already high throughput, it is additionally well suited for GPU computing, motivating unbounded data-analysis.
- Our algorithms work together to form a novel system for the very important problem of separating out tumoral from stromal regions via application of LMS.

1.3 Organization of Thesis

The structure for the thesis is as follows:

Chapter 1 In this chapter we have introduced Computer Aided Diagnostics and discussed the problem scope and challenges. We developed in abstract terms an overview of the material in this thesis, breaking it down into two parts, segmentation and classification. We explained on a high level how these two algorithms fit together.

Chapter 2 To clearly illustrate how our theories and algorithms differentiate themselves from existing work, in this chapter we provide a targeted literature survey. By transparently exam-

ining the weak and strong points of the various but similar approaches, our contribution becomes clearly identifiable.

Chapter 3 We develop the theory for the HNCut algorithm which combines a hierarchical data structure with normalized cuts to extract colors of interest from the color space. In this chapter we also give explicit algorithms for the implementation of HNcut using a frequency weighted mean shift for a complete high-throughput approach.

Chapter 4 To validate the theory presented in the previous chapter on HNCut, we conducted numerous experiments which validate not only the speed and efficiency but also the reproducibility and robustness of our approach. Further we compare our approach to both an unsupervised (k-means) and supervised learning algorithms (Probabilistic Boosting Trees).

Chapter 5 In this chapter we explain the Local Morphologic Scale, an approach which quantizes local morphology as a feature descriptor which could be used in registration, segmentation, classification or retrieval. This explanation includes the necessary theoretical background and algorithms and is further elucidated by a discussion on the various properties it possesses.

Chapter 6 The properties discussed in the previous chapter are fully vetted by using a synthetic-data set across numerous experiments. These experiments quantitatively prove the properties which were proposed in theory section. Additionally, an experiment comparing the efficiency parameter versus computation time is presented.

Chapter 7 We introduce the combination of these two algorithms to the real world application of TIL identification. This chapter explains the training and testing methodology used across 5 different data domains. A thorough discussion and set of experiments which demonstrate the impact of the various parameters is also provided.

Chapter 8 The final chapter contains a summary of the previous chapters and how this work is seated in the histopathology domain. We identify the contributions made in this thesis as well as discuss the possible shortcomings of our techniques. We conclude by indicating ways to extend some of the proposed ideas in future works.

Chapter 2

Literature Survey

As discussed in the previous chapter, the desire to use computer aided tools for diagnostic purposes is not unfounded. In this chapter we develop three sections. The first two sections are for the core aspects of this thesis, namely the two algorithms (HNCut and LMS) mentioned in the previous chapter. *We compare and contrast them to previous works and identify the building blocks used for the theory in the upcoming chapters*. The third section is devoted to previous approaches for tumor identification. This allows for the study of previous shortcomings and thus motivate the need for novel solutions.

2.1 Relevant Work in Segmentation

Most previous computerized image analysis algorithms for TMAs have involved thresholding based schemes [18], [19], [20]. These methods are known to be highly sensitive to even slight changes in color and illumination. Clustering based approaches, including k-means [18], have also been investigated for the analysis of TMAs. However, k-means is a non-deterministic algorithm and is highly sensitive to the initial choice of cluster centers [21]. Active contour schemes [22], while suitable for cell and nuclear segmentation in digital pathology, are not ideally suited to the problem of pixel level classification. Additionally they are typically infeasible for problems where hundreds of objects need to be concurrently segmented on very large images [23].

While supervised learning methods such as Probabilistic Boosting Trees (PBT) [24, 25] have become popular for image classification and segmentation, these methods are constrained by the difficulty [26] in obtaining ground truth segmentations from experts for classifier training of the object of interest. Manual annotation of the data, apart from being time-consuming and

laborious, can also be expensive if only a medical practitioner is capable of providing accurate annotations. Additionally, if the target of interest changes, considerable effort might be required to generate new annotations and re-train the classifier.

Normalized Cuts (NCut) [27] is among the final mature descendants from a series of graph cutting techniques ranging from max cut to min cut [28, 29, 30, 31]. It is a popular scheme in spite of its main drawbacks: (1) the large number of calculations needed for determining the affinity matrix and (2) the time consuming eigenvalue computation. For large images the computation and overhead of these border on the infeasible [27]. Consequently, a significant amount of research has focused on avoiding their direct calculations [32, 33].

The mean shift algorithm (MS) [34] has been employed and modified in [35] as an unsupervised technique for mode discovery instead of k-means. MS attempts to identify the cluster mean within a pre-defined bandwidth. By using a steepest gradient approach, a fast convergence to the set of true means of the statistical data can be found [36]. The improved fast Gauss transform (IFGT) implementation of the MS algorithm [37] allowed computation times for large images to become reasonable. For the rest of this thesis, we will make no distinction between IFGT-MS and MS.

The attempt to merge NCuts and mean shift is not new [38]. To overcome the computational issues associated with NCut, a novel approach of combining both the MS and NCut algorithms was presented in [38]. Clustering the image by running the MS algorithm to convergence produced class assignments for the pixels. By taking the average intensity value of the regions obtained via the MS clustering step and using them as the vertices in the NCut algorithm, a significant speed improvement was obtained.

It was later noticed in [39] that when points of similar values are within an ϵ neighborhood of each other, their contribution to the overall system can be merged, providing an efficiency improvement by reducing the number of computations needed per iteration. *We use this to extend the MS work of [38] in a hierarchical fashion which is more pertinent and amenable to problems in digital pathology and biomedical imaging*. This allows us to perform the same detection or segmentation task in less than one third the time (under .5 seconds for HNCut as compared to the reported 1.78 seconds for [39]).

While there are similarities between our approach and [38, 39], there are also significant differences. The proposed algorithm is specifically designed for rapid extraction of pixels of interest in a minimally supervised manner, as opposed to unsupervised clustering which is in-

sensitive to the user's domain knowledge as the aforementioned approaches take. Thus, we first manually identify the desired target class based on individual representative colors (referred to as a swatch) selected from the target class by a user. This swatch, which can be changed based on the desired target class or domain, lends HNCut significant flexibility and ease of use.

2.2 Relevant Work in Localized Scale

The notion of scale in the context of image processing has been routinely employed over the last few decades to facilitate multi-resolution feature analysis; the assumption being that certain pertinent image features are only discernible at certain image scales and hence a spectrum of image resolutions needs to be considered for object recognition. Multi-scale approaches (scale-space [40] and hierarchical pyramids [41]) envisioned image processing operations being applied on a single image at varying levels of resolution; homogeneous regions being operated on at a lower resolution, with more heterogeneous regions being examined at higher resolutions. A limitation of these multi-scale techniques is that an "optimal" image resolution needs to be selected from within the image pyramid [41]. Additionally, some approaches [42] might require selection of multiple image scales for classification of a single image region.

To overcome these difficulties, the idea of locally adaptive scale emerged [43]. The concept of local scale was introduced to characterize varying levels of image detail so that localized image processing tasks could be performed, yielding an optimal result globally. Pizer et al. [44] suggested that having a locally adaptive definition of scale was necessary even for moderately complex detailed images. By quantifying these image details, an adaptive local scale image could encode implicit information present in the image intensity values. Locally adaptive scale has seen application in a variety of image processing tasks including MRI bias field correction[45], image segmentation[46], image registration [47], and image coding [48].

With that said, in the context of this thesis, we break away from that definition of scale and extend it in a different context. The idea of scale can also be defined as "a graduated range of values forming a standard system for measuring or grading something" [1]. Examples of common *scale spaces* include temperature (heat), barometric (pressure) and altitude (elevation). This in itself is not novel in the imaging domain. Saha and Udupa introduced the notion of ball-scale [49] which at every spatial location was defined as the value corresponding to the radius of the

[1] Oxford Dictionary, Oxford University Press, 2013

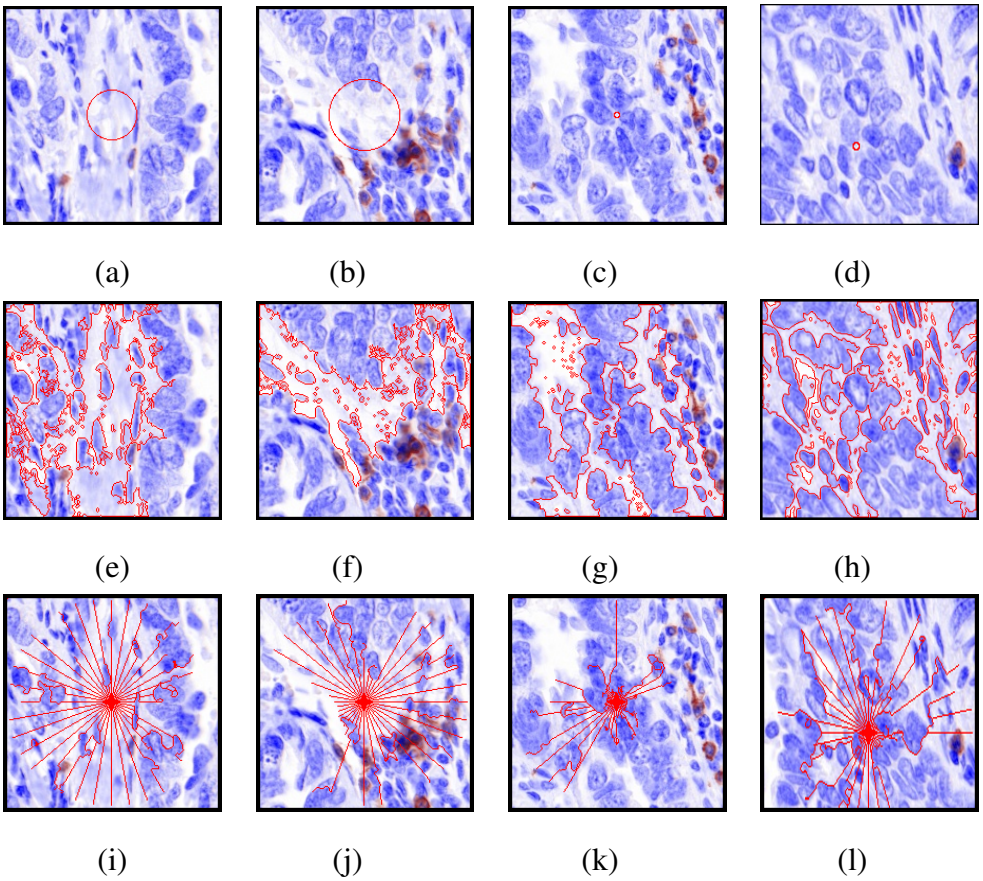
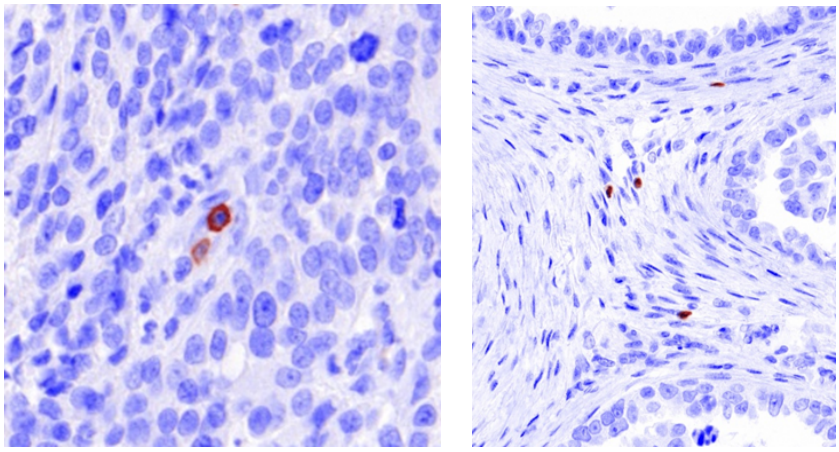


Figure 2.1: The associated b-scale ((a)-(d)), g-scale ((e)-(h)), and (proposed in this thesis) LMS signatures ((i)-(l)) shown for a candidate image location on an OCa biopsy image in red.

largest ball encompassing all locations neighboring the location under consideration and satisfying some pre-defined homogeneity criterion. In [50], Saha extended the ball-scale idea to a tensor-scale (*t*-scale), where the *t*-scale was defined as the largest ellipse at every spatial location where the pixels within the ellipse satisfied some pre-defined homogeneity criterion. The shape constraints of both (*b*-scale) and (*t*-scale) were overcome by Madabushi and Udupa with the introduction of generalized scale (*g*-scale)[46]. *g*-scale was defined as the largest connected set associated with every spatial location, such that all spatial locations in this set satisfied a pre-defined homogeneity criterion. Finally, with these scale space definitions in place, we refer back to the value of the multi-scale idea and note that similar pixels in these scale spaces can still be treated similarly, motivating the creation of scale spaces which are relevant to individual features so that groupings can be leveraged and exploited for efficiency.

The common thread among these space scale concepts was that they were defined based on some homogeneity criterion linking the pixels neighboring the spatial location under con-



(a) (b)

Figure 2.2: An example of a (a) TIL and (b) a non-TIL. We pose the hypothesis that perhaps homogeneity is not as interesting in histopathology as the ability to quantify heterogeneity. In the two images, we can see that the local regions surrounding the lymphocytes (stained in red) are both indeed heterogeneous, all be it in a different manner. A quantification of that difference, which is what our LMS aims to provide, is intended to lead to a valuable classifier.

sideration. Figure 2.1 reveals that both the b-scale ((a)-(d)) and g-scale ((e)-(h)) representations for a specific spatial location (located in the center of the image) attempts to identify the largest ball and set of pixels, respectively, that is homogeneous with respect to the pixel under consideration. For the stromal regions the associated b-scale ((a), (b)) and g-scale ((e), (f)) regions are large, reflecting the relative image homogeneity in that location of the image. Note however that the corresponding g-scale set is affected by the presence of local heterogeneity (g-scale has multiple cavities). The LMS ((i), (j)) for stromal regions has particles radiating far out in certain directions, but is also locally constrained (to the right) on account of neighboring nuclei. For the tumor regions ((c), (d)), the corresponding b-scale is small, with g-scale resulting in an amorphous shape with multiple cavities. Additionally the g-scale sets for the tumor regions in ((g), (h)) appears dramatically different. The corresponding LMS ((k), (l)) while not constrained by a prior shape model, yields a local structural signature that is consistent across both ((k), (l)) and distinctly different from the corresponding non-tumor LMS signatures ((i), (j)). Note that the initial motivation of both b-scale, and g-scale was from the perspective of noise filtering and bias field correction [45], image processing operations that warranted identification of locally connected homogeneous regions.

Using an example of a TIL and non-TIL in Figure 2.2, *we pose the question: is homogeneity interesting in this domain, or is the ability to quantify heterogeneity of greater value*? Consequently the scale definitions are not necessarily optimized to capture local heterogeneity, except as a very small ball (in the case of *b*-scale) or set (in the case of *g*-scale) of homogeneous pixels in image regions with significant complexity.

It is important here to note that the *b*-, *t*-, and *g*-scale formulations were not devised with the purpose of object classification in mind. While both the *b*- and *t*- scale definitions assign a feature vector (ball radius and parameters of ellipse respectively) to each image location, which could then potentially be used to perform pixel-level classification, it is not clear whether these *b*-, *t*- scale related parameters are discriminatory enough. For instance, for an ovarian cancer (OCa) biopsy image (Figure 2.1), the *b*-scale at two different locations in the image (Figures 2.1 (c), (d)), are identical in spite of significantly different local structural attributes. Similarly the *g*-scale representations at 4 different image locations from within the OCa biopsy image does not appear to yield a signature that is amenable to object or pixel level classification.

2.3 Relevant Work in Tumor Identification

While the historical works in the field of tumor versus stroma identification are varied, they tend to fall under three categories. As a result, we focus on one representative work from each category to provide the reader with a sufficient breadth of knowledge.

2.3.1 Specialized Staining

One of the common approaches to tackling this challenge revolves around using specialized staining. In some cases, it is possible to stain directly for a specific tumor type of interest allowing for a clear separation of regions. Since this is often not the case, the authors in [51] present an approach which requires specially stained fluorescence images from which they extracted the DAPI (49,6-diamidino-2-phenylindole) channel. They formed cell graphs based on the topological distribution of the tissue cell nuclei and extracted the corresponding graph features. By using topological, morphological and intensity based features they built a supervised classier using support vector machines which obtains an accuracy of $88\% \pm 6.68$. *We can contrast this with our approach*, obtaining on par results with a much smaller variance $86\% \pm .000354$ while operating on solely industry standard Hematoxylin (H) or Hematoxylin & Eosin (H&E) stained

images, allowing broader usage in pre-existing tissue repositories.

2.3.2 Computationally Expensive

There are notable approaches to the domain which given our current technological knowledge and infrastructure are intractable for immediate application in a clinical setting. For example, an N-point correlation function [52] (N-pcfs) for constructing an appropriate feature space for achieving tissue segmentation in histology-stained microscopic images was presented. The N-pcfs estimates microstructural constituent packing densities and their spatial distribution in a tissue sample. Afterwards, they represented the multi-phase properties estimated by the N-pcfs in a tensor structure. Using a variant of a higher-order singular value decomposition (HOSVD) algorithm, they realize a classifier that provides a multi-linear description of the tensor feature space. While the approach wasn't used directly for tumor versus stroma identification, they showed $> 90\%$ accuracy of their segmentations in a case-study that focuses on understanding the genetic phenotyping differences in mouse placentae. Unfortunately, the authors note in their discussion section the need to invest additional research in finding more optimal data structures and algorithms to reduce the overall time associated with computations. *We show that our feature set can be generated in as little as .0058s per sample, motivating the immediate usage in a high-throughput system.*

2.3.3 Full-Featured

C-path, as described in [53], first performed an automated, hierarchical scene segmentation that generated thousands of measurements, including both standard morphological descriptors of image objects and higher-level contextual, relational, and global image features. Using the concept of superpixels, they measured the intensity, texture, size, and shape of the superpixel and its neighbors. Afterwards, to produce more biologically meaningful features, they classified superpixels as epithelium or stroma. Using these classified superpixels they created more than 6600 features. Their approach found a set that were associated with samples from patients who had a shorter survival period. The key aspect of this analysis was that these features were not predefined by a pathologist as being relevant to cancer; instead, the software itself found the cancer-related features among the very large set of measurements of the image and obtained an accuracy of 89% in the superpixel classification. Their novelty was defined by their successful

combination of existing features. Our work presents a significantly lower dimensional novel morphological feature set, which obtains 86% accuracy on the same task, *clearly indicating a competitive approach*. Additionally, since we have the ability to perform on par using only a single feature, we believe that our approach is far more scalable and thus applicable in a clinical setting.

2.3.4 Summary

We can see from the three categories presented above that while this specific problem domain has previously been considered, no proposed solution which covers all the required facets currently exists. The core components of speed, scalability and cost-effectiveness are needed in order to have a clinically viable solution. The LMS framework presented in later chapters was specifically designed to meet these needs and thus advance the technological influence of computer aided diagnostics.

Chapter 3

Hierarchical Normalized Cuts (HNCut): Theory

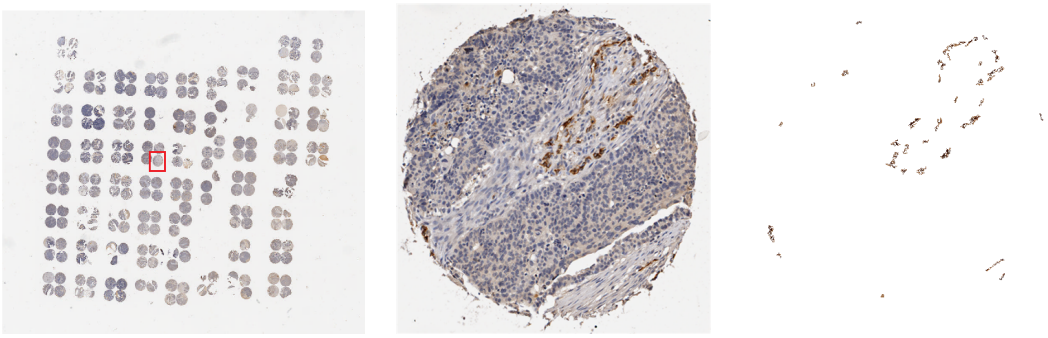
3.1 Introduction to Stain Quantification

Cells are inherently hard to see under a microscope due to their highly transparent nature and thus require staining in order to increase their visibility. Additionally, many chemical properties are not visible at all unless they are specifically targeted via chromatic staining. During this chemical process a visible chromatic dye binds to the desired molecule, leaving a visual marker of the molecule's presence. A valuable property of staining, as indicated by the Beer-Lambert law [54], is that the staining will visually manifest itself in locations where there is a higher concentration of the tagged molecule. As a result, by examining the intensity and extent of the stain, it becomes possible to have a quantitative measure of an otherwise immeasurable chemical property.

One main clinical result of this chemical reaction is that it becomes possible to stain for biomakers. Biomarkers are unique chemical signatures which are indicative of an important underlying chemical process. Pertinent to this chapter on Ovarian cancer (OCa), there are specific tumor vascular biomarkers (TVMs) [3, 4, 5, 6, 7] which are suggested as having prognostic significance, helping to not only predict the survival rate, and disease risk factors, but also help determine a more specific course of treatment.

Unfortunately, biomarkers are typically discovered by staining explicitly for individual molecules, essentially requiring a vast study for each biomarker of interest. This creates an undue burden on pathologists to manually annotate or review each stained biopsy to quantify

the stain presence. While tissue mircoarrays (TMAs) help to alleviate some of the burden in the preparation and reviewing of the specimens, the actual quantification process is still unnecessarily time consuming, worsened by the size of the image data being too large for typical automated approaches.



(a) (b) (c)

Figure 3.1: (a) A TMA and (b) a representative magnified tissue cylinder drawn from (a) with the extracted stained TVM presented in (c). A typical TMA could contain over 500 individual cylinders, making the biomarker detection via traditional image analysis algorithms a challenge.

As an introduction to TMAs, the OCa specimens mentioned above are produced by taking needle biopsies as small as 0.6 mm in diameter from regions of interest. These tissue cores are then inserted in a recipient paraffin block in a precisely spaced array pattern where sections are cut using a microtome and mounted on a microscope slide. As shown in Figure 3.1, each slide can contain 100–500 samples, are over 26 gigabytes (uncompressed), and have typical sizes of 11500 x 78000. There is research underway to push the amount of these samples (or spots) on a TMA to over 10,000 [55]. For illustrative purposes, consider having 10 patient studies each with a TMA of 500 cylinders. Overall, there are 5000 1500×1500 images to analyze. An expert clinician could expect to invest 5 minutes per image, thus resulting in over 400 hours to analyze all of the data.

As an end result, in order to provide any type of statistical analysis for patient prognosis, a large number of these TMAs must be analyzed in a high throughput, yet reproducibly accurate manner. The new tools developed to combat these difficulties should provide accurate results that could lead to a standard approach, such that the results can be shared comfortably between institutions.

Computationally speaking, a grid alignment technique for TMAs has already been proposed by [18] making it trivial to analyze this one large TMA image as a set of smaller spots

by extracting each of them as a separate image. Since each of these specimens is operated on individually but has the identical algorithm performed on it, the process becomes an obvious choice for parallel computing, further motivating the need for high-throughput, accurate algorithms. When we consider our proposed algorithm, using a laptop, we could complete the same aforementioned work in about 10 hours. Using a standard 8 core machine, the analysis could be completed in just over an hour.

3.2 Challenges and Contributions

The major contribution of this chapter is a fast, novel, hierarchical unsupervised segmentation method (HNCut), which we demonstrate with an application in ovarian TMAs. Unlike traditional clustering algorithms, we aim to extract a single cluster pertaining to the stained region, while ignoring pixels in all other clusters. In traditional algorithms, pixels are allocated to the cluster that they are least dissimilar with, as opposed to being removed. Our setup encourages cuts that confidently trim away these undesired pixels. The unsupervised aspect of HNCut is particularly desirable for image analysis applications in histopathology and TMAs where obtaining annotated samples for training a supervised classifier depends on the annotations provided by an expert and hence difficult to obtain.

Our region of interest, the reactive area, is a chemically stained dark brown region (Figure 3.1). The light brown areas are to be ignored as they are considered to be artifacts. Our goal then becomes the robust rapid extraction of this stained region given minimal domain knowledge from a layman.

3.2.1 Contributions

This is the first attempt at combining a frequency weighted MS (FWMS), also a contribution in this thesis, with an existing partitioning algorithm for the task of segmentation. With FWMS accomplishing the same clustering task as MS, but doing so significantly more efficiently, previously intractable images become tractable. FWMS exploits the fact that as each iteration of MS completes, more points converge. We demonstrate how the convergence of our novel FWMS scheme allows us to perform clustering, the first step of our HNCut segmentation approach, 15 times faster than the traditional MS algorithm [38]. We can see from the run times presented in [38] that for a 240x160 pixel image the running time is 2.18 seconds. By working directly

in the color space, and using FWMS as the first stage, we can perform a similar segmentation operation in 6 seconds on an image 58 times larger.

We can thus summarize the important methodological and clinical contributions juxtaposed by their challenges as:

- A novel hierarchical segmentation approach that marries our innovative Frequency Weighted Mean Shift with the well-known Normalized Cuts (HNCut). This speaks directly to the nature of the large data size and efficiency needed as HNCut not only operates on large (1.5 million or greater) images in under 10 seconds, but is easily scalable to entire TMAs. The affinity matrix can now take advantage of multiple features, and multiple color spaces efficiently across large window sizes.

- Parameter insensitive segmentation for large images and the ability of HNCut to discriminate between regions with similar color values provides a robust approach in a domain where there are great variances in lighting, staining, and human preparation protocols. The parameter for the Gaussian kernel in the affinity matrix of NCut is automatically computed. The parameters for the mean shift are automatically adjusted based on the variance of the output.

- Layman initialization of the system is possible, obviating the need for detailed ground truth annotation from an expert that is required for more sophisticated supervised classifiers. With this constraint removed, the investment required for each additional stain/domain is minimal.

- The first attempt, to our knowledge, to precisely quantify in a reproducible manner a vascular marker on OCa TMAs. Given a wide range of input parameters we show small variance in results, making the approach amenable to the ultimate objective of creating a trustworthy quantitative image based metric for OCa prognosis and survival.

3.3 Overview

Figure 3.2 presents a high level overview of the 4 stages associated with the HNCut algorithm. Each of these stages are discussed in detail in the following subsections. We present an overview here to guide the reader through the various stages.

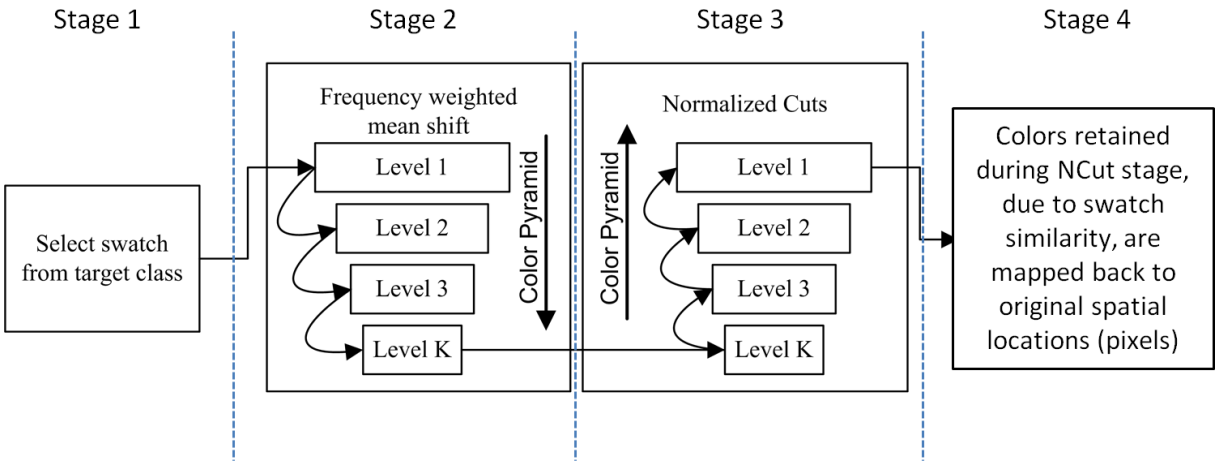


Figure 3.2: A flow chart of the HNCut process. Proceeding left to right, the user selects the domain swatch, which then gets fed into our FWMS with the image's pixel values. This results in the original image being decomposed into multiple levels of color resolution, which is then followed by the application of NCuts at each of the color resolutions generated. At each pyramid level colors not deemed to be part of the swatch are eliminated. Following the application of NCuts on the color pyramid (from the lowest to the highest color resolution), the color values that have not been eliminated are mapped back to the spatial domain via their original pixel locations, and the final segmentation is obtained.

We start by requiring the user to select a few sample pixels from the target class, termed a *swatch*, from an image. We use these pixels to guide the subsequent pixel classification process across all images in the same domain. Intuitively, one may think of this sample as the foreground.

Next, we employ our novel version of a mean-shift algorithm, frequency weighted mean shift (FWMS), on the color values in the image to form a hierarchical data structure (represented by the levels in the color pyramid in the second box in Figure 3.2). Intuitively, the FWMS algorithm allows for identification of color values which are within some specified tolerance of each other and assigns them to the same mode. A popular graph partitioning algorithm, normalized cuts [27] (NCuts), then operations on only the unique values at each level of the pyramid, as opposed to all possible color values, allowing for a factorization resulting in significantly fewer computations. This reduction is key as it allows NCuts to operate on images which were previously infeasible due to data size constraints. An illustration of the application of the scheme to an OCa TMA, for detecting a TVM, is illustrated in Figure 3.3. We then compute the weight for each unique mode, which reflects the actual frequency of the number of pixels associated with it.

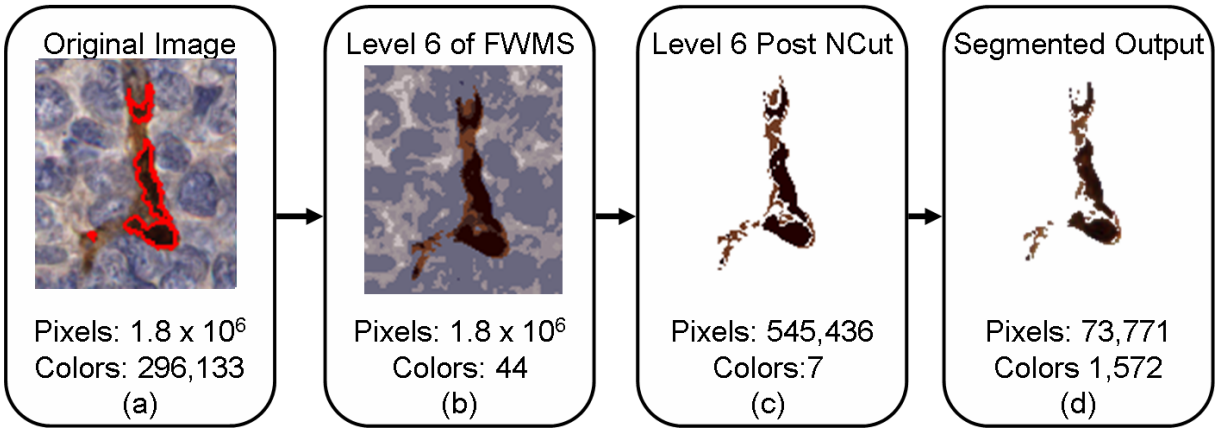


Figure 3.3: (a) Original image with desired TVM stain enclosed in red, (b) image at the bottom of the color pyramid during FWMS, (c) image at the bottom of the color pyramid following application of NCuts, (d) final segmentation results obtained by mapping colors not eliminated by HNCut spatially onto the original image. Note that between (a) and (b) a significant reduction in color resolution occurs, which allows NCuts to be performed on an image with several orders of magnitude fewer colors compared to the original image (a). NCuts is then applied at progressively higher color resolutions, while at each pyramid level colors not deemed to be part of the swatch are eliminated. The colors retained at the highest resolution are then spatially mapped onto the corresponding pixels to yield the final segmentation.

Using this pyramid we can drastically reduce the large segmentation problem in the color space to a set of much smaller graph partitioning problems (the third box from the left in figure 3.2), which we show can be solved far more efficiently by NCuts. By starting at the bottom of the pyramid, we partition the unique values (typically on the order of 10 values) into two sets such that all of the values selected by the user in the first step are assigned to the first partition. Subsequently, we eliminate the second partition and map the colors in the first partition to an immediately higher color resolution level in the pyramid. This process continues until the entire pyramid is traversed. The last step involves mapping the color values not eliminated back into the spatial domain.

The hierarchical set of operations described above makes for an extremely efficient and accurate algorithm; thus applying the NCut at the lowest levels of the pyramid is relatively simple to do and encourages a more sophisticated definition of pixel affinity. While in this work only chromatic information was leveraged, the method is easily and efficiently extensible to incorporate additional image features (e.g., texture).

Figure 3.3 displays an image from our dataset undergoing the HNCut procedure, with the

intent of quantification of the vascular marker stain (brown color). The numbers shown in the boxes in Figure 3.3 represent the reduced number of colors and pixels generated by the HNCut scheme at different levels of the pyramid within a single cylinder (1500 $\times$ 1500 pixels, 300,000 colors) from a TMA.

3.4 Theory and Algorithms

3.4.1 Notation

An image scene is defined as $\mathcal{C} = (C, \mathbf{f})$ where C is a 2D Cartesian grid of N pixels, $c \in C$, where $c = (x, y)$. $\mathbf{f}$ is a color intensity function, where $\mathbf{f} \in \mathbb{R}^3$. We define as $\mathbf{F}_1 \in \mathbb{R}^3$ the vector of colors associated with all pixels $c \in C$ at the full color resolution (top of the color pyramid). The elements of $\mathbf{F}_1$, namely $f_{1,i}$, are derived such that for pixel c_i, $f_{1,i} = \mathbf{f}(c_i)$ and $f_{1,i} \in \mathbb{R}^3$.

3.4.2 Integrating Domain Knowledge to Guide Normalized Cuts

A user via manual selection defines a small color swatch $S_1 = \{f_{1,\alpha_\tau} | \alpha_\tau, \tau \in \{1, \ldots, N\}\}$ where α_τ is an index value to the original color vector. Note that S_1 is easily obtained by annotating (manually) a few pixels from the object of interest on a representative image and may be easily changed based on the application. As we will describe in further detail later, S_1 is only used to identify which color partition (A or B from Eq. 3.8) to retain during NCut. It is important to note that since S_1 is a reference to a subset of the color values in the original image, it is available at, and undergoes, all steps of the HNCut algorithm. will undergo all of the MS and NCut operations presented below. Note that S_1 is the swatch originally defined by the user at the full resolution, $k = 1$.

3.4.3 Frequency Weighted Mean Shift (FWMS)

Theory

The mean shift algorithm is used to detect modes in data using a density gradient estimation. For a more detailed explanation of the algorithm we refer the reader to [36].

We start with the fixed point iteration update $\forall j \in \{1, \ldots, N\}$ in MS (described in [36]) as

$$f_{k+1,j} \leftarrow \frac{\sum_{i=1}^{N} f_{k,i} G(f_{k,j} - f_{k,i})}{\sum_{i=1}^{N} G(f_{k,j} - f_{k,i})}, \tag{3.1}$$

where G is a Gaussian function with a bandwidth parameter σ_{MS}, which is used to compute the kernel density estimate at data point c_j, $G(f_{k,j} - f_{k,i})$=$\exp(-\frac{||f_{k,j} - f_{k,i}||_2}{\sigma_{\mathrm{MS}}^2})$, with $||\cdot||_2$ representing the $L2$ norm. $k \in \{1, \ldots, K\}$ represents various levels of color resolution produced at each iteration. The overall computation time for Equation 3.1 is $O(N^2)$. By employing the Improved Fast Gauss transform (IFGT) [37], we can reduce the computation complexity to $O(N)$ with minimal precision loss.

It becomes possible to exploit the fact that after each iteration of the MS many of the data points, in our case color values, converge. If we consider what that convergence means mathematically, essentially two points c_{β_1}, c_{β_2}, where $\beta_1, \beta_2 \in \{1, \ldots, N\}$ meet the requirement that $|f_{k,\beta_1} - f_{k,\beta_2}| \leq \epsilon$ where ϵ is a pre-defined tolerance value. We can thus rewrite the numerator of Eq. 3.1, which is

$$\begin{aligned} f_{k,\beta_1} G(f_{k,j} - f_{k,\beta_1}) + f_{k,\beta_2} G(f_{k,j} - f_{k,\beta_2}) \\ + \sum_{i=1, i \neq \beta_1, \beta_2}^{N} f_{k,i} G(f_{k,j} - f_{k,i}), \end{aligned} \tag{3.2}$$

in the form:

$$2 f_{k,\beta_1} G(f_{k,j} - f_{k,\beta_1}) + \sum_{i=1, i \neq \beta_1, \beta_2}^{N} f_{k,i} G(f_{k,j} - f_{k,i}), \tag{3.3}$$

thereby avoiding the explicit calculation of $G(f_{k,j} - f_{k,\beta_2})$ where $j, \beta_1, \beta_2 \in \{1, \ldots, N\}, k \in \{1, \ldots, K\}$. This results in one less computation for the Gaussian, which is by far the most expensive operation in the entire MS clustering process. The formulation in Equation 3.3 results in a significant computational efficiency improvement. The computational savings apply to the denominator as well, as it follows the same reduction.

As a result, we may rewrite the update presented in Equation 3.1 as a multi step update. Initially, we determine the unique values in iteration k (i.e., $\mathbf{F}_k$) under the constraint that any color values $|f_{k,i} - f_{k,j}| \leq \epsilon$ are considered equivalent. Thus from $\mathbf{F}_k = \{f_{k,1}, f_{k,2}, \ldots, f_{k,|\mathbf{F}_k|}\}$ we can construct the vector $\hat{\mathbf{F}}_k$, where $\hat{\mathbf{F}}_k \subset \mathbf{F}_k$ and $\hat{\mathbf{F}}_k$ is a set of only unique values in $\mathbf{F}_k$, with $|\hat{\mathbf{F}}_k| = M_k$. A weight vector $\mathbf{w}_k = \{w_{k,1}, \ldots, w_{k,M_k}\}$ is then computed for $\hat{\mathbf{F}}_k$ as

$$w_{k,j} = \sum_{i=1, f_{k,i}=\hat{f}_{k,j}}^{|\mathbf{F}_k|} w_{k-1,i}, \tag{3.4}$$

where $j \in \{1, \ldots, M_k\}$ and $\mathbf{w}_0 = \mathbf{1}$, since at $k = 1$ each color value has equal weighting. Equation 3.4 is summing the weights from the previous level into the new unique values that resulted from the next iteration of mean shifting. As a result, $w_{k,j}$ contains a count of the number of original pixels that have migrated to $f_{k,j}$ through mean shifting. Now, the number of points in the system that have converged to some intensity (color) value $\hat{f}_{k,j}$ is represented by $w_{k,j}$. It is important to note the following definition of M_k where

$$|\mathbf{w}_k| = |\hat{\mathbf{F}}_k| = |\mathbf{F}_{k+1}| = M_k, \tag{3.5}$$

and

$$\sum_{i=1}^{M_k} w_{k,i} = N, \tag{3.6}$$

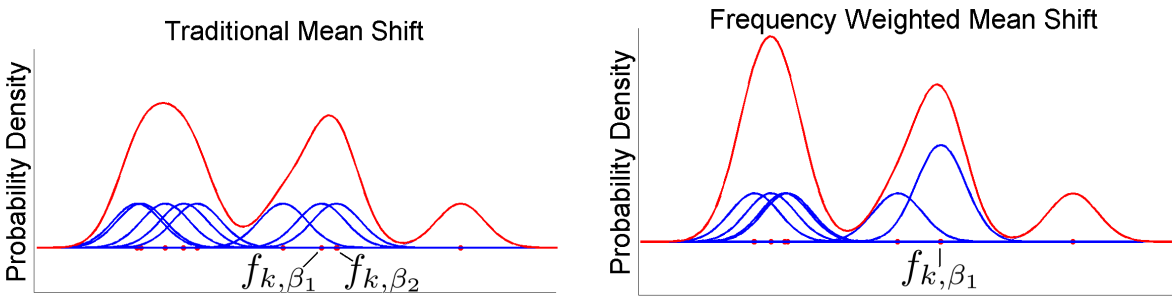
which leads us to the update of Equation 3.1:

$$f_{k+1,j} \leftarrow \frac{\sum_{i=1}^{M_k} w_{k,i} \hat{f}_{k,i} G(\hat{f}_{k,j} - \hat{f}_{k,i})}{\sum_{i=1}^{M_k} w_{k,i} G(\hat{f}_{k,j} - \hat{f}_{k,i})}, \tag{3.7}$$

for $j \in \{1, \ldots, M_k\}$.

An illustration of the steps described in Equations 3.3-3.7 is presented in Figure 3.4. The images depict a standard probability density function (PDF in red) computed from the Gaussian contributions (in blue) from the 1 dimensional data points (red circles). From Figure 3.4(a) we can see that colors f_{β_1} and f_{β_2} will converge in the next iteration of the MS. We exploit the fact that once f_{β_1} and f_{β_2} converge, it becomes possible to factor out f_{β_2} from the system, and move its contribution into f_{β_1}, without altering the distribution (Figure 3.4(b)).

We call this new approach the Frequency Weighted Mean Shift (FWMS). FWMS helps to produce a pyramidal scene representation $\mathcal{C}_k = (C, \mathbf{F}_k)$, where $k \in \{1, \ldots, K\}$ represents K levels of the color pyramid. Note that $M_1 \geq M_2 \geq \ldots \geq M_K$, indicating level 1 has the most colors and M_K the least. In other words, FWMS results in a series of scenes $\mathcal{C}_k$, all mutually aligned, but with a smaller number of colors in $\{\mathcal{C}_K, \mathcal{C}_{K-1}, \ldots\}$ compared to $\{\mathcal{C}_1, \mathcal{C}_2, \ldots\}$, which allows for NCut to be tractable. The FWMS algorithm is given in Algorithm 1.



(a) (b)

Figure 3.4: A visual representation of the probability density functions (pdf) illustrating the difference between the (a) traditional MS and the (b) frequency weighted MS. The red circles on the x-axis are the given values in a 1 dimensional system, the blue arcs are the associated Gaussian contributions, while the red line above represents the summation of all of the contributions, i.e., the pdf. In (b), when points f_{β_1} and f_{β_2} converge, f_{β_2} is removed from the system, and its contribution is moved into f_{β_1} as a multiplication, avoiding an additional expensive step in the computation of the Gaussian pdf.

Algorithm

The convergence requirement stated in line two of Algorithm 1 may be specified via three possible criteria. The first is the maximum number of iterations, a number specified by the user. The second more common approach is to stop the algorithm when the difference between any two iterations falls below a pre-defined threshold (i.e., the amplitude of the migrations associated with each point reduces significantly). Lastly, convergence can be reached when the number of elements in $\hat{\mathbf{F}}$ becomes small enough that additional clustering provides no efficiency benefit as the overhead in the NCut starts to outweigh the computation time. This process was illustrated in Figure 3.3 as the sequence of steps going from (a) to (b). It may be seen from Figure 3.3 that the overall color resolution is significantly reduced as the algorithm proceeds from level 1 to level K. In this example, the original image containing about 300,000 unique color values was reduced to 44 unique values. This significantly smaller set of values makes the NCut step tractable since we operate directly in the color space.

Algorithm 1 Frequency Weighted Mean Shift to Generate Color Pyramid

Input: $\mathbf{F}_1$ of $\mathcal{C}_1$

Output: $\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \ldots, \hat{\mathbf{F}}_K$

1: $k = 1$

2: **while** not converged **do**

3: Compute the unique values of $\mathbf{F}_k$ and store them in $\hat{\mathbf{F}}_k$

4: Compute frequency of colors in $\hat{\mathbf{F}}_k$ as they appear in $\mathbf{F}_k$ using Eq 3.4, store in $\mathbf{w}_k$

5: Generate $\mathbf{F}_{k+1}$ using Eq 3.7

6: $k = k + 1$

7: **end while**

8: **return** $\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \ldots, \hat{\mathbf{F}}_K$

3.4.4 Normalized Cuts on Frequency Weighted Mean Shift Reduced Color Space

Theory

Normalized cuts [27] is a graph partitioning method, used to separate data into disjoint sets. For our problem, the hierarchical pyramid created by FWMS at various levels of color resolution ($\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \ldots, \hat{\mathbf{F}}_K$) serves as the initial input to the NCut algorithm. NCut takes a connected graph **G=**(E, V), with vertices (V) and edges (E) and partitions the vertices into disjoint groups. By setting V equal to the set of color values $\hat{\mathbf{F}}_K$, and having the edges represent the similarity (or affinity) between the color values, we can separate the vertices into groups of similar color values. A normalized cut is defined as the process by which the removal of edges leads to two disjointed partitions A and B such that the variance of values (in our case colors) in A and B are minimized and the difference in average value (intensity of colors) between A and B is maximized. We present the high level formulation as described in [27]:

$$\text{NCut}(A, B) = \frac{\text{cut}(A, B)}{\text{assoc}(A, V)} + \frac{\text{cut}(A, B)}{\text{assoc}(B, V)}, \tag{3.8}$$

where cut describes the affinity between the sets, encouraging higher dissimilarity between sets, and assoc describes the affinity between a set and the whole system, encouraging sets of significant size. The ψ function is used to define the affinity between two points. Our ψ function is defined as:

$$\psi(\hat{f}_{k,i}, \hat{f}_{k,j}) = \exp(-\frac{||\hat{f}_{k,i} - \hat{f}_{k,j}||_2^2}{\sigma_{Ncut}}) \tag{3.9}$$

with σ_{Ncut} as a bandwidth parameter. It is worth noting that in the traditional NCut paper [27], their affinity calculation took into account both a spatial and color component. For even small images, this made the affinity matrix intractable. As a result, the ψ function had a spatial constraint introduced such that Equation 3.9 is set to zero if the associated pixels are farther away than a user specified distance. This constraint forced the affinity matrix $\mathbf{\Psi}$ to typically be sparse, making its storage and subsequent operations applied to it less burdensome. Nevertheless, for large images, the affinity matrix is still too large (in spite of the spatial constraints), and as such we choose to operate solely in a significantly reduced color space, without the imposition of spatial constraints. In Figure 3.3, we can see at the bottom of the hierarchical pyramid for a color image with original dimensions of 1200×1200 with about $300,000$ unique colors, we would have an affinity matrix of only 7×7, and at the highest level a size of 1572×1572 instead of the naive NCuts implementation resulting in an affinity matrix of $300,000 \times 300,000$.

Algorithm

The main steps comprising the HNCut technique are shown in Algorithm 2. We begin by applying NCut on the lowest image resolution generated in the above section, by setting $k = K$, $V_k = \{\hat{f}_{k,1}, \hat{f}_{k,2}, ..., \hat{f}_{k,M_k}\}$, i.e., the set of unique color values present at level K from FWMS.

Step 1: We apply NCut to partition the scene into two disjoint color sets A and B, where $A, B \subset V_k$. To perform this partition, we compute the affinity matrix $\mathbf{\Psi}_K \in \mathbb{R}^{M_k \times M_k}$ using Equation 3.9 for all $i, j \in \{1, \ldots, |V_k|\}$. σ_{NCut} is a scaling parameter set to some initial value.

Step 2: As a result of the partitioning, we need to identify if either A or B uniquely contains all colors in S_k, our user selected color swatch at level k of our hierarchy. Hence if $S_k \subseteq A$ and $S_k \cap B = \emptyset$ then eliminate all colors in B by setting $V_k = A$. If $S_k \subseteq B$ and $S_k \cap A = \emptyset$, similarly eliminate A by setting $V_k = B$. However if S_k is not uniquely contained in either A or B, we increase σ_{NCut} and proceed back to Step 1. We keep incrementing σ_{NCut} until S_k is uniquely contained within either of A or B, and set V_k to that partition.

Step 3: Begin the process again with the new V_k until no further partitioning of the color space at level k is possible; that is until S_k cannot be contained uniquely within a single color partition for any value of $\sigma_{\text{NCut}} < \sigma_{\max}$.

Step 4: Using this process, we sequentially climb the hierarchical data structure $\hat{\mathbf{F}}_k$ where $k \in$

$\{1, ..., K\}$. Thus, we migrate to the next higher image resolution, level $k-1$ and set V_{k-1} to V_k, i.e., the set of colors retained at resolution level k, and repeat the process again. We return to Step 1 until k = 1.

Step 5: At level 1, V_1 contains a subset of values from $\hat{\mathbf{F}}_1$, which are considered to be the chromatic values of the region of interest. Thus the final image is computed by retaining all pixels $j \in \{1, \ldots, N\}$ such that $f_{1,j} \in V_1$, and eliminating the others.

Algorithm 2 NCuts on FWMS Reduced Color Space

Input: $\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \ldots, \hat{\mathbf{F}}_K, S_1$

Output: $\mathbf{V}_1$ is returned, which contains all retained color values

$k = K$
$V_k = \hat{\mathbf{F}}_k$
Using Equation 3.9 build $\mathbf{\Psi}_k$ from V_k
while $k \neq 1$ **do**
  $\sigma_{\text{NCut}} =$ intial σ value
  while $\sigma_{\text{NCut}} < \sigma_{\max}$ **do**
    Solve for A, B by using Eq. 3.8
    if S_k is not uniquely contained in A or B **then**
      Increase σ_{NCut} by a factor of 10
    else
      $V_k = \begin{cases} A, & \text{if } S_k \subseteq A \\ B, & \text{if } S_k \subseteq B \end{cases}$
    end if
    Using Equation 3.9 re-construct $\mathbf{\Psi}_k$ from V_k
  end while
  $k = k - 1$
  $V_k = \hat{f}_{k,i}, \forall i$ where $f_{k+1,i} \in V_{k+1}$
  Using Equation 3.9 re-construct $\mathbf{\Psi_k}$ from V_k
end while
return V_1

Chapter 4

HNCut: Experiments & Results

4.1 Dataset

Our image database comprises of a total of seven digitized TMAs of ovarian cancer (OCa), in turn comprising a total of over 500 tissue cylinders from 100 patients, from which 130 were randomly selected for performing quantitative evaluation (qualitative evaluation was done on all 500). Only 130 of them were submitted to our pathologist for annotation due to the laborious nature of the work, which further motivates the utility and clinical motivation for HNCut.

The TMAs were obtained by sampling OCa tissue and were stained for the presence of the TVM ESM-1, resulting in vascular regions with the antibody to ESM-1 staining brown. The digitized version of the TMAs were obtained by scanning the slides at 40x resolution on a whole slide digital scanner, but subsequently these were down-sampled and stored at 20x magnification. This resulted in over 500 digital images of individual cylinders, each of which were approximately 1,500 $\times$ 1,500 pixels in dimension. An expert pathologist annotated the precise spatial extent of the TVM on all of the 130 tissue cylinders considered for the test. Care was taken by the pathologist to include even those instances where only a few isolated pixels were picked up by the TVM.

4.2 Implementation

All experiments were run on a 2.8Ghz Linux machine running Matlab 2008b with 32Gb of RAM. The setup of HNCut was as follows. All experiments were performed after converting the RGB input images to the HSV colorspace, though the algorithm is extensible to scalar valued

images (such as grayscale images) as well. [1] The FWMS was performed using $\sigma_{\mathrm{MS}} = .05$. NCut was subsequently performed using the Silverman function [56] to determine the value for the initial σ_{NCut}, which was then incremented by a factor of 10 as prescribed in step 9 in Algorithm 2. The Improved Fast Gauss Transform's clustering variable, as suggested by [37], was set to the square root of the number of data points. When the number of remaining clusters fell below this value, it was reset to the square root of the number of remaining clusters.

The procedure that we used to enforce the ϵ distance for the generation of $\hat{\mathbf{F}}_k$ from $\mathbf{F}_k$ as discussed in Section 3.4.3 was implemented as follows. Since the human visual system is unable to easily discriminate between subtle variations of the same color, we can set ϵ to a relatively large value. The easiest way to apply this ϵ requirement in an algorithmic form is to simply choose the desired precision level (such as 10, 0, .01 or .001, depending on the format of the data) and then simply round to that level. Since our data is stored using double precision in the range [0, 1], we have used the thousandths decimal place. The subsequent procedure of locating unique values and computing their frequencies is as simple as generating a histogram of the data values with each unique value occupying its own bin. This is a significant benefit, as the production of histograms is not only well studied but easily transformable into a parallel computing problem [57].

4.3 Evaluation Description

A total of 4 experiments were conducted to evaluate the accuracy, efficiency, and reproducibility of the HNCut algorithm, specifically in terms of its ability at (1) identifying pixels whose colors are within the swatch and also in terms of (2) identifying contiguous vascular regions annotated by the pathologist. It was felt that both pixel level and region level statistics were required to comprehensively and reliably evaluate HNCut performance.

[1] Experiments were run using RGB space with less success, the main issue being that the colorspace is not flat or intuitive, so as the colors were mean-shifted unnatural combinations were created. This is especially notable when trying to obtain brown stain in an H&E image where the red dye becomes quite similar to the brown stain in RGB space.

Region level metric

We define $R^{a,\varsigma}$ as the regions identified by HNCut and $R^{b,z}$ as the corresponding expert annotated regions, with $z \in \{1, \ldots, Z\}$ and $\varsigma \in \{1, \ldots, \urcorner\}$, where Z and $\urcorner$ are the number of regions in the HNcut and expert annotated images, respectively. If for any $R^{b,z}$, $\frac{|R^{b,z} \cap R^{a,\varsigma}|}{|R^{b,z}|} > 0.3$ then $R^{a,\varsigma}$ is identified as a true positive (TP). If for any $R^{a,\varsigma}$ there is no $R^{b,z}$ for which this condition is satisfied then $R^{a,\varsigma}$ is identified as a false positive (FP). If there is a $R^{b,z}$ for which no $R^{a,\varsigma}$ can be found that satisfies the above condition, $R^{b,z}$ is deemed to be a false negative (FN). The .3 threshold was experimentally determined based on interactions with our expert pathologist. The complex nature of the stain shapes necessitated a lower threshold.

Pixel level metric

Pixel-level statistics are defined using the basis of P^a and P^b, a collection of all pixels in the segmented result ($\cup_{\varsigma=1}^{\urcorner} R^{a,\varsigma}$) and the ground truth ($\cup_{z=1}^{Z} R^{b,z}$), respectively. From there we can define, the True positive rate (i.e., sensitivity, $\frac{|P^a \cap P^b|}{|P^b|}$), Positive predictive value ($\frac{|P^b|}{|P^b|+|P^a-(P^a \cap P^b)|}$), False negative rate ($\frac{|P^b-(P^a \cap P^b)|}{|P^b|}$) and True negative rate (i.e., specificity, $\frac{|C-(P^a \cup P^b)|}{|C-P^b|}$). In all cases the $|\circ|$ notation defines the cardinality of the set.

4.4 Comparative Strategies

PBT was implemented as described in [24] using suggested default values for both of PBT's variables θ and ϵ (.45 and .4 respectively). PBT iteratively generates a hierarchical tree structure in the training stage where each node of the tree is converted into an Adaboost classifier [58] constituting 7 weak classifiers. During testing, the conditional probability of the sample belonging to the target class is calculated at each node based on the learned hierarchical tree. The discriminative model is obtained at the top of the tree by combining the probabilities associated with probability propagation of the sample at various nodes. Unlike other commonly used classifiers, such as AdaBoost [58] and decision trees [59], which provide a hard binary classification, PBT generates a posterior conditional probability value $p(1|c), p(-1|c) \in [0, 1]$, for each sample c as belonging to one of two classes. The feature vector was created by taking a 3×3 window around every $c \in C$, across all 3 color channels in HSV space, resulting in a 27 dimensional vector. 1000 random positive (stained) samples and 1000 random negative (unstained and spuriously stained) samples were selected from 25 randomly selected images,

resulting in a total training vector of size 27 $\times$ 50,000. Training and testing was done via 50 runs of cross validation. This consisted of randomly selecting 25 images and training the classifier as described above, followed by testing on the other 105 images. The probabilities returned by the PBT were subjected to thresholds at 92% and 97% (represented via the first two columns in Figures 4.2, Figures 4.3 and Figures 4.4). The choice of thresholds was determined as follows. During each run of the randomized cross validation, a receiver operating characteristic (ROC) curve (representing the tradeoff between sensitivity and specificity) was generated and the threshold was set at the determined operating point. This value was found to range between 92% and 97%.

4.5 Experiment 1: Comparison of HNCut to PBT and k-means

Design

We compared the detection performance of HNCut with k-means and PBT. A standard k-means algorithm [60] was performed using 10 clusters. Since k-means is not deterministic and is notoriously sensitive to the choice of cluster centers, offline experiments were performed to identify initial cluster centers (cluster centers being identified both within and outside of the target object of interest), which were qualitatively determined as being optimal.

A subset of qualitative segmentation results are presented in Figure 4.1. The first column represents the original stained TVM OCa image cropped to an area of interest, with the boundary of the ground truth highlighted by the pathologist labeled in red. The first row illustrates a case where all of the algorithms performed comparatively. The second and third rows illustrate instances where the HNCut algorithm performs better compared to PBT and k-means, both of which yield several false positives. The final row is used to illustrate a scenario where false negatives occur for all three methods. The middle region for the image in Figure 4.1(m) is correctly segmented in all algorithms, while the three other regions are incorrectly rejected. This specific image is a very difficult case as the stain in those regions is only barely visible to an expert. k-means results in the largest number of positives compared to the two other methods; a consequence of k-means requiring all pixels to be assigned to a cluster.

Figures 4.2,4.3, and 4.4 quantitatively illustrates the mean and variance of false negatives,

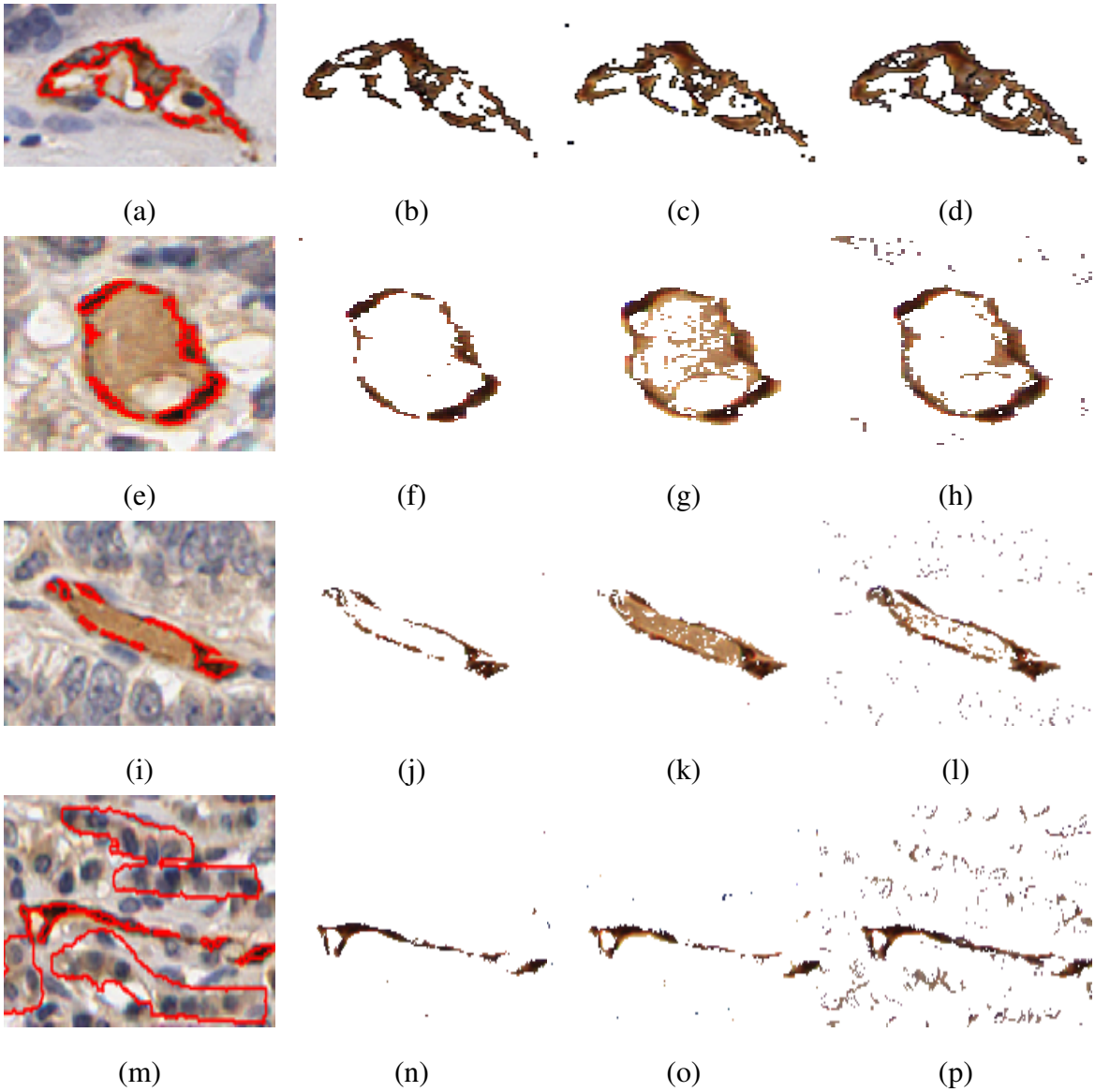


Figure 4.1: The first column ((a), (e), (i), (m)) represents the ground truth annotations of the vascular stained areas on 4 different cylinders. Columns 2-4 (left to right) represent corresponding segmentation results from HNCut ((b), (f), (j), (n)) for $\sigma_{\mathrm{MS}} = .05$, PBT ((c), (g), (k), (o)) at the 97% threshold, and k-means ((d), (h), (l), (p)) using 10 clusters. It can be seen that k-means always overestimates the stain extent, resulting in a large number of false positives. While PBTs perform better compared to k-means, (g) and (k) show how the PBT can occasionally retain spuriously stained pixels. On the other hand, HNCut's results closely resemble the ground truth. Note however that none of the algorithms are able to correctly identify the faintly stained regions in the upper portion of (m), since the stain there is barely discernible.

false positives and true positives, respectively, for the region level metric for the different algorithms across 10 runs. The red line indicates the mean value across all 10 runs, the blue box marks the positions where 25% of the 10 values on either side of the mean are encapsulated, and the black line extends to where 75% of the values that are on either side of the mean are contained. Thus, the closer the blue and black markers are to the red mean line, the more consistent the algorithm was able to perform. For the PBT this process involved 10 runs using different training and testing sets, while for HNCut we selected 10 different swatches. Finally, we note HNCut provides a similar mean for false negatives, while still providing a similar percentage for true positives. The false positive rate for HNCut versus PBT reveals that HNCut on average yields better performance, with a much smaller variance. The threshold of 92% for the PBT encourages few false negatives at the cost of many false positives. Figure 4.5 reveals that HNCut significantly outperforms both the PBT and k-means algorithms in terms of execution time.

Interestingly, randomly generating the training set for the PBT from the ground truths provided by the expert seems to lead to a larger variance in the false positive metric. This can be as a result of human error in performing the ground truth annotation, or in the selection of pixels that are not truly representative of the rest of the desired class.

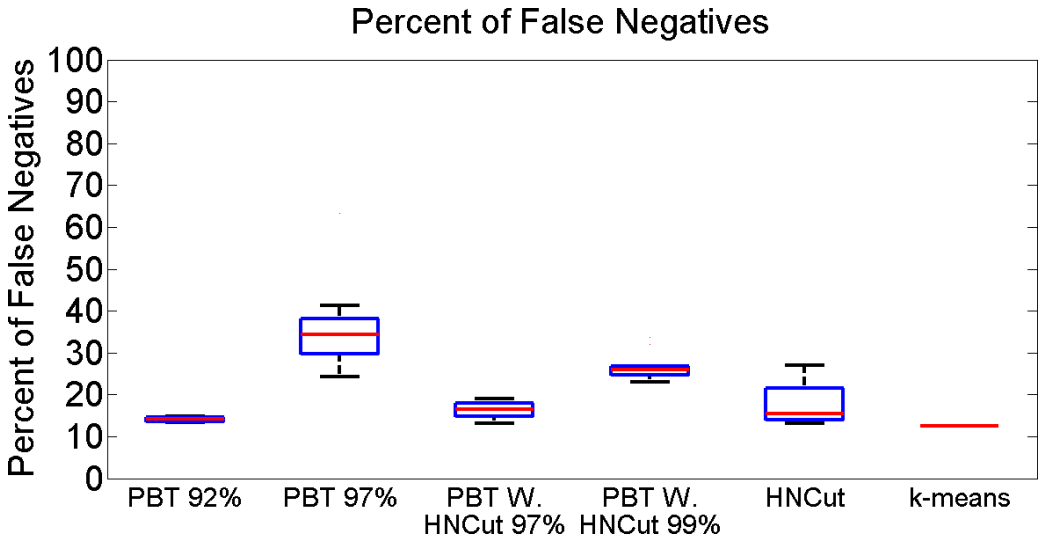


Figure 4.2: Mean and variance of the region-based performance measure for False Negatives over 10 runs for the PBT classifier (92% and 97% threshold), PBT classifier trained using HNCut (97% and 99% threshold), HNCut and k-means over 130 images.

It is also worth noting that k-means does quite poorly. There is no variance associated with the algorithm since we determined the optimal centers offline, thus removing the non-

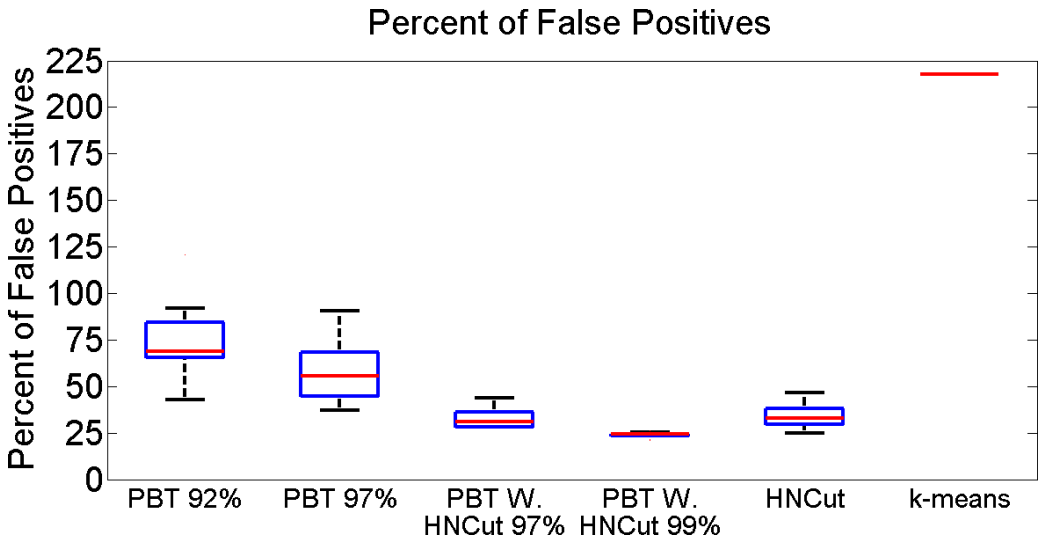


Figure 4.3: Mean and variance of the region-based performance measure for False positives over 10 runs for the PBT classifier (92% and 97% threshold), PBT classifier trained using HNCut (97% and 99% threshold), HNCut and k-means over 130 images.

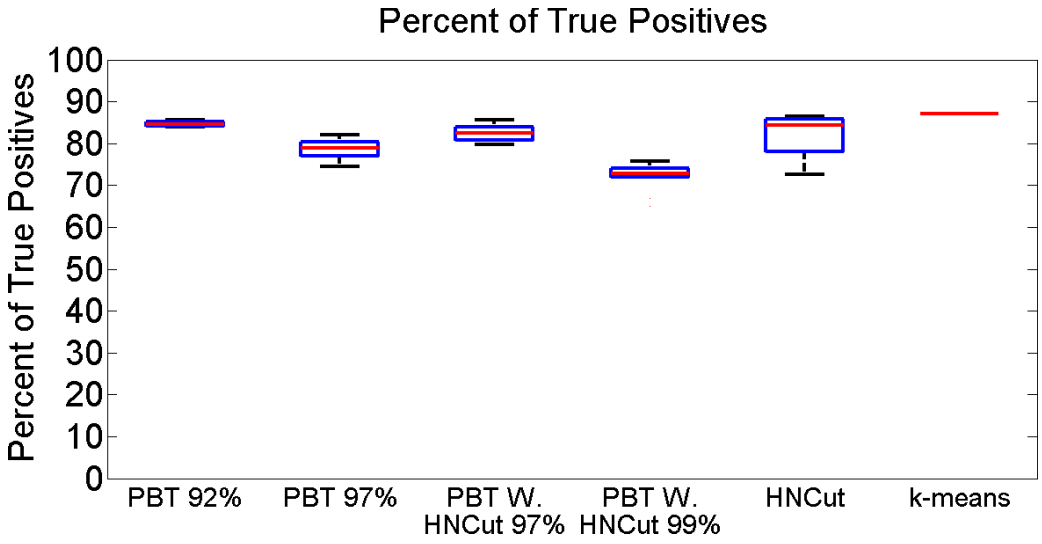


Figure 4.4: Mean and variance of the region-based performance measure for True Positives and over 10 runs for the PBT classifier (92% and 97% threshold), PBT classifier trained using HNCut (97% and 99% threshold), HNCut and k-means over 130 images.

deterministic aspect of the scheme. Figures 4.1 and 4.6 reveal the reason for the large number of false positives associated with k-means, it tends to retain many spuriously stained pixels (as visible by the light brown pixels) as being part of the target class.

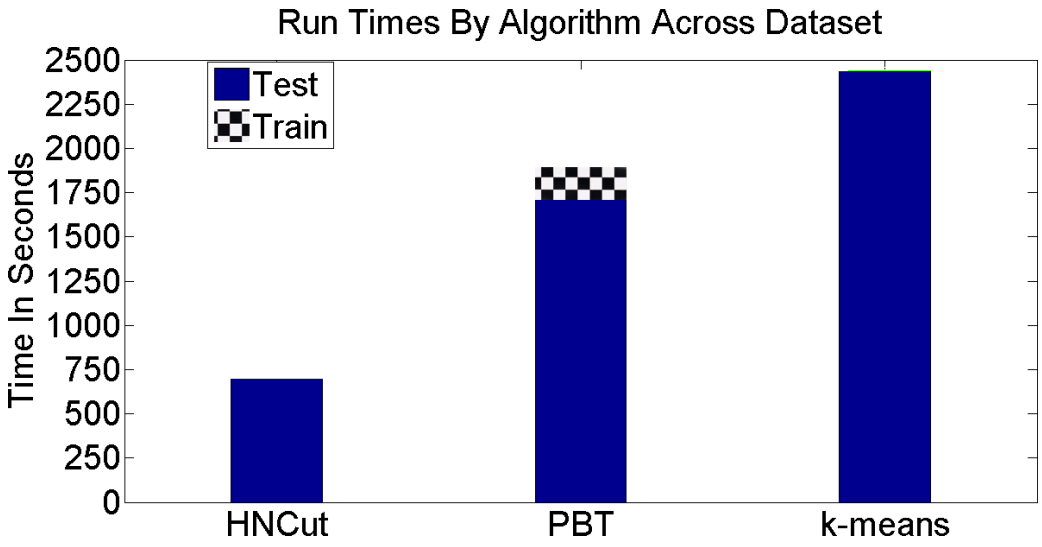


Figure 4.5: A comparison of computation times of each algorithm across 130 images reveals that HNCut significantly outperforms both the PBT and k-means algorithms in terms of execution time.

Pixel level performance measure

Table 4.1 quantitatively illustrates the mean and variance of the pixel level performance measure for the different algorithms across 10 sets of randomly selected training and test sets. HNCut's mean true positive rate (59%) places it in between the two PBT setups (63%, 51%), while still outperforming (99%) the true positive rate associated with the PBT on the two trials (98%, 98.3%). HNCut was intermediate in performance to the two runs of PBT in terms of positive predictive (36% versus 35% and 46%) and false negative rates (40% versus 36% and 47%). HNCut, thus, appears to provide a good balance between precision and recall.

4.6 Experiment 2: Reproducibility of HNCut with Respect to Swatch and Parameter Sensitivity

The results produced by HNCut are dependent upon the selection of the swatch and the size of the σ_{MS} bandwidth parameter. Clearly if there is a great deal of heterogeneity within the target class and the choice of swatch is not representative of the target class, the quality of the segmentation will be sub-optimal. Consequently, the user has the choice of either (a) sampling additional values corresponding to the target class, or (b) repeating the segmentation with HNCut a few times with different swatches until the desired target class segmentation is ob-

	True Positive Percentage	True Negative Percentage	Positive Predictive Percentage	False Negative Percentage
HNCut	59.24%±7.36	**99.01% ± 0.56**	36.34% ±52.30	39.95% ±7.36
PBT (92%)	**62.65% ± 0.87**	98.09% ±0.16	35.23% ±14.42	35.71%±0.87
PBT (97%)	51.72%±3.26	98.33%±0.07	46.70% ±5.08	46.65%±3.26
PBT W. HNCut (97%)	58.70%±4.09	98.46% ±0.06	56.18% ±6.44	39.66% ±4.09
PBT W. HNCut (99%)	46.03% ±2.53	98.56% ±0.02	**67.60% ± 2.29**	52.34% ±2.53
k-means	71.89%	97.56%	14.34%	**27.3%**

Table 4.1: Quantitative results presented for the pixel-level performance measure across all of the algorithms. The ± value is the percent variance associated with the difference in running the algorithms with 10 different training sets or swatches. From Equation 15, we see it is possible to obtain a value greater than 100% when the number of pixels identified are greater than the total number of pixels in the target region.

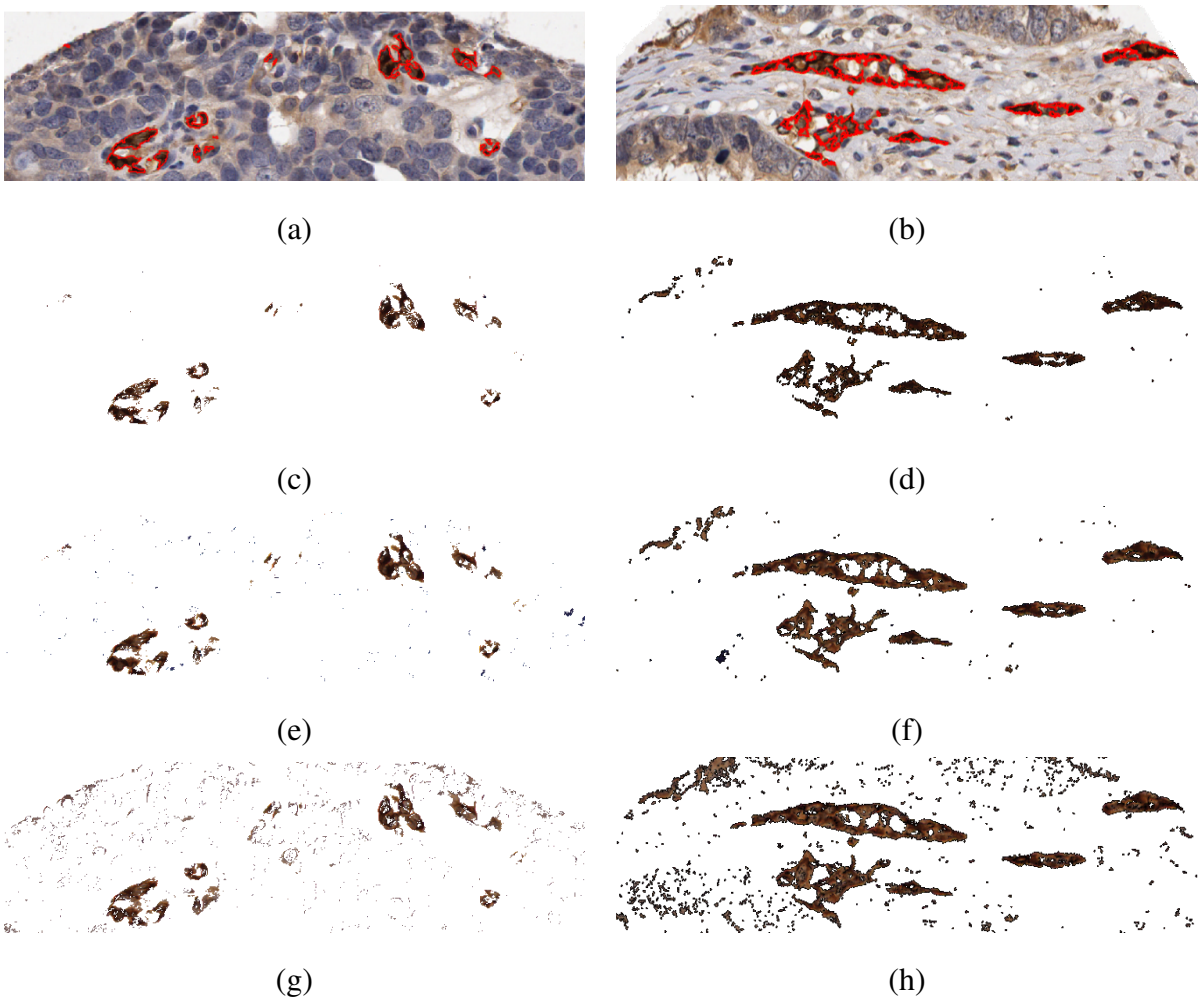
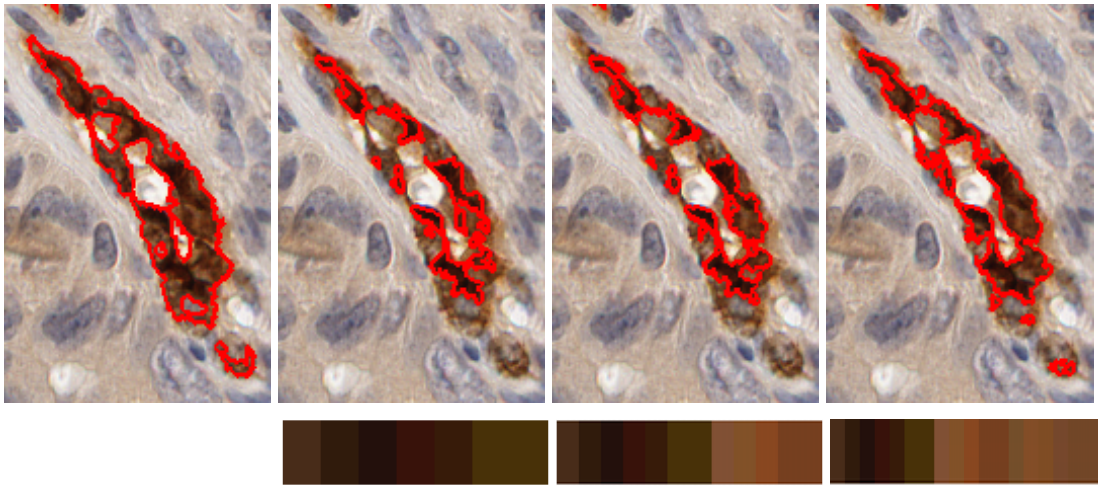


Figure 4.6: Two bands across selected TMA cylinders are presented. The (a), (b) original input, with the annotated ground truth in red, is presented on the top, followed by (c), (d) HNCut with $\sigma_{\mathrm{MS}} = .05$*, (e), (f) PBT at the 97% threshold and (g), (h) k-means using 10 clusters.*

tained. Note that both tuning procedures are only made possible by the superior computational efficiency of HNCut.

Swatch selection

Figure 4.7 shows qualitative results reflecting the sensitivity of the segmentation as a function of the choice of the swatch. A small patch was randomly selected from the desired class by a non-expert user. The resulting segmentation was overlaid using a red boundary on the original image. Subsequently, a few additional pixels were added to the swatch, and the segmentation repeated. In Figure 4.7(b), we can see that when the user selects dark pixels within the swatch, the segmentation focuses on the darker aspects of the stain. When the swatch shown in Figure



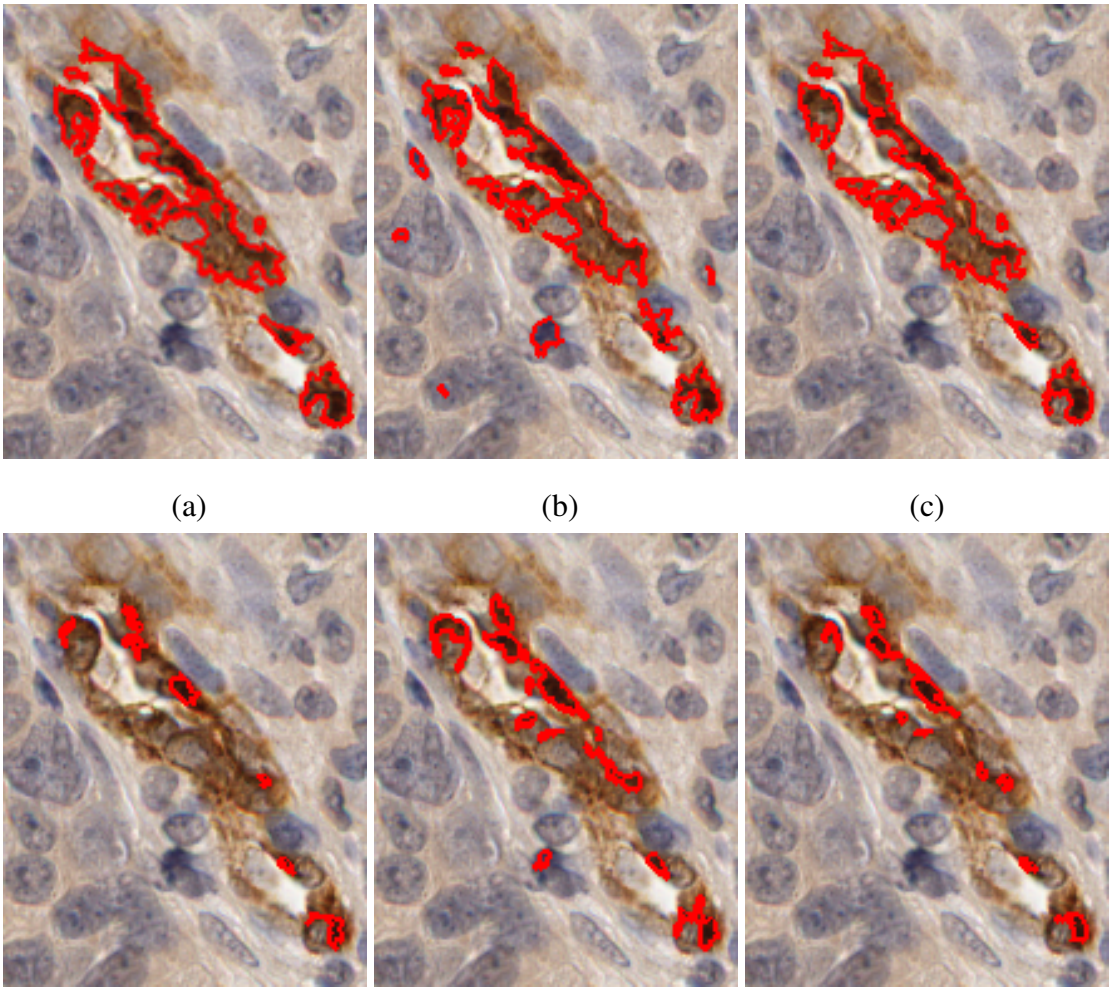
(a) (b) (c) (d)

Figure 4.7: (a) Ground truth annotation of stain extent obtained from an expert pathologist. The segmentation result shown in (b) was created using a swatch comprising 7 selected pixels. The next column (c) contains the same values as (b) with the addition of another 5 values. The final column (d) has 18 values selected from the original image. The red line encapsulates the results of the segmentation algorithm. We can see that the first set of results (b) are reasonable, but as more class representative samples are used to construct the swatch, the results improve ((c), (d)). Red boundary delineates perimeter of selected region for clear viewing.

4.7(d) was used (a true representation of the variance in the target class) the results approached the annotations of the expert. Note that a non-expert user could easily determine which areas of the target class were not sampled from, and include those in the s watch. This iterative process could be repeated until the non-expert user observes the results that match the exact desired output. Once the domain swatch is selected, it can safely be used for the rest of the images in the TMA set.

Parameter sensitivity

σ_{MS} is the parameter, used in Gaussian function of the FWMS, which determines the variance of the modes and is thus dependent upon the dataset considered. In Figure 4.8, the importance of selecting the correct σ_{MS} becomes apparent. In the case where the σ_{MS} value is too large, the FWMS aggregates together pixels not contained within the swatch. As a result, they can never be pruned away as shown in (b). The highlighted blue section is dark enough in color that it becomes associated with the stain due to the large bandwidth selection. On the other hand,



(a) (b) (c)

(d) (e) (f)

Figure 4.8: (a) Ground truth (pathologist) segmentation of stain extent, (b), (c) above show segmentation outputs for two different σ_{MS} values ($\sigma_{\mathrm{MS}} = .01, .3$). The algorithm rarely experiences unacceptable segmentations except in the case when an intentionally inappropriate value of σ_{MS} for the domain swatch is chosen. Figures 4.8 (d), (e), (f) are illustrated with σ_{MS} values of .01, .3, and .05 respectively, except that for these cases, a non-representative swatch for the target class was deliberately selected.

when the appropriate swatch representative of the desired target class is selected, almost any σ_{MS} value becomes acceptable, as shown with $\sigma_{\mathrm{MS}} = .01$ in Figure 4.8(c). Unfortunately, in the case where a swatch that is not representative of the target class is selected, as in Figures 4.8(d), (e) and (f), the results tend to be more sensitive to the choice of value for σ_{MS}.

In our specific application, using HNCut on 500 discs, about 10 of them failed to converge properly (as determined by qualitative, visual inspection), resulting in very poor segmentations. Interestingly, these 10 images all had little to no stain present. By computing the variance of

the color pixels in the segmented output against the domain swatch, we can assess the performance of HNCut and make relevant adjustments in an unsupervised manner. For instance, if the variance is larger than desired, adjusting σ_{MS} to a smaller value will produce new output that is more similar to the domain swatch. For all 10 images considered in this experiment, the scheme for automatically adjusting σ_{MS} resulted in excellent results.

4.7 Experiment 3: Efficiency and Speed Considerations of HNCut

A crucially important property of HNCut is the efficiency of FWMS compared to the traditional MS. To quantitatively evaluate the computational savings in using FWMS compared to MS, the MS and FWMS procedures were executed over a total of 20 iterations and the corresponding iteration times graphed. Additionally, we compared the time it took for PBT, k-means, and HNCut to segment the 130 tissue cylinders for which quantitative evaluation was performed.

In order to clearly illustrate the high-throughput capabilities of HNCut, we compared its runtime to PBT and k-means. Figure 4.5 illustrates a graphical representation of the results. From the onset we can see that PBT's training time of 181 seconds accounts for 25% of HNCuts 643 second run time. Typically this training time is divided amongst all of the tested samples; thus the more samples that are tested, the cheaper it becomes to train the system. Regardless, even upon excluding the training time for PBT, HNCut still performs significantly faster. The average of 16 seconds per sample by PBT is easily beaten by the runtime of 6 seconds per sample by HNCut (for each 1500×1500 cylinder on the TMA). This implies that HNCut is roughly 62% faster compared to PBT.

In Table 4.2 we can see the expected time taken to perform the classification task on images of different sizes. For larger images, the difference in execution time becomes even more apparent. When we compare the time needed for HNCut versus that of a human expert performing the same task, the need for a technological approach becomes apparent.

Figure 4.9 shows the numerical advantages to using FWMS over MS. When the initial number of points is large, after each iteration, fewer computations need to be performed. The larger ϵ is selected, the faster FWMS will converge, on the other hand, when ϵ is selected to be extremely small the execution time for FWMS begins to approach that of MS.

Image Dimensions

(Number of pixels)

	323x323 (104,329)	646 x 646 (417,316)	1292 x 1292 (16,69,264)	2584 x 2584 (96,677,056)
HNcut	0.6s	1.2s	7.2s	25s
K-means	6.9s	30s	104s	504s
PBT	0.678s	2.9s	12.5s	46s
Ncut	43s	585s	Insufficient Memory	Insufficient Memory
Est. Manual Segmentation	30s	160s	600s	2600s

Table 4.2: Run times for segmentation of various sized images. We can see in all cases the HNCut algorithm provides the best run times. Additionally, there are two cases in which NCut is unable to finish because it exceeds the maximum amount of memory, a strong limitation for large scale usage. It also becomes apparent that using an algorithm, as opposed to manual segmentation, is certainly a more efficient process. The mentioned timings were performed using a 2GHz dual-core laptop having 8GB of RAM.

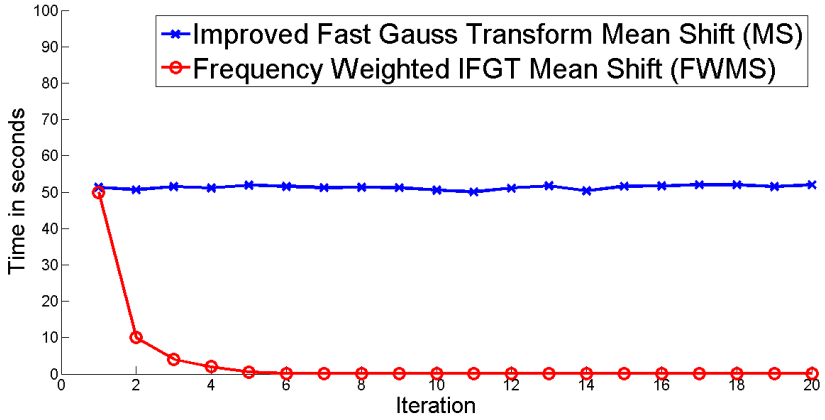


Figure 4.9: A graph showing the typical computation time in seconds for each iteration of the MS and FWMS procedures. The original Improved Fast Gauss Transform (MS) Mean shift (top, in blue) has constant time for each iteration. The benefits of the Frequency Weighted Mean Shift (FWMS) algorithm (bottom, in red) become apparent within a few iterations of the clustering procedure as each additional iteration requires significantly less time as additional data points converge to the cluster mean.

4.8 Experiment 4: Comparing a Supervised Classifier driven by expert annotations versus HNcut

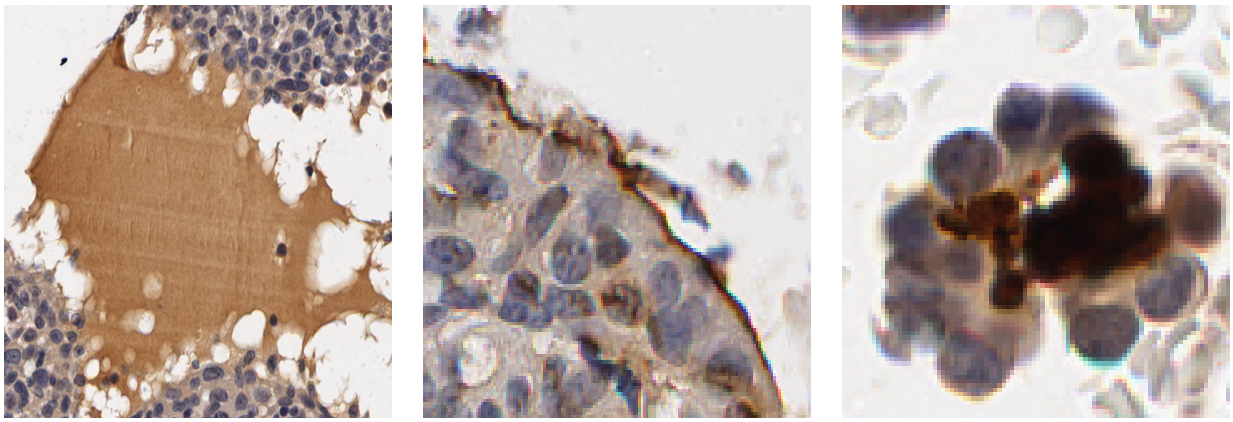
Since the production of the ground truth datasets by experts for training is laborious, we pose the question: is it possible to differentiate a supervised classifier trained with expert human annotated data from a supervised classifier trained with HNCut segmented data? In general, supervised methods are viewed as more dependable because they rely on training data. However, the question we pose is whether it is possible to use a minimally-supervised method to train a supervised method and obtain on par or better results to a supervised classifier trained with manually annotated data. Towards this end, we performed 10 iterations of the training/testing procedure using the HNCut output as the ground truth for training in the PBT, and compared it against the PBT output resulting from the pathologist annotated data. The choice of thresholds was determined in a similar fashion as Experiment 1, except the operating point was found to range between 97% and 99%, and thus we chose those two values.

The results presented in Figure 4.5 (using the "PBT W. HNcut" label) suggest that when a PBT is trained with the results from the HNCut, the results are actually superior to all other classifier configurations considered (including PBT, k-means, and HNCut), with a much smaller standard deviation. In the case of FP, the variance at the 99% threshold is almost negligible, giv-

ing a high confidence of reproducibility. As a result, the output suggests that it is possible to use HNCuts layman's initialization to produce data that is of a similar quality to the expert's laborious annotation work, minimizing user interaction. This is especially interesting because it means that the combination of the two outperforms a supervised method trained with expert data. This result suggests that supervised classifier methods can be employed for accurate quantification of biomarker extent by using HNCut to create the training set. This would be highly beneficial, avoiding the extremely expensive overhead of laboriously and manually annotating the target class. Based on these results, HNCut would appear to produce results which are on-par with an expert for usage as training data in a supervised classifier.

4.9 Discussion of Segmentation Errors

As with any segmentation algorithm, HNCut is also subject to FP and FN errors. Below, we briefly discuss some of these errors and possible reasons for these errors.



(a) (b) (c)

Figure 4.10: Typical reasons for false positive and false negative errors. Stain tends to fill the (a) void where tissue is absent causing a re-active presence. The (b) rim of spots tend to stain darkly, these are easily ignored by adding a distance from border threshold. Psamommas (c) are calcifications which absorb stain and thus appear similar to target staining, the specific nature of the biological anomaly makes it difficult to classify it correctly using only chromatic information.

Since the stain severity is proportional to the quantity of the biomarker, the stain will vary greatly in intensity of color across not only all cylinders but also across all stained areas themselves (Figure 4.10(a)). This high variance is one of the reasons why thresholding and k-

means type algorithms tend to do poorly. Additionally, the rims of the cylinders (Figure 4.10(b)) are often corrupted with noise which manifests as a dark stain. The removal of these artifacts could be done by simply choosing to ignore pixels that lie on or very close to the cylinder boundary. In the situation where the disc is not well formed, either on account of tissue tearing or an absence of cells, there is the possibility for large scale pooling of FP stain within the void. Since the chromatic qualities of the FP regions are very similar to true positive areas, this specific type of error is difficult to identify and eliminate.

Psammomas (Figure 4.10(c)) are calcified material within the center of a laminated whorl of elongated fibroblastic cells [61]. Unfortunately, psammomas are exactly the same in color and texture as the true positives, making it difficult for all save an expert reader to identify. In the absence of additional domain knowledge, it would be impossible for any segmentation algorithm (let alone HNCut) to distinguish these FP errors from the true positives.

Chapter 5

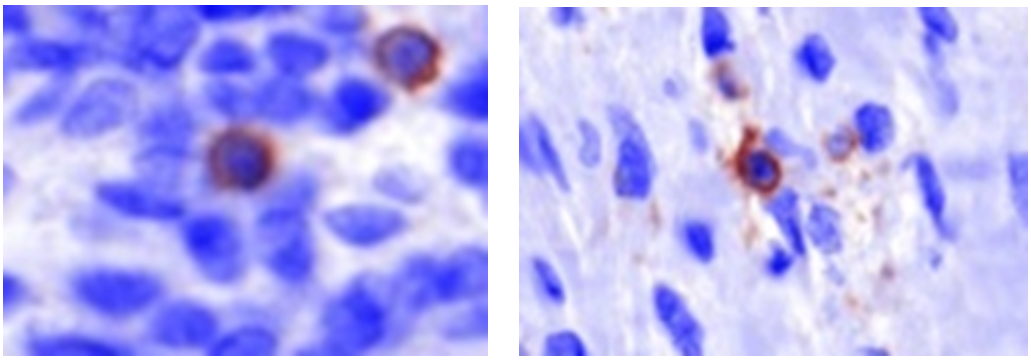
Local Morphologic Scale (LMS): Theory

5.1 Introduction to Region Classification

Another common task in the field of digital pathology is the classification of a region as either tumoral or stromal. An important application of the said task is the identification of tumor infiltrating lymphocytes (TILs) versus non-TILs. Recent work [8, 9, 10, 11, 12, 13, 14, 15, 16, 17] has suggested that a valuable prognostic indicator is based on the extent to which the patient's own immune response has detected the cancer. This response can be characterized by the behavior of lymphocytes. A lymphocyte is a type of white blood cell that is sent to the proximity of objects which the body considers foreign, the more lymphocytes the greater the perceived risk. In this case, the object of interest is a tumor, thus creating the dichotomy of lymphocytes into tumor infiltrating lymphocytes (TIL) and a non-TIL.

Lymphocytes are fairly easily identified and segmented using a combination of targeted staining and advanced image analysis techniques. Unfortunately, lymphocytes of both classes appear morphologically similar, as can be seen in Figure 5.1. Thus, after segmentation, the challenge of classification of the lymphocytes into the aforementioned two groups, TILs and non-TILs, arises. Since they visually appear similar, it is impossible to determine the class of a lymphocyte by merely examining the lymphocyte itself and thus a more global approach is needed in order to accurately determine their embedding. This placement falls into two parts: either the lymphocyte is in the tumor or the lymphocyte is in the stroma (see Figure 5.3). This creates the dual problem of detecting stroma versus tumor regions.

To summarize: (a) lymphocytes are the body's natural defense mechanism against foreign bodies, (b) the number of lymphocytes present in a particular region is proportional to the per-



(a) (b)

Figure 5.1: Under high magnification the TIL in (a), identified by the red stain in the center, appears morphologically similar to the red non-TIL in (b). We know that internally the objects are the same, but their classification is dependent upon their neighborhood, i.e., tumor for TILs and stroma for non-TILs.

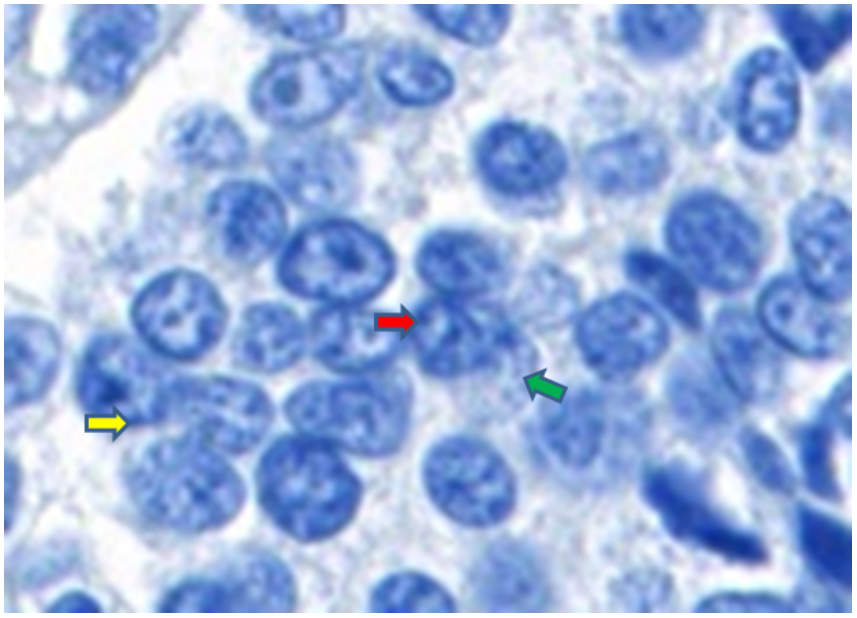


Figure 5.2: Since histology images are 2D representations of a 3D structure, overlap artifacts are inevitable. The red arrow identifies an in-focus cell which occludes an out of focus cell(green arrow). The yellow arrow is focused on a region which exhibits cell clumping. Both situations make identifying cellular boundaries challenging.

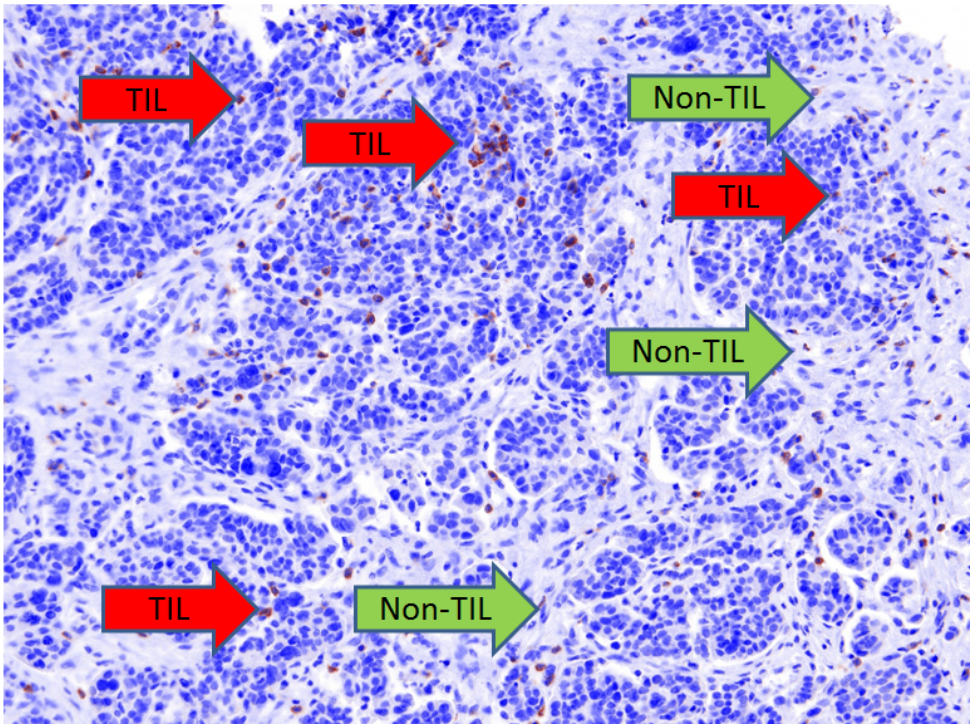


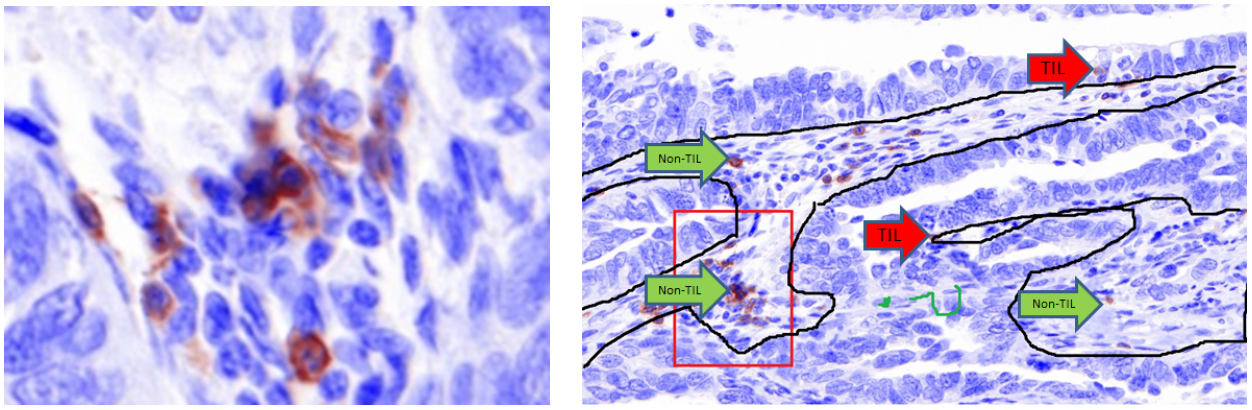
Figure 5.3: The defining features of the TILs identified by red arrows are (a) surrounded by larger more circular tumor cells, (b) tumor cells tend to be a bit more hollow and (c) contain dark nuclei. The non-TILs identified by green arrows also have their own properties. They tend to be (a) in a more sparse region, (b) not embedded in tumor cells and (c) surrounded by spindle shaped endothelial cells

ceived danger and (c) if there are a large number of lymphocytes present in a tumor, indicating a strong immune response, there is potentially a better prognosis for the patient. Again, as in Section 1.1.1, to accurately test this hypothesis, and further to successfully model the exact quantification of lymphocyte configuration, the challenge of data size arises. In order to make any definitive conclusions, a vast number of patients, and thus samples, must be analyzed. As expected, the cost and time associated with an expert performing these tasks is not only insurmountable, but duplicated upon each inquest per new property analyzed. As such, off-loading such laborious and repetitive tasks to computers provides significantly improved efficiency and reproducibility.

5.2 Challenges and Contributions

In the context of biological images [62] (such as in microscopy applications or histopathology imagery), the objective is often to identify local regions of heterogeneity (e.g., cancer nuclei, lymphocytes), whereas larger homogeneous regions (e.g., benign stroma) may be less interesting or informative from a diagnostic or prognostic perspective [63, 64]. Additionally, the shape of the physical manifestation of this local heterogeneity may be highly predictive of a pathologic process (e.g., architectural arrangement of nuclei and glands in prostate cancer reflects the Gleason grade and hence aggressiveness of the disease [65]). Hence for images where the most interesting information is encoded in the local heterogeneity, and where the objective is to spatially assign distinctive quantitative scale signatures to characterize and classify these regions, it would appear that a new local scale definition is warranted.

The major contribution of this chapter of the thesis is the presentation of the concept of a novel *scale space*, as described in Chapter 2.2, Local Morphologic Scale (LMS) which model local heterogeneity and associate a quantitative local morphologic signature with every spatial image location. Since lymphocytes in both the stromal and tumor regions appear identical, identifying TILs requires identifying the kind of tissue that the lymphocyte is embedded in. In other words, discriminating TILs from non-TILs has to do with quantifying the appearance of the local neighborhood within which a lymphocyte is present. The difference in the LMS signatures for the lymphocytes in Figures 2.1 (i), (j) and Figures 2.1 (k), (l), respectively could be exploited to distinguish all TILs and non-TILs in the image, a laborious task for a human to perform manually. As such, our objective is to train a supervised classifier to use LMS



(a) (b)

Figure 5.4: The highly disorganized nature of the cells in (a), a high magnification field of view, makes it seem like a cancerous region. When looking at a (b) bigger field of view, with (a) indicated by a red box, we see that the area is actually non-cancerous and simply appears cancerous as a result of a biopsy artifact.

signatures to accurately discriminate between stroma and non-stroma regions in OCa histology images.

5.2.1 Challenges

There are numerous challenges, especially in the ovarian cancer domain, which inhibit the classification of lymphocytes as TILs or non-TILs. First, we can see in Figure 5.2 that the segmentation of individual cells would be difficult since they clump together as a result of a 3-dimensional tissue sample being scanned in two dimensions. This artifact often leads to cellular boundaries which are occluded or are ill-defined as they lay on top of other cells (Figure 5.2 yellow arrow). A great difficulty then arises in attempting to split these cells in order to define clear boundaries required for domain specific feature extraction approaches. Needless to say, these algorithms are often computationally expensive and complex [66, 67].

The selection of an appropriate window size and shape for typical algorithms is also notably challenging. We can see from the varying optimal window shapes and sizes shown in Figure 5.5 that the systematic determination of these inconsistent regions of interest, used in algorithms such as texture features, is challenging. A simple square, as would normally be used, is not an appropriate fit to the displayed diversity. This great variance in regions of interest leads to a breakdown of standard approaches as an incorrect selection fails to encompass

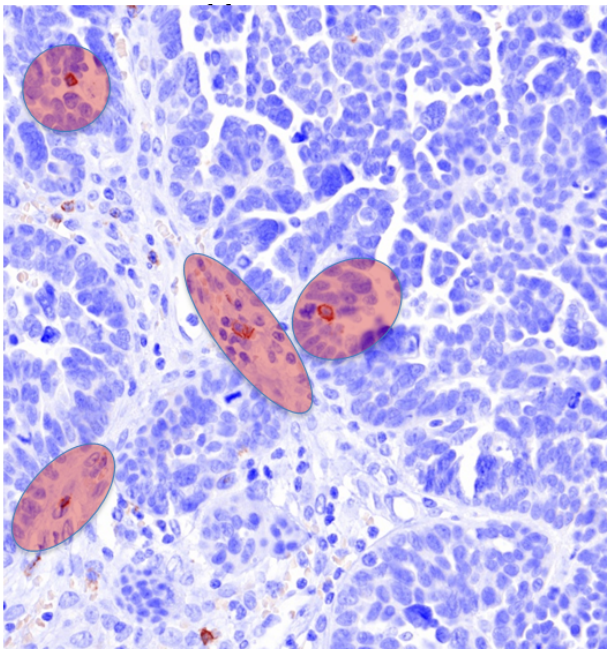
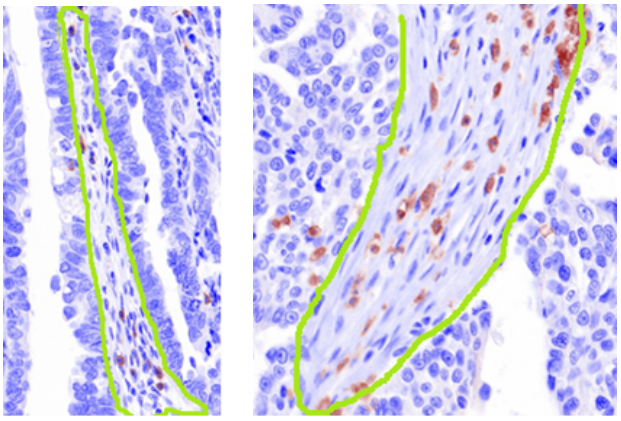


Figure 5.5: Optimal ellipses overlaid in red for their associated lymphocytes. Although the ellipse is a rather simplistic shape, we can see the large complexity in variations among 4 examples in a single image.



(a) (b)

Figure 5.6: Stroma region circled in green. Lymphocytes stained in red. Notice the stark difference in sizes between the two green regions in (a) and (b). Consistently selecting an appropriate region of interest for typical image techniques would be challenging. In (b) we can see some lymphocytes inside the green delineated region, making them non-TILs, while their TIL counter parts lay outside.

all the necessary information, and perhaps even incorrect information, leading to an incorrect determination of the lymphocytes TIL or non-TIL status.

Lastly, to further complicate things, the stroma region is often nestled between areas of tumor, making not only its boundaries not clearly defined but the size of the associated region even more difficult to pre-determine as shown in Figure 5.6 and Figure 5.4(b). This is to say that a lymphocyte must be firmly embedded in the tumor (i.e., surrounded by tumor cells) in order for it to be a lymphocyte. When sitting on the boundary it becomes notably more difficult to ascertain which side of the boundary the lymphocyte is sitting on, thus putting its TIL or non-TIL status into question, hence requiring an *orientation sensitive* algorithm to make the correct determination.

5.2.2 Contribution

In this chapter we present a new definition of local morphologic scale (LMS), which is appropriate for images with high degrees of local complexity, where existing local scale definitions governed by satisfying homogeneity criterion break down. Our innovative approach models the heterogeneity of a local region, allowing for the definition of local, quantitative signatures of heterogeneity. Since LMS motivates the definition of local regions as regions which are topographically similar, pixel level features can be defined from the corresponding LMS at that location, features which can then be used for segmentation, registration, or classification. By converting these signatures to feature vectors by using Fourier descriptors [68], we obtain the valuable properties of scale, translational and rotational invariance.

The novel framework we present in this thesis is highlighted by the following important contributions:

- A novel morphological signature definition that allows for quantitatively characterizing local heterogeneity, unlike other local scale definitions focused on capturing local homogeneity.
- The LMS yields a rotationally invariant quantitative signature at the pixel level which can be used for region classification, segmentation, and registration, especially relevant in the context of highly heterogeneous images such as in histopathology.
- This signature generalizes to higher order class labels and is accurate across a range of window sizes, overcoming common downfalls of texture and template matching based

classifiers.

- We develop a novel approach to the important problem of separating out tumoral from stromal regions via application of LMS which we show to be computationally efficient and scalable to the large number of repositories in existence.

5.3 Overview

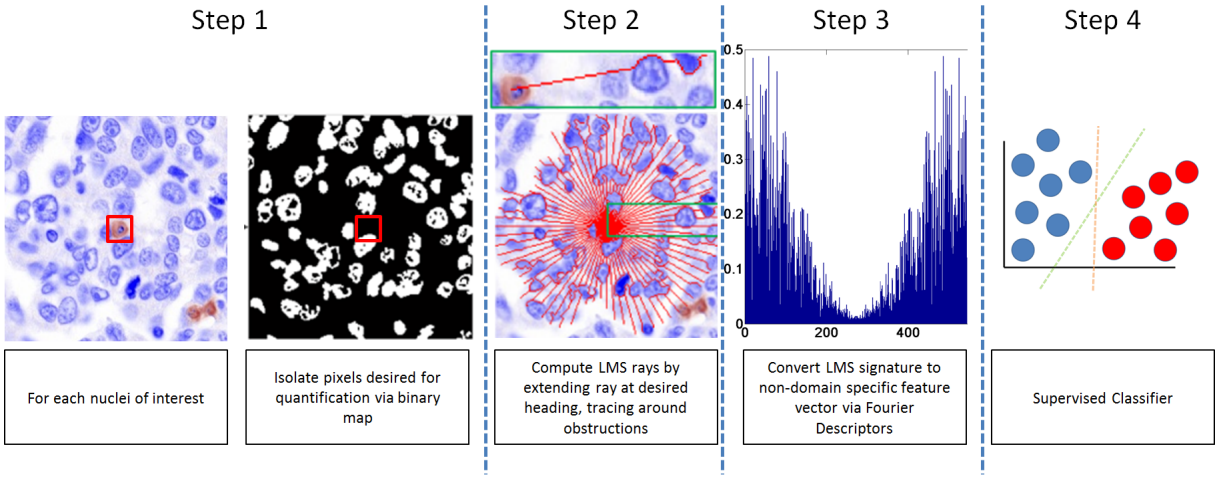


Figure 5.7: Overview of the LMS signature creation process. In Step 1 we create a binary map using HNCut which indicates which pixels will be used to define local morphology. Step 2 produces the LMS rays by extending a ray outwards from the point of interest and circumventing any obstructions in its path. We quantify these rays using Fourier Descriptors in Step 3 and lastly train a supervised classifier to differentiate between the two classes using our feature vectors.

Figure 5.7 presents an overview of the LMS creation process, with Figure 5.8 demonstrating a TIL and non-TIL undergoing said process for comparison. These steps are described below.

Step 1: Identify lymphocytes using a *red* swatch and HNCut, a hybrid mean shift and normalized cuts algorithm [69]. Next, perform HNCut using a *blue* swatch to produce the resulting binarized images (as seen in column B of Figure 5.8) and in Figure 5.9. Binarized images indicate which pixels will be incorporated in the morphologic signature of the point of interest (POI).

Step 2: Calculating LMS involves projecting connected paths, radially outwards from the POI (nucleus center identified in step 1). Column C in Figure 5.8 shows the LMS signature (in

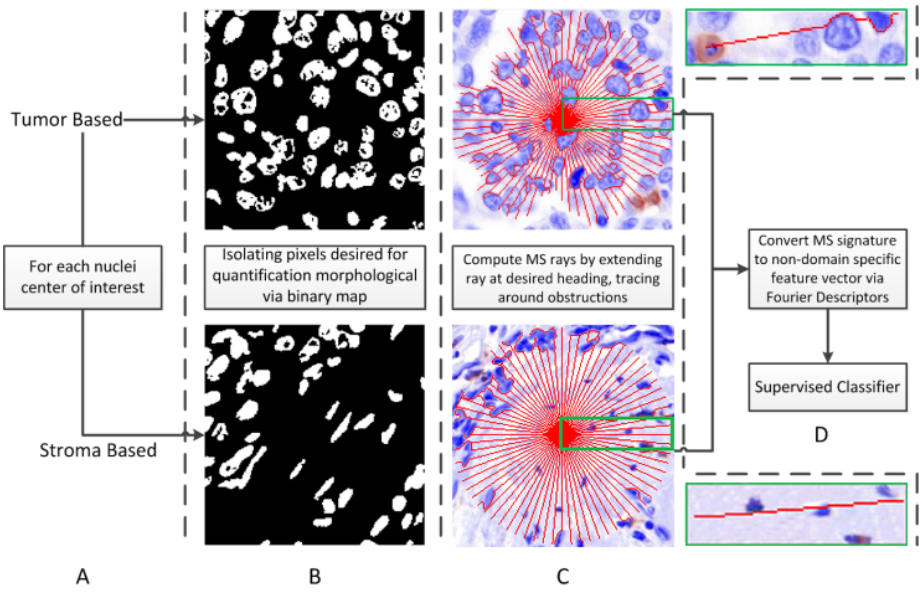


Figure 5.8: Overview of the LMS signature creation process as it applies to the two classes of interest. From column B we can directly see that the organization and properties of the cells as captured by the binarized map are indeed visibly different. Next we can see that the tumor LMS signature contains a noticeably increased number of deviations from the straight line trajectory, on account of the rays attempting to take the path of least resistance and hence overcome obstacles along the way. In its associated green box, the ray is forming a larger circular path indicative of encountering a tumor cell. On the other hand, the LMS signature for the non-tumor region is much smoother as a result of comprising fewer and smaller objects. In its green box, we can see that the obstructions are shorter and more spindle like, the classical definition of endothelial cells residing in a stroma region.

red) for a TIL (top) and a non-TIL (bottom). In each image, a green box is used to illustrate the path of a single ray more clearly. It can be seen that the different classes produce noticeably different path characteristics. Morphologic features that characterize the local topography of the binarized image via individual rays/paths should be able to quantify these characteristics to provide class labeling for unseen samples.

Step 3: The quantification of the local topography of all these paths (via Fourier descriptors [68]) yields a measure of local heterogeneity.

Step 4: Use the LMS signature vector created by the Fourier Descriptors to train a supervised classifier to identify signatures as either located in tumor or stromal regions.

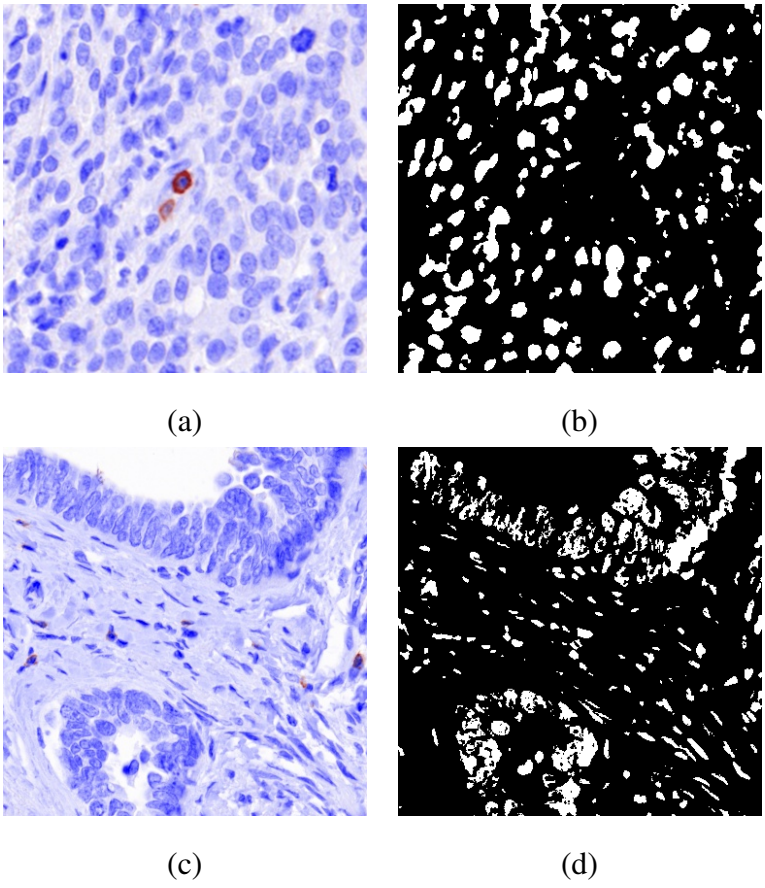


Figure 5.9: A tumor region (a) and non-tumor region (c) with their associated binary masks ((b) and (d), respectively), as produced by using a blue swatch with HNCut. Note that the red lymphocytes are absent.

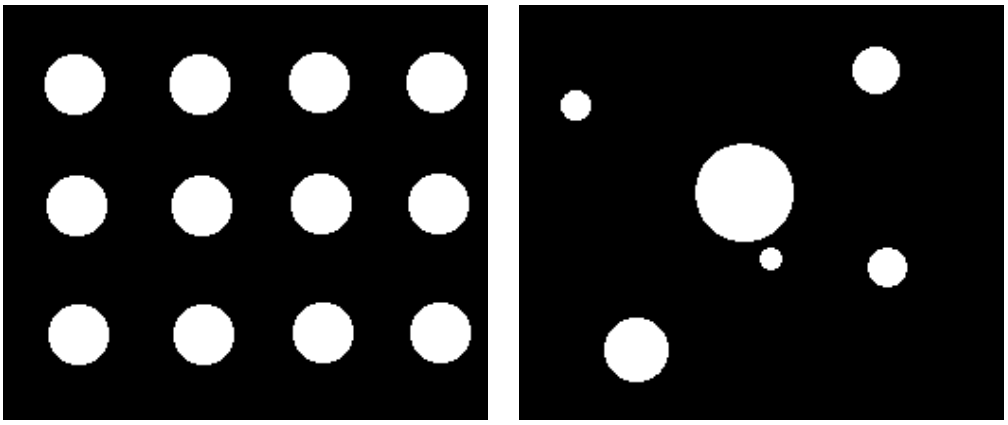
5.4 Theory and Algorithm

5.4.1 Notation

An image scene is defined as $\mathcal{C} = (C, \mathbf{g})$ where C is a 2D Cartesian grid of N pixels, $c \in C$, where $c = (x, y)$. $\mathbf{g}$ is a Boolean function ($g(c) \in \{0, 1\}$), where $g(c) = 1$ indicates a foreground pixel, identified in our implementation by using a hybrid mean shift and normalized cuts algorithm, contributing to the local morphological signature. The notation $< c^{(1)}, \ldots, c^{(m)} >$ denotes an ordered set of m pixels, with $|\circ|$ denoting the number of elements in a set. $||\circ||_p$ denotes the standard p-norm. Lastly $\mathcal{N}_4(c) = \{(x + xx, y + yy)|xx \in \{-1, 1\}, yy \in \{-1, 1\}\}$ and $\mathcal{N}_8(c) = \{(x + xx, y + yy)|xx \in \{-1, 0, 1\}, yy \in \{-1, 0, 1\}\}$, or more commonly known as the 4- and 8-neighborhood, respectively, around c.

5.4.2 LMS Signature

Theory of Signature Generation



(a) (b)

Figure 5.10: Image (a) has consistent spacing and sizes for the white circles, implying a very low amount of entropy, and thus indicating a homogeneous structuring. We can contrast this with a heterogeneous image in (b) which is associated with a larger value of entropy, due to the large variance in sizes and relative spacing.

Qualitative *homogeneity* (and thus *heterogeneity*) definitions can be quantified by low (or high) entropy. Entropy, in this case, is defined classically in information theory literature [70] as a measure of the uncertainty associated with a random variable. Without loss of generality, if we consider only size and relative distance of objects as the random variables in our scene we can get a feel for these qualitative definitions. From the simple example constructed in Figure 5.10, we can appreciate visually the concept that the more homogeneous a region is, as shown in Figure 5.10(a), the less the observable object variables (such as size or proximity to other objects) differ. On the other hand, the more heterogeneous a local structuring of objects is, as shown in Figure 5.10(b), the greater the variables will vary resulting in higher entropy.

Going forward, we aim to justify that the LMS rays are linked to the random variables associated with the local entropy of the point of interest.

Definition A *path* $p_{r,s} = \{< c^{(1)} = r, c^{(2)}, \ldots, c^{(m-1)}, c^{(m)} = s >: ||c^{(i)} - c^{(i+1)}||_1 = 1, i \in \{1, \ldots, m\}\}$ is a connected set of pixels which starts at r and ends at s.

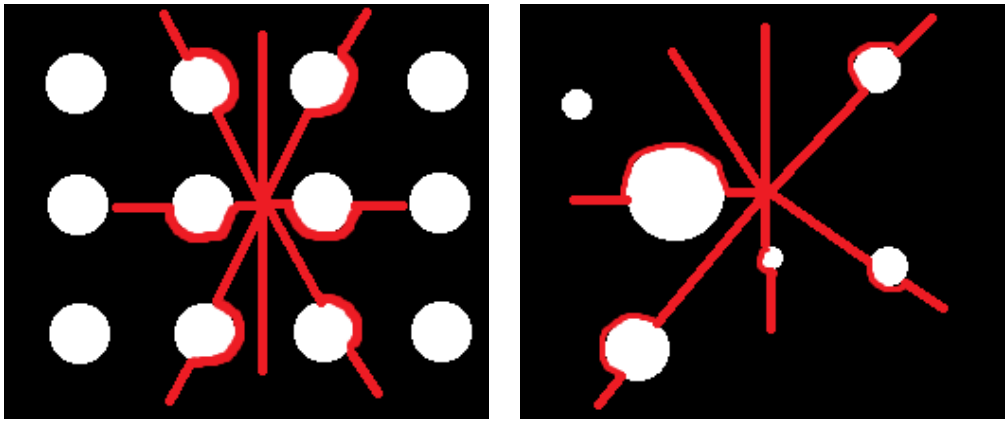
Definition A *μ-path* is a $p_{r,s}$ path, denoted as $\dot{p}_{r,s}$ such that the affinity constraint function $\mu(c^{(i)}, c^{(i+1)}) \equiv g(c^{(i)}) + g(c^{(i+1)}) = 0$, is met for $i \in \{1, \ldots, |\dot{p}_{r,s}|\}$, implying sequential

pixels are both background pixels.

Definition A *minimal* μ*-path* $\hat{p}_{r,s}$ is a specific $\dot{p}_{r,s}$ which is defined as $\hat{p}_{r,s} = \min_{|\dot{p}_{r,s}|} \dot{p}_{r,s}$, intuitively making it a minimal μ-path from r to s such that the pixels are not only connected, forming a path, but are also all background pixels.

Definition The *LMS path* $R_{w,\theta}(q)$ is $\hat{p}_{q,\delta}$ with q as the point of interest and δ positioned at a distance w with angle θ. To simplify notation, we refer to it as $R_\theta(q)$ since w is user defined and thus held constant.

Definition The *unobstructed path* $O_\theta(q)$ is similarly defined as $R_\theta(q)$, except it has no affinity constraint, i.e., $\min_{|p_{q,\delta}|} p_{q,\delta}$, making it the straight set of connected pixels from q to δ.



(a) (b)

Figure 5.11: Revisiting the previous figure with possible LMS signatures overlaid on images in red. In image (a) we can see that the rays have consistently interacted with the obstructions providing curves of similar amplitude. On the other hand, with the heterogeneous image in (b), we can see that the rays are notably different in amplitude and periodicity leading to a state of greater entropy.

Proposition $R_\theta(q)$ models local morphology, and thus an entropy surrogate, at point q in direction θ via deviations from $O_\theta(q)$. As the affinity function μ constrains the LMS path to background pixels, we can expect $R_\theta(q)$ to become more tortuous when the objects it encounters require modeling additional entropy in the local neighborhood, resulting in a direct correlation between the pixels selected for $R_\theta(q)$ and its associated heterogeneity.

- We can see that if the region contains no obstructions, $R_\theta(q) = O_\theta(q)$, indicating the most trivial case of homogeneity
- If $R_\theta(q)$ is computed in an image such as Figure 5.11(a), we would expect to see curves

of similar amplitude at equally spaced intervals. These curves are indicative of the size, shape and orientations of objects in the rays path; thus the curves encode the homogeneity of those implicit variables.

- If $R_\theta(q)$ is computed in an image such as Figure 5.11(b), we would expect to see curves of dissimilar amplitudes as the ray must circumvent circles of various sizes. These amplitudes would be unevenly spaced as the circles are not uniformly placed, as a result modeling the heterogeneity of these two variables.

- Additionally, since we are not defining explicitly the domain specific attributes (in our example of shape and size above), the ray is also subject to varying shapes, concavities, etc.

Definition The *LMS signature* of a query pixel q is then $R(q) = \{R_0(q), \ldots, R_{2\pi}(q)\}$, or the set of rays in the desired sampling directions, giving a sampled view of the surrounding regions.

Morphologic Scale Computation

We present Algorithm 3 for computing the discrete LMS signature $R(q)$ for a query pixel $q \in C$. R_q is defined by a series of sampled paths at ϵ interval from the query point q outward, sampling the morphology of the surrounding region.

Algorithm 3 LMS Signature Creation

Input: A query pixel $q \in C$, binary function g, interval size ϵ, window size w

Output: $R(q)$

$S = C_{-q_x,-q_y}$, a transformation such that q is located at the origin
for $\theta = 0 : \epsilon : 2\pi$ **do**
  $\delta = (\cos(\theta) * w, \sin(\theta) * w)$
  if $\nexists \hat{p}_{q,\delta}$ **then**
    $\delta = \underset{d \in C, g(d)=0}{\operatorname{argmin}} ||d - \delta||_2$
  end if
  $R_\theta(q) = \hat{p}_{q,\delta} = \min_{|\dot{p}_{q,\delta}|} \dot{p}_{q,\delta}$
end for
$R(q) = \{R_\theta | \forall \theta\}$
return $R(q)$

We can see from the examples in Figure 5.8, the end result of our algorithm is set of connected pixels (shown in red), which travel from the POI q to δ. The algorithm proceeds by translating the image such that q is placed at the origin. For each of the R_θ , we determine the location of the end point δ by casting it on a unit circle and multiplying by the window size to get the appropriate magnitude, as shown in step 3 of Algorithm 3. There is often the case where the desired end point δ is not a background pixel ($g(\delta) = 1$) and thus we assign δ to the closest possible pixel, in the Euclidian sense, which has the required property ($g(\delta) = 0$). Afterwards, we identify the minimal μ-path, $\hat{p}_{q,\delta}$, which is intuitively $O_\theta(q)$ with minimal divergence to circumvent obstacles. This is to say when the path hits an object, the resulting affinity function threshold criterion is exceeded and hence the path continues along a new direction of lower resistance (satisfying the affinity criterion). We apply these steps multiple times, each time moving δ at an interval of ϵ radians, and produce $R(q)$ the set of individual rays.

Figure 5.13 presents the LMS signature in red over a TIL and non-TIL. In both images, we can see that as the complexity of the local region increases a noted change in the LMS occurs, which is to say as the entropy of the neighborhood structures rises, the rays become more chaotic. We can see in Figure 5.13(b) homogeneous regions have few obstructions, and thus the LMS paths form straighter lines. In Figure 5.13(a), we can see for a very complex region the LMS signature becoming increasingly tortuous as it adapts to the local heterogeneity. By quantifying this notable change, we are able to train a supervised classifier to differentiate between the two classes.

LMS Algorithmic Implementation

The implementation of the LMS signature computation comes with a few points of note. First, to ensure that the origin point meets the criteria that it is a background pixel, we apply a circular mask around q of a user specified size forcing $g(c)$ to return 0. Next, computing the globally optimal $\hat{p}_{q,\delta}$ (i.e., such that we are guaranteed the shortest past), is notably computationally expensive. This expense is as a result of need to use algorithms such as Dijkstra [71] or a Fast Marching [72] approach to compute the global minimum length path. On the other hand, we can sacrifice some precision, and obtain a "minimal path" but not the *shortest* path, and benefit from orders of magnitude improvement in efficiency. To do so, we present Algorithm 4, which is an iterative greedy approach towards solving for $\tilde{\hat{p}}_{q,\delta}$, a sampled approximation of $\hat{p}_{q,\delta}$.

The algorithm proceeds by defining the first pixel in the path ($c^{(0)}$) as the query pixel (q),

Algorithm 4 Compute Sampled Minimal Path

Input: A query pixel q , end pixel δ, valid pixel indicating function f_θ, maximum iterations M

Output: $\tilde{\hat{p}}_{q,\delta}$

1: $m = 0$

2: $c^{(m)} = q$

3: **while** $c^{(m)} \neq \delta$ and $m < M$ **do**

4: $\quad m++$

5: $\quad c^{(m)} = \underset{t}{\operatorname{argmin}} ||t - \delta||_2$, subject to $t \in \mathcal{N}_8(c^{(m-1)}), f(t) = 1, t \notin \{c^{(0)}, \ldots, c^{(m-1)}\}$

6: **end while**

7: **return** $\tilde{\hat{p}}_{q,\delta} =< c^{(0)}, \ldots, c^{(m)} >$

then it iteratively selects the next pixel in the path by determining, using Euclidian distance, which of the possible pixels in the 8-neighborhood is closest to its goal. If there are two pixels which meet the same criteria in the path, we sample from them with equal probability. The key to this approach is the valid pixel indication function $f_\theta(c)$, which is specific to the θ being considered. The value of $f_\theta(c)$ is defined by the logical equation

$$f_\theta(c) = (c \in O_\theta(c) \vee \Sigma_{d \in \mathcal{N}_4(c)} g(d) > 0) \wedge \neg g(c). \tag{5.1}$$

Intuitively, this function returns a true value for pixels which are on the edge of the objects or if the pixel is on the unobstructed path $O_\theta(q)$, causing the path to be constrained to objected borders or the line from q to δ.

We also note that the selection of ϵ in Algorithm 3 acts as a Monte-Carlo type parameter, allowing for the selection of appropriate sampling in proportion to the total computation time. Additionally, not only is the algorithm computationally straight forward (i.e., requiring only the most basic of arithmetic or logic operations), it can be seen that each ray is computed in a deterministic manner not dependent upon its peer rays. These two properties are beneficial as it allows for parallel computation of the individual rays on GPU technology resulting in calculation of rays en masse.

5.4.3 Fourier Descriptors of LMS

Fourier descriptors (FD) [68] are a technique for quantifying the morphologic structure of a closed curve as a feature vector which is scale, translationally and rotationally invariant. These

properties are a necessity in domains such as biomedical image analysis where information is represented in an orientation-free plane. We use a slightly modified version, as the scale invariance is not important to our domain as all samples are drawn from the same magnification. Unfortunately, at first glance, $R(q)$ is not a closed curve and thus these techniques cannot be directly applied. In the following section we define a theoretical foundation for conversion of the open set of $R(q)$ to a closed curve $J(q)$.

Theory of Conversion from $R(q)$ to a Closed Curve

Definition *Return path* $K_\theta(q)$ is a minimal path from δ returning to q, essentially $O_\theta(q)$ reversed.
Definition *Returned path segment* $J_\theta(q)$ is $< R_\theta(q), K_\theta(q) >$, i.e., the concatenation of $R_\theta(q)$ and $K_\theta(q)$, forming a path which starts at q and forms the shortest constrained path to δ and then returns in a straight line to q forming a closed curve.
Definition *LMS closed signature* $J(q) =< J_0(q), \ldots, J_{2\pi}(q) >$
Proposition $J(q)$ forms a closed curve. By induction, consider the path $J_0(q) \equiv< R_0(q), K_0(q) >$ = $< c^{(1)} = q, \ldots, c^{(m)} = q >$. We can see that it both begins and ends at the same point, q, thus forming a local closed contour. Now we take $< R_1(q), K_1(q) >$ = $< e^{(1)} = q, \ldots, e^{(m)} = q >$ we can see $c^{(m)} = e^{(1)}$ and $c^{(1)} = e^{(m)}$, thus each $< R, K >$ combination starts where the previous ended, forming a closed contour. We can therefore generalize to $< R_0(q), K_0(q), \ldots, R_{2\pi}, K_{2\pi} >$ (i.e., $< J_0(q), \ldots, J_{2\pi}(q) >$). Intuitively, we can see that this is no different than the petals of a flower all starting from the center and proceeding in their respective angles outwards; the center in this case is q and the petal is each individual ray concatenated with its return path.
Definition $S(q) =< rot_0(R_0(q)), rot_0(K_0(q)), \ldots, rot_{-2\pi}(R_{2\pi}(q)), rot_{-2\pi}(K_{2\pi}(q)) >$, where rot_p is a rotation function which rotates the function around the domain p radians, making $S(q)$ a 1D representation of $J(q)$.
Proposition $J(q)$ is a non-domain specific signal of the heterogeneity.

Fourier Descriptor's Computation

The steps followed for the creation of the feature vector from $R(q)$ using FD are presented below and are illustrated by Figure 5.12.

Step 1: Rotate each $J_\theta(q)$ by $-\theta$, resulting in all rays originating at q and having an orientation of 0°.

Step 2: Concatenate each J_θ to create $J(q)$. We can see the result of this in Figure 5.12(b) is a 1D signal representation of $R(q)$.

Step 3: Compute the magnitude of $\mathcal{F}(J(q))$, the Fourier transform of $J(q)$, as $F(q)$. From the proof demonstrated in [68], this leads to a rotationally invariant representation of $J(q)$ in the frequency space, $F(q)$.

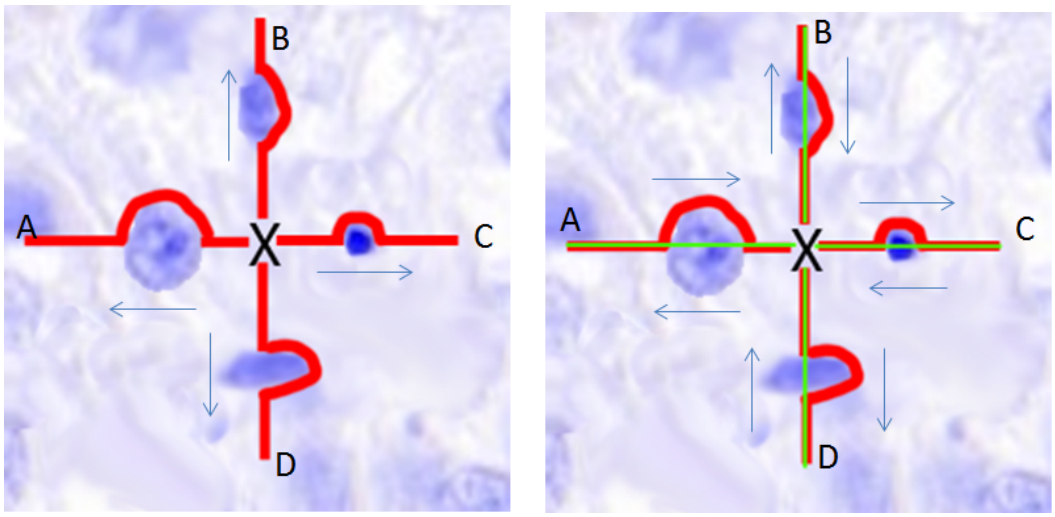
Algorithmic Notes

Typically the Fourier transform quantifies the frequency presence in a signal. In the case of using image data, which is discrete, there is often a high amplitude sawtooth wave like property as shown in the blue curve of Figure 5.14(a). These sawtooth signals require a high number of FFT coefficients to accurately represent them without adding a large amount of residual noise due to the approximation. To compensate, we smooth $F(q)$ using a simple moving average filter. Figure 5.14 shows a zoomed in example of this smoothing operation. we can see in Figure 5.14(b) that the red curve is indeed less susceptible to the discrete nature of the image scene, and thus could be more easily represented by fewer FFT coefficients.

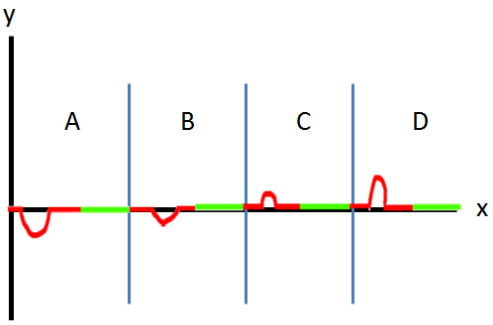
As the end points of the 1D signal tend to suffer the most from approximation, we compensate by wrapping the signal, e.g., joining the last 25% of the signal to the beginning, and the first 25% of the signal to the end. Post smoothing we undo this process by removing the added information. This creates noticeably smoother boundary values.

5.4.4 Training Classifier for Differentiation

For all q of interest, as identified by HNCut using the red swatch in Step 1 of Section 5.3, we arrange their respective $F(q)$ row-wise to form a matrix M and compute the t rank truncated Singular Value Decomposition such that $\hat{M} = U_t \Sigma_t V_t^T$, and thus represent each $F(q)$ as its dimensionality reduced $U_t(q)$ counterpart. This allows for the training of classifiers that are simultaneously accurate and computationally efficient.

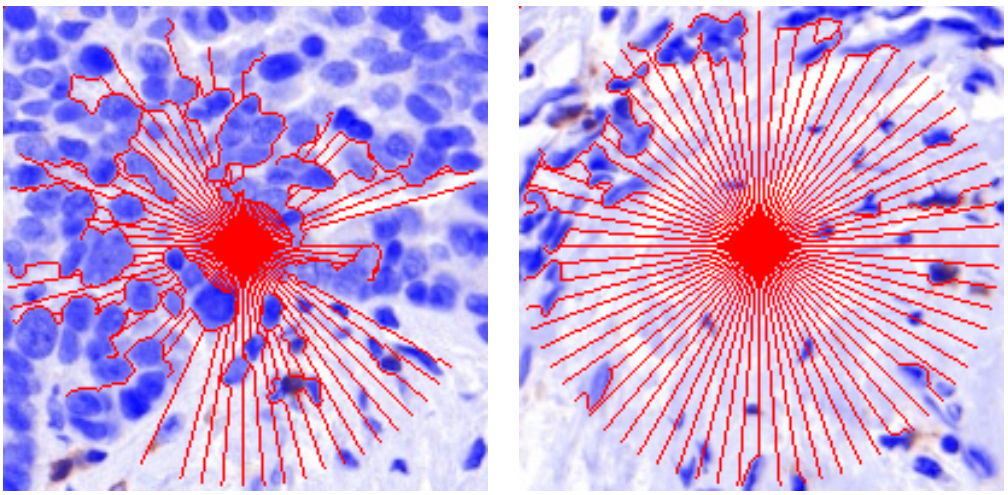


(a) (b)



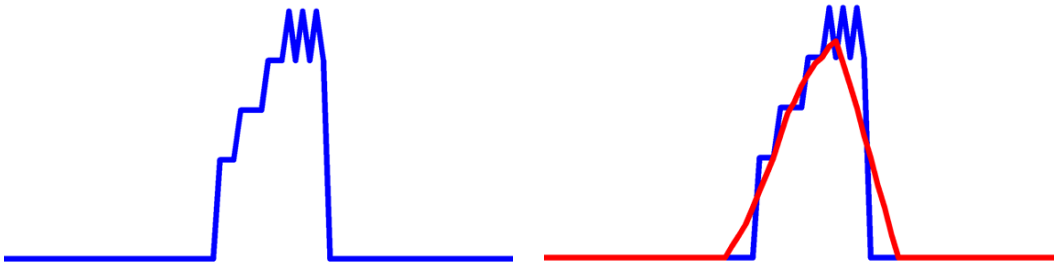
(c)

Figure 5.12: Visual example of the conversation of R to J. In (a) and (b) the blue cells represent obstructions causing local heterogeneity. As the μ-paths are minimized, we can see avoidance of these objects. Afterward we form a closed contour J, in (b) by the red and green lines, by concatenating R, the red lines in (a), and K, the green lines in (b). The S computed from J is displayed in (c). This results in a 1 dimensional signal which can be used to compute feature vector.



(a) (b)

Figure 5.13: Overlaid red LMS signatures in both (a) TIL and (b) a non-TIL image. We can see that the homogeneity of (b) is higher, and as a result the LMS rays appear more smooth and less contorted. On the other hand, in (a), we see a TIL which results in notably more tortuous LMS rays.



(a) (b)

Figure 5.14: A selected piece of (a) $R(q)$ shows that it tends to be subject to the discrete nature of the pixel image domain. On the other hand, after applying a (b) smoothing filter (in red) we can see that the function possesses qualities which are better suited for Fourier transform representation, namely a stronger signal with less fluttering.

Chapter 6

LMS: Experiments & Results

6.1 Dataset Description

To quantitatively evaluate the properties of LMS associated with differentiation between various classes using our non-domain specific features, we created 10 synthetic datasets, as shown in Figure 6.1. The production of the 1000 250×250 images per datasets is described in Table 6.1. Further, we compare the results to both texture features and ball-scale, which as was discussed in Chapter 1.1.2 and Chapter 2.2, are potential other solutions to the TIL versus non-TIL classification problem.

6.2 Experimental Setup

Only a *single point* was used for the classification of each image. For all queried points in the datasets ($q \in Q$), 50% were used as training data (Q_{tr}) while the remaining 50% were used as test data (Q_{te}) such that $Q_{te} \cup Q_{tr} = Q$. For the datapoints in Q_{tr}, the respective feature vector was fed into a naive Bayesian supervised classifier [73] which fits a multivariate normal density to each class, using a pooled estimate of covariance. The classes which were used were stromal = 0 and tumoral = 1, thus we attempted to provide a high probability for TILs. Q_{te} were classified as belonging to the first of the dataset classes and a Receiver Operating Characteristic curve (ROC) was computed. The Area Under the ROC (AUC) was computed for the run. This procedure was performed 50 times, each time a new training and testing set was randomly chosen. Mean and variance of the AUC was calculated.

Set	Description
Z_1	10 circles of size 10 ± 2, randomly placed. The goal was to model tumor regions.
Z_2	10 ellipses of size 10 ± 2 with a ratio of .2 of major to minor axis, spindle shape in nature, randomly orientated and placed. The goal was to model the stroma region.
Z_3	Similar to Z_1, except all circles are forced to be in quadrant IV.
Z_4	Similar to Z_1, except all circles are forced to be in quadrant I.
Z_5	Similar to Z_1, except all circles are forced to be within 60 pixels from the center.
Z_6	Similar to Z_1, except all circles are forced to be greater than 80 pixels from the center.
Z_7	Combination of Z_1 and Z_2 with the first 2/3 allocated to objects from Z_2 and the remaining 1/3 solely for Z_1
Z_8	Combination of Z_1 and Z_2 with the first 2/3 allocated to objects from Z_1 and the remaining 1/3 solely for Z_2
Z_9	Combination of Z_1 and Z_2 with quadrants I, II, and III allocated to objects from Z_1 and quadrant IV allocated to objects from Z_2
Z_{10}	Combination of Z_1 and Z_2 with quadrants I, II, and III allocated to objects from Z_2 and quadrant IV allocated to objects from Z_1

Table 6.1: Explanation of the 10 synthetic datasets presented in Figure 6.1.

6.2.1 LMS Setup

The LMS was generated for the center point. The LMS signature $L(q)$ was 2048 Fourier coefficients large and their respective U_t, as computed in the previous chapter, with $t = 5$. The other parameters used were $\epsilon = 5$, the smoothing average was set to a neighborhood of 5, and the initial mask was set to a 10 pixel radius from the point of interest. This mask size was selected as intentionally as the average expected size of the objects.

6.2.2 Texture Features

To provide some qualitative comparison, we performed the same experiment as described above with the standard 16 Haralick texture features as described in [74]. These consisted of {angular second moment, contrast, correlation, sum of squares, variance, inverse difference moment, shade, prominence, sum average, sum variance, sum entropy, entropy, difference variance, dif-

ference entropy, information measures of correlation, maximal correlation coefficient}. Since these are binary images, the co-occurrence matrix was computed for 2 gray levels (black and white). Again each feature vector of size 16 was fed into the same type of naive Bayesian supervised classifier.

6.2.3 Ball Scale Feature

We implemented the *b*-scale algorithm as per [49], and computed the radius around the POI encapsulating their described homogeneity constraint. This radius was used as the *b*-scale feature, and was used in its naive Bayesian supervised classifier. The feature "vector" then, in this case, was a scalar value.

6.3 Experiment 1: Examination of 10 Set Results

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
Z_1	n/a	1.0±4e-8	1.0±8e-7	1.0±5e-7	0.99±3e-6	0.95±3e-5	0.96±2e-5	**0.78±1e-4**	**0.86±7e-5**	0.97±2e-5
Z_2	1.0±5e-8	n/a	1.0±1e-6	1.0±6e-7	1.0±3e-9	1.0±7e-7	1.0±4e-6	1.0±2e-7	1.0±2e-6	1.0±1e-6
Z_3	1.0±6e-7	1.0±1e-6	n/a	**0.50±2e-4**	1.0±5e-8	1.0±2e-7	0.96±2e-5	1.0±4e-7	1.0±2e-7	**0.63±2e-4**
Z_4	1.0±5e-7	1.0±8e-7	**0.49±2e-4**	n/a	1.0±1e-7	1.0±9e-7	0.93±4e-5	1.0±3e-7	1.0±1e-7	**0.61±2e-4**
Z_5	0.99±4e-6	1.0±3e-9	1.0±5e-8	1.0±1e-7	n/a	1.0±1e-6	1.0±2e-7	0.99±5e-6	1.0±2e-6	1.0±9e-7
Z_6	0.95±3e-5	1.0±7e-7	1.0±2e-7	1.0±7e-7	1.0±1e-6	n/a	0.99±8e-6	0.99±3e-6	0.99±7e-6	1.0±3e-6
Z_7	0.96±3e-5	1.0±5e-6	0.96±3e-5	0.93±4e-5	1.0±2e-7	0.99±9e-6	n/a	0.95±3e-5	**0.90±5e-5**	**0.82±1e-4**
Z_8	**0.78±9e-5**	1.0±1e-7	1.0±3e-7	1.0±2e-7	0.99±7e-6	0.99±3e-6	0.95±3e-5	n/a	**0.76±1e-4**	0.97±2e-5
Z_9	**0.86±6e-5**	1.0±1e-6	1.0±2e-7	1.0±1e-7	1.0±1e-6	0.99±5e-6	**0.90±6e-5**	**0.76±1e-4**	n/a	0.96±3e-5
Z_{10}	0.97±2e-5	1.0±1e-6	**0.63±1e-4**	**0.61±2e-4**	1.0±9e-7	0.99±4e-6	**0.81±1e-4**	0.97±2e-5	0.96±3e-5	n/a

Table 6.2: Local Morphologic Scale AUC results using $\epsilon = 5$, across 50 runs with variance, indicating the success in differentiating pair-wise classes. For ease of viewing, scores less than or equaled to .90 are highlighted in bold.

There are many interesting things to note regarding the comparison of texture features results (see Table 6.3) versus local morphological scale results (see Table 6.2). Mainly we focus on the ability for LMS to differentiate between classes which are morphologically different without specifically coding features for them. We present the results from *b*-scale for completeness (see Table 6.4), though it is evident that the algorithm doesn't form a strong classifier for separating these classes in a robust fashion. This is unsurprising as the basis for the approach was to define local homogeneity for use in denoising applications, as such we focus on

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
Z_1	n/a	1.0±0e+0	**0.90±5e-5**	**0.90±3e-5**	0.96±2e-5	**0.62±2e-4**	1.0±3e-8	1.0±4e-7	1.0±1e-9	0.99±5e-6
Z_2	1.0±0e+0	n/a	1.0±0e+0	1.0±0e+0	1.0±0e+0	1.0±0e+0	1.0±0e+0	1.0±0e+0	1.0±0e+0	1.0±0e+0
Z_3	0.90±6e-5	1.0±0e+0	n/a	**0.49±3e-4**	**0.70±1e-4**	**0.86±8e-5**	1.0±3e-10	1.0±6e-13	1.0±4e-11	1.0±2e-7
Z_4	**0.90±4e-5**	1.0±0e+0	**0.49±2e-4**	n/a	**0.71±2e-4**	**0.85±5e-5**	1.0±3e-11	1.0±2e-10	1.0±6e-11	1.0±2e-7
Z_5	0.96±1e-5	1.0±0e+0	**0.70±1e-4**	**0.70±1e-4**	n/a	0.94±3e-5	1.0±0e+0	1.0±0e+0	1.0±0e+0	1.0±1e-7
Z_6	**0.61±1e-4**	1.0±0e+0	**0.86±7e-5**	**0.86±8e-5**	0.94±3e-5	n/a	1.0±6e-9	1.0±2e-7	1.0±1e-8	0.99±3e-6
Z_7	1.0±2e-8	1.0±0e+0	1.0±1e-11	1.0±7e-11	1.0±3e-12	1.0±2e-8	n/a	**0.64±2e-4**	**0.87±7e-5**	**0.87±6e-5**
Z_8	1.0±4e-7	1.0±0e+0	1.0±3e-12	1.0±9e-11	1.0±3e-13	1.0±2e-7	**0.64±2e-4**	n/a	0.92±4e-5	0.92±4e-5
Z_9	1.0±1e-9	1.0±0e+0	1.0±6e-11	1.0±4e-11	1.0±0e+0	1.0±1e-8	**0.87±8e-5**	0.92±4e-5	n/a	**0.90±4e-5**
Z_{10}	0.99±5e-6	1.0±0e+0	1.0±2e-7	1.0±3e-7	1.0±7e-8	0.99±3e-6	**0.87±6e-5**	0.92±3e-5	0.89±3e-5	n/a

Table 6.3: Texture features confusion matrix of AUC, across 50 runs with variance, indicating the success in differentiating pair-wise classes. For ease of viewing, scores less than or equaled to .90 are highlighted in bold.

comparing LMS to texture features.

- In all cases, Z_2 is differentiable from the other classes by both algorithms. This leads us to believe that in the clear cut cases, both algorithms are able to identify a non-TIL with ease. This further motivated us to create the other 8 sets as to delve further into the specific edge cases where the algorithms don't perform comparatively.

- Comparing Z_1 and Z_6 we can see that the LMS is able to differentiate the two classes quite well, with an AUC of .95 as compared to texture features obtaining only a .61. This is similar to: Z_4 versus Z_5 (1 LMS vs .7 texture), Z_4 versus Z_6 (1.0 LMS vs .86 texture) and Z_3 versus Z_6 (1.0 LMS vs .86 texture). In all cases, texture features is only concerned with the information content represented by the pixels, in these cases a non-discriminating feature, while LMS specifically models the location and size of the objects with respect to the pixel of interest allowing it to perform significantly better in those cases.

- Comparing Z_7 to Z_8 (.95 LMS versus .64 texture features) is valuable as it specifically aims at modeling the boundary of regions. Specifically for our TILs, it is important to identify if they are embedded *in* the tumor as opposed to simply next to it. By creating a scenario which has the pixel of interest on either side of the boundary, we can see that LMS is more successful at identifying the correct class as opposed to texture features.

- Comparing Z_9 to Z_{10} we see that LMS is obtaining a .96 AUC versus .89 from texture features. This is relatively unsurprising as both classes contain the same number and size

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
Z_1	n/a	0.87±1e-4	0.77±1e-4	0.62±2e-4	0.62±2e-4	0.49±4e-4	0.60±2e-4	0.69±1e-4	0.59±1e-4	0.59±2e-4
Z_2	0.87±1e-4	n/a	0.72±1e-4	0.89±9e-5	0.81±9e-5	0.85±8e-5	0.88±7e-5	**0.90±6e-5**	0.89±8e-5	0.89±7e-5
Z_3	0.77±2e-4	0.72±1e-4	n/a	0.80±1e-4	0.70±2e-4	0.73±1e-4	0.80±2e-4	0.82±1e-4	0.79±1e-4	0.80±1e-4
Z_4	0.62±1e-4	0.89±5e-5	0.80±1e-4	n/a	0.69±2e-4	0.59±1e-4	0.47±9e-4	0.57±2e-4	0.53±4e-4	0.53±2e-4
Z_5	0.62±2e-4	0.81±8e-5	0.70±2e-4	0.69±2e-4	n/a	0.59±2e-4	0.72±2e-4	0.73±2e-4	0.68±2e-4	0.67±2e-4
Z_6	0.49±4e-4	0.85±8e-5	0.73±2e-4	0.58±2e-4	0.59±2e-4	n/a	0.55±2e-4	0.65±2e-4	0.55±2e-4	0.56±2e-4
Z_7	0.60±2e-4	0.88±1e-4	0.80±2e-4	0.47±1e-3	0.71±2e-4	0.55±2e-4	n/a	0.62±2e-4	0.50±2e-4	0.50±2e-4
Z_8	0.69±1e-4	**0.90±5e-5**	0.82±1e-4	0.57±1e-4	0.73±2e-4	0.65±1e-4	0.62±2e-4	n/a	0.62±1e-4	0.60±2e-4
Z_9	0.59±1e-4	0.89±9e-5	0.80±1e-4	0.53±9e-4	0.69±2e-4	0.55±1e-4	0.50±1e-4	0.62±1e-4	n/a	0.49±1e-4
Z_{10}	0.59±1e-4	0.89±6e-5	0.80±2e-4	0.53±1e-4	0.67±2e-4	0.56±1e-4	0.50±2e-4	0.60±2e-4	0.49±2e-4	n/a

Table 6.4: Ball Scale feature confusion matrix of AUC, across 50 runs with variance, indicating the success in differentiating pair-wise classes. For ease of viewing, only scores greater than or equaled to .90 are highlighted in bold.

of objects, except that the topology is different as a result of the quadrants which the circles are confined. Interestingly though Z_8 to Z_{10} (.97 LMS versus .92 texture features) seem to do better in both algorithms, though their structure isn't greatly different from Z_9.

- Comparing Z_1 to Z_8 we notice that LMS tends to struggle to differentiate them very well (.78 AUC). On deeper examination, this is as a result of the larger objects overpowering the signal from the small ellipses. While they appear in the LMS signature, the Fourier descriptor based representation struggles to isolate the frequencies specific to those pieces as they are a subset of the frequencies of the larger objects. This is especially notable in comparison of Z_{10} to Z_3 and Z_4.

- Comparing Z_1 to Z_9, this overpowering characteristic is less prevalent (.86 AUC), we believe this is because the ellipses are isolated and not shadowed by larger objects. This is to say they have their own LMS rays associated solely for them, instead of a single LMS ray encountering both circles and ellipses.

6.4 Experiment 2: Rotational Invariance

As we can see from the comparison of set Z_3 and set Z_4, LMS is unable to differentiate them obtaining 0.5 AUC. This is notable as these two sets are actually 90° rotations of one another.

We can thus infer, that since LMS cannot distinguish between them, that the approach does indeed have the property of rotation invariance.

6.5 Experiment 3: Efficiency

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
Z_1	n/a	1.0±3e-7	0.99±1e-6	1.0±2e-6	0.99±5e-6	0.96±3e-5	0.93±5e-5	**0.77±1e-4**	**0.81±1e-4**	0.96±2e-5
Z_2	1.0±2e-7	n/a	1.0±2e-6	1.0±1e-6	1.0±4e-9	0.99±5e-6	0.99±5e-6	1.0±6e-9	0.99±3e-6	1.0±2e-6
Z_3	0.99±1e-6	1.0±3e-6	n/a	**0.48±2e-4**	1.0±7e-8	1.0±1e-6	0.95±2e-5	1.0±9e-7	1.0±7e-7	**0.63±2e-4**
Z_4	1.0±1e-6	1.0±1e-6	**0.48±2e-4**	n/a	1.0±1e-7	1.0±2e-6	0.95±3e-5	1.0±6e-7	1.0±8e-7	**0.64±1e-4**
Z_5	0.99±5e-6	1.0±5e-9	1.0±6e-8	1.0±1e-7	n/a	1.0±1e-6	1.0±1e-6	0.99±8e-6	0.99±2e-6	1.0±9e-7
Z_6	0.96±3e-5	0.99±5e-6	1.0±1e-6	1.0±2e-6	1.0±2e-6	n/a	0.97±2e-5	0.99±3e-6	0.98±1e-5	0.99±8e-6
Z_7	0.93±5e-5	1.0±6e-6	0.95±3e-5	0.95±3e-5	1.0±8e-7	0.97±2e-5	n/a	0.93±5e-5	**0.86±1e-4**	0.80±2e-4
Z_8	**0.77±1e-4**	1.0±4e-9	1.0±9e-7	1.0±6e-7	0.99±1e-5	0.99±3e-6	0.93±3e-5	n/a	**0.74±1e-4**	0.96±2e-5
Z_9	**0.81±1e-4**	0.99±3e-6	1.0±7e-7	1.0±8e-7	0.99±2e-6	0.98±1e-5	**0.87±8e-5**	**0.74±1e-4**	n/a	0.94±4e-5
Z_{10}	0.96±2e-5	1.0±1e-6	**0.63±1e-4**	**0.64±2e-4**	1.0±8e-7	0.99±6e-6	**0.79±1e-4**	0.96±2e-5	0.94±3e-5	n/a

Table 6.5: Local Morphologic Scale AUC results Using $\epsilon = 10$, across 50 runs with variance, indicating the success in differentiating pair-wise classes. For ease of viewing, scores less than or equaled to .90 are highlighted in bold.

Using a Matlab implementation on a laptop with a 2.50GHz processor, we were able to generate 1,000 LMS signatures in 17 seconds or 0.017 sec/sample. We note that the time for creation of the LMS indicator function $f(q)$ is 0.008 seconds, and the generation of the LMS signature from it is 0.005 seconds. When using all 4 cores of the same machine, computation time dropped to 8.398 seconds per 1,000 or .00839 sec/sample. We note that the speed up is only 2x faster as opposed to the theoretical 4x due to various other system bottlenecks such as disk reading of the images, setting up and distributing the work across the Matlab pool. If additional efficiency is needed, it is possible to reduce ϵ. For example, when $\epsilon = 10$, the time to generate 1,000 LMS signatures drops to 10.38 seconds (40% improvement in speed) in the serial format, and 7.81 seconds using 4 cores. Of course, this comes with some slight degradation of results as shown in Table 6.5. We present a more complete discussion on the effects of ϵ using real-world data in the next chapter.

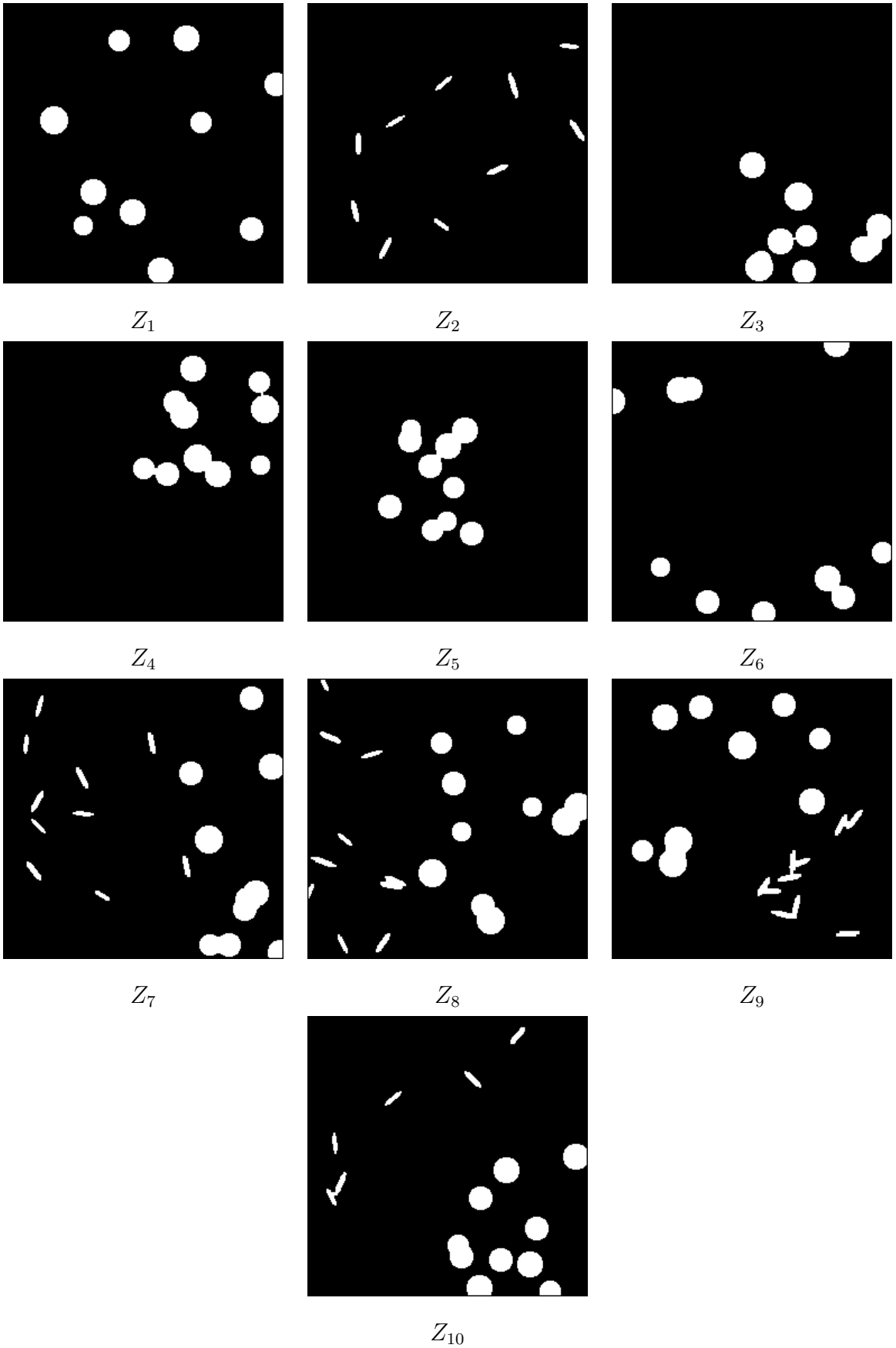


Figure 6.1: We present 10 different synthetic datasets, each of 1000 250×250 *images, all containing the same object density of 10 or 20 objects per image. These images are a suitable test ground to display the properties associated with the LMS approach. For a full description see Table 6.1*

Chapter 7

Detection of Tumor Infiltrating Lymphocytes

7.1 Introduction to Real World Applications

We apply the framework which was presented in the previous chapters to the real world application of detecting tumor infiltrating lymphocytes. Due to the expensive nature of acquiring annotated information from pathologists, we only have lymphocyte information for the ovarian cancer (OCa) domain; as such we perform extensive vetting of the system solely in this domain in Section 7.2.1. For comparison in 3 other domains, we investigate a parallel experiment in Section 7.3: *if* a randomly selected pixel *were* a lymphocyte, would it have been correctly identified as a TIL or non-TIL?

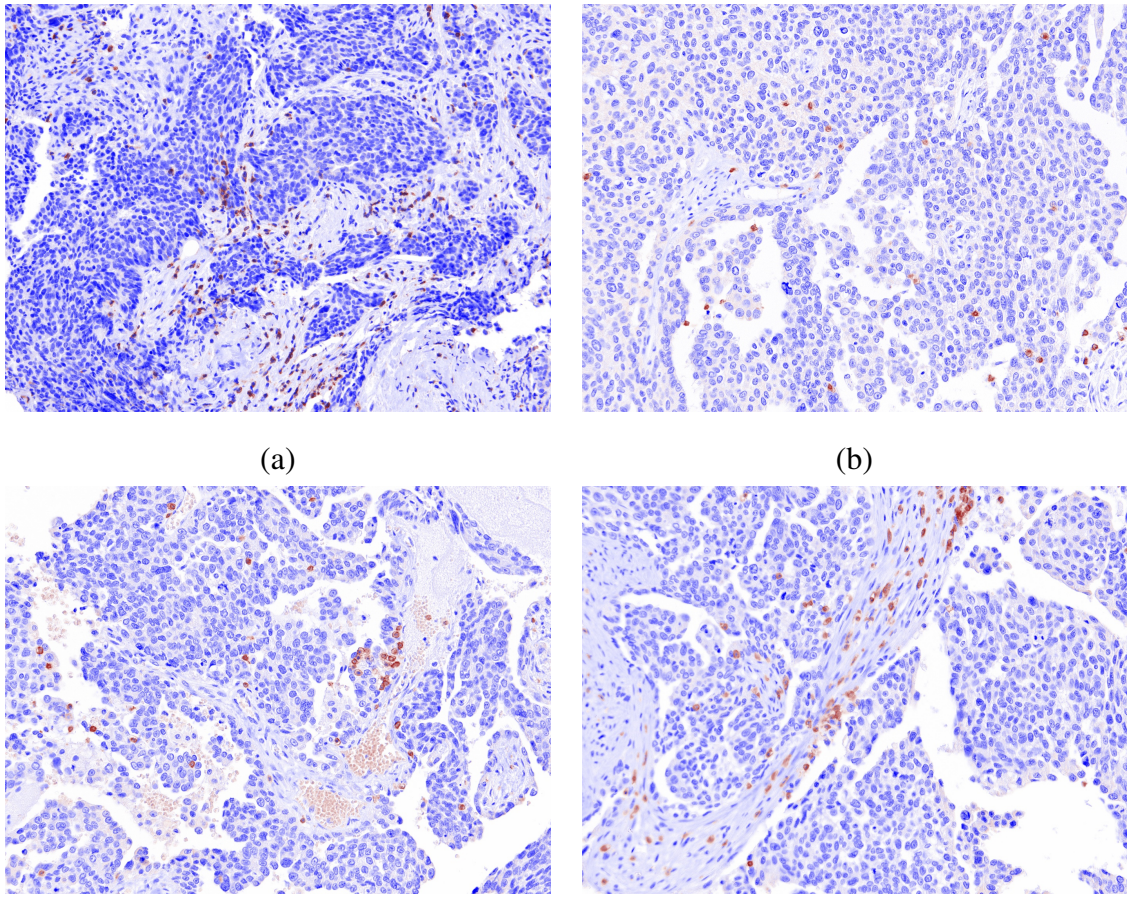
7.2 Experiments In TIL Identification

7.2.1 Training and Testing Methodology

The training and testing methodology were performed exactly the same as in Chapter 6.2. Please refer to this section for all necessary details.

7.2.2 Data Set Description

The data set consisted of 60 slide mount ovarian cancer images of size 1400 x 1050. Each slide was stained with hematoxylin, which makes the tumor and endothelial cells appear blue



(a) (b)

(c) (d)

Figure 7.1: Four sample images from the OCa data set. Each image is 1400 x 1050, and the blue endothelial and tumor cells are visible and contrasted with the red stained lymphocytes. The homogeneous white regions are areas without cells as a byproduct of the biopsy and mounting procedures.

and a CD3 positive T stain which caused the lymphocytes to appear in red. The images were then scanned using a 40x magnification. We display some of the example images in Figure 7.1. In total, 4320 lymphocytes were identified using HNCuts. Although the segmentation was straight forward due to the strong contrast between the red CD3 stained lymphocytes and the blue hematoxylin counter-stain, each lymphocyte was still reviewed by an expert to ensure it was correctly delineated. Typically, it is a lot easier to review automatically annotated data for correctness than it is to perform the segmentation manually. As a result, we can ease the burden on the expert while still ensuring a high quality data set. As part of the review, the lymphocytes were divided into 1402 TILs and 2918 non-TILs, forming the ground truth.

7.2.3 Experiment 1: Ovarian Cancer TIL Identification

We aim to train our classifier to differentiate between TILs and non-TILs in the OCa data set. We compare the results from LMS to texture features and ball scale, which as discussed in 1.1.2 and Chapter 2.2 are potential approaches to solving this problem as well. We ensured that training samples were never pulled from images which appeared in the test set.

Algorithm Setup

- **Local Morphologic Scale**: The algorithm proceeds as per the flowchart presented in Figure 5.7. The only additional processing performed was to use a watershed algorithm [75] to quickly separate large regions. This was necessary as the cell clumping tended to interfere with the LMS signature generation and watersheds quickly breaks large monolithic areas into over-segmented patches. We mention, though, that these over-segmentation patches ideally would be individual cells, but that was often not the case. To determine the optimal operating parameters, a grid search was conducted seeking the best settings for the algorithm. We present the searched domain in Table 7.1.

Variable	Searched Values
Degree sampling (ϵ)	1, 5, 10, 15, 30
Number of FFT coefficients	$2^8, 2^9, 2^{10}, 2^{11}, 2^{12}$
Smoothing Neighbors	1, 5, 10
Singular Value Dimension (t)	2, 5, 10, 25
Window Size (w)	25, 50, 100, 125, 200

Table 7.1: Description of all grid-searched variables and their associated attempted values.

The optimal values were found to be: $\epsilon = 1$, $t = 5$, $w = 50$, number of coefficients for the Fourier transform = 2^9, number of points used for smoothing = 10, mask size = 10. This means that each signature vector starts off as 512 and then is reduced down to 5 via SVD.

- **Texture Features**: Through a grid search of the variables presented in Table 7.2, we were able to identify the optimal parameters for the texture features as window size set to 50 and the number of gray levels as 16 and the singular value space as 25.

Variable	Searched Values
Number of Gray Levels	8, 16, 32, 64
Singular Value Dimension (t)	2, 5, 10, 25, 50
Window Size	25, 50, 100, 125, 200

Table 7.2: Description of all grid-searched variables and their associated attempted values.

- **Ball Scale**: For this we used the same implementation as in [49], and did not enforce a window size. This had little effect on the output as the radius of the ball of homogeneity never grew large enough for concern.

Results

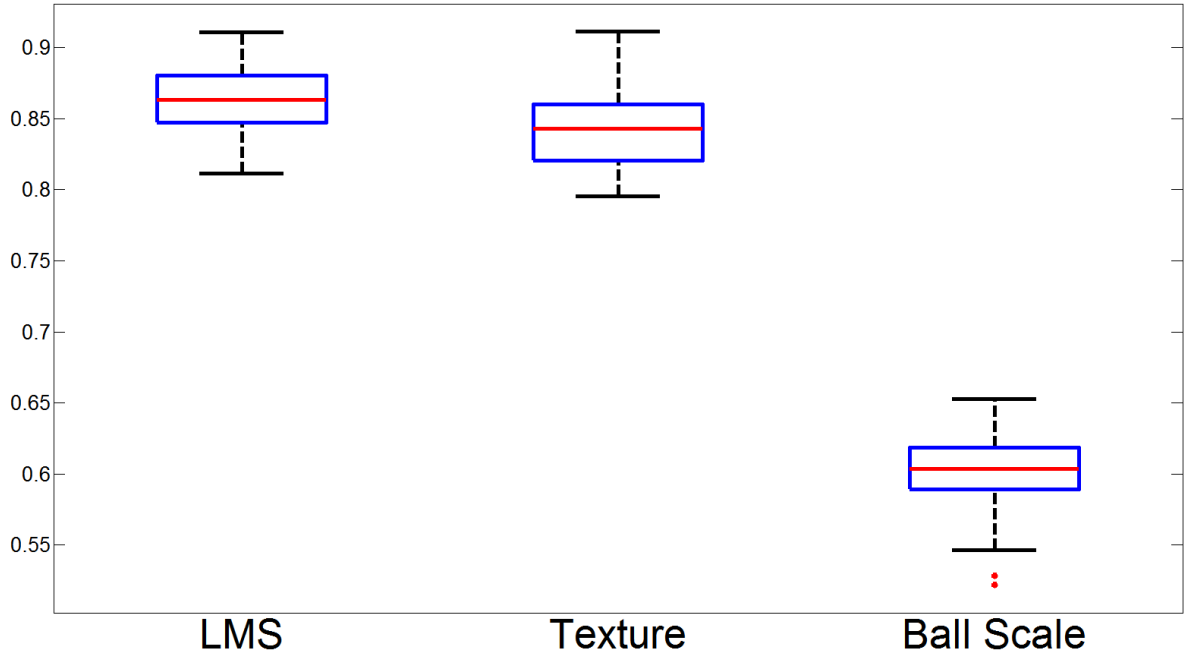


Figure 7.2: Box plots for the AUC across 50 runs from all 3 algorithms. The red line identifies the mean, the blue box encompasses 25th percentile, with the black whiskers extending to the 75th percentile. Red dots are indicative of outliers. We can see that the LMS provides a higher mean AUC than texture features with a smaller variance. On the other hand, ball scale seems to produce a poor classifier.

We present the box plots for the 3 approaches in Figure 7.2, with the associated true positive average in Figure 7.3 and true negatives in Figure 7.4. We can see that with a mean

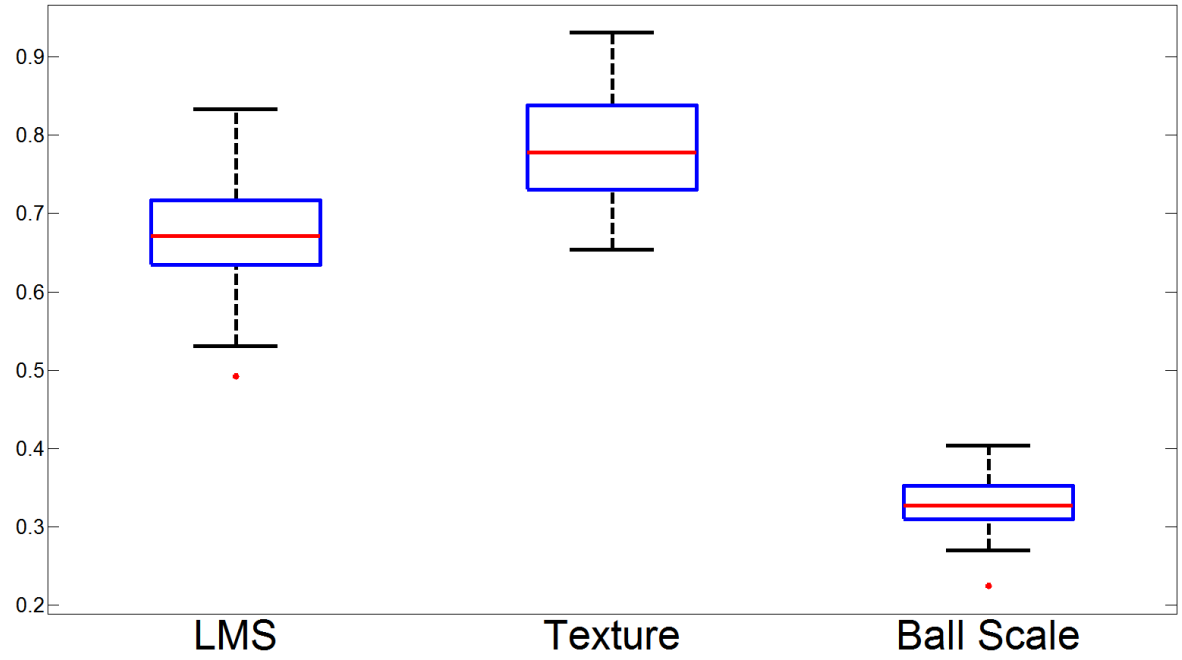


Figure 7.3: Box plots for the true positives across 50 runs from all 3 algorithms. The red line identifies the mean, the blue box encompasses 25th percentile, with the black whiskers extending to the 75th percentile. Red dots are indicative of outliers. We can see that b-scale is struggling to identify TILs with a rate of just 30% correct. While texture features appears to be performing better than LMS here, we note that this is only the case when the most optimal parameters (found via an expensive search procedure) are used. Experiment 2, shows LMS's resilience to a wide range of parameter settings.

AUC of .866 LMS provides a slightly better classifier than texture features with .842. These are comparable to the current state of the art approach [53] with their self-stated .88 accuracy. We draw attention to the significantly lower dimension of our approach as compared to their 6,000 features. Lastly, we can see that homogeneity is not an ideal separating characteristic as *b*-scale fares rather poorly in this classification task.

7.2.4 Experiment 2: Impact of Window Size

The question of how window size impacts the quality of the results is important. To test this, we perform the identical grid search as Experiment 1 except report the mean AUC across each tested window sizes (25, 50, 100, 125, 200) for both LMS and texture features.

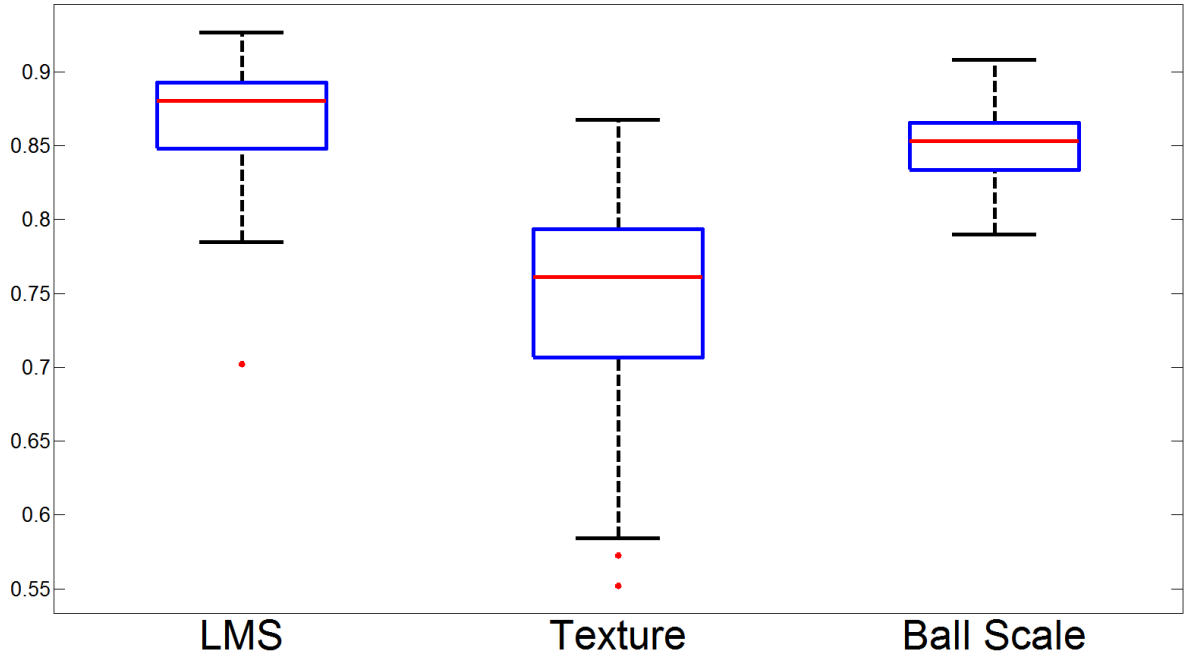


Figure 7.4: Box plots for the true negatives across 50 runs from all 3 algorithms. The red line identifies the mean, the blue box encompasses 25th percentile, with the black whiskers extending to the 75th percentile. Red dots are indicative of outliers. We can see that LMS notably provides the best identification of non-TILs from the 3 algorithms. Texture features seems to produce a very wide variance in its ability to correctly identify true negatives.

Results

We present in Figure 7.5 the mean AUC across 50 runs using the *optimal* parameters for each window size. From this experiment, we can clearly see that as the window size varies, the LMS approach keeps a consistent AUC while texture features degrade as additional information becomes available. This is a critical point. Since a grid search was used to find optimal operating variables for both LMS and texture features, which was extremely computationally expensive, an algorithm which has resilience to a very wide range of settings is preferred over one which has a much smaller range and is thus susceptible to constant ”tweaking”. LMS meets this criteria by producing consistently good results within a wide range of operating parameters, while texture features notably degrades outside of its optimal parameters. So while Figure 7.3 shows texture features outperforming LMS, it will only do so after an extensive computational investment to find its optimal settings.

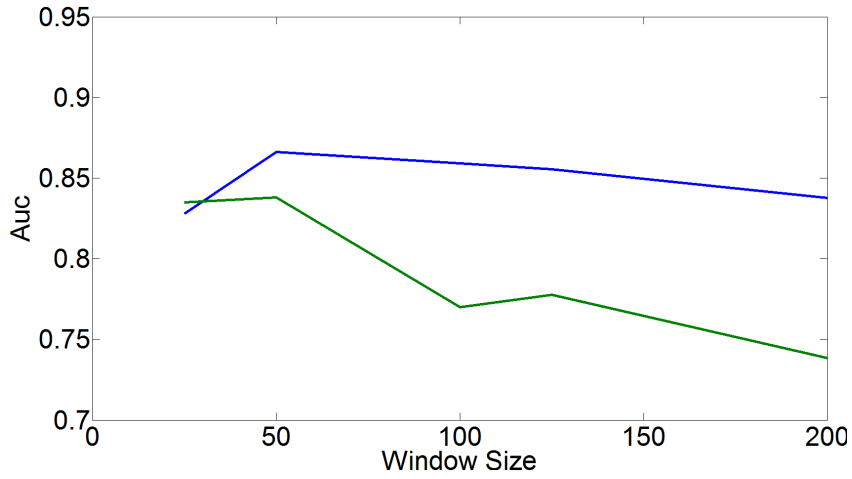


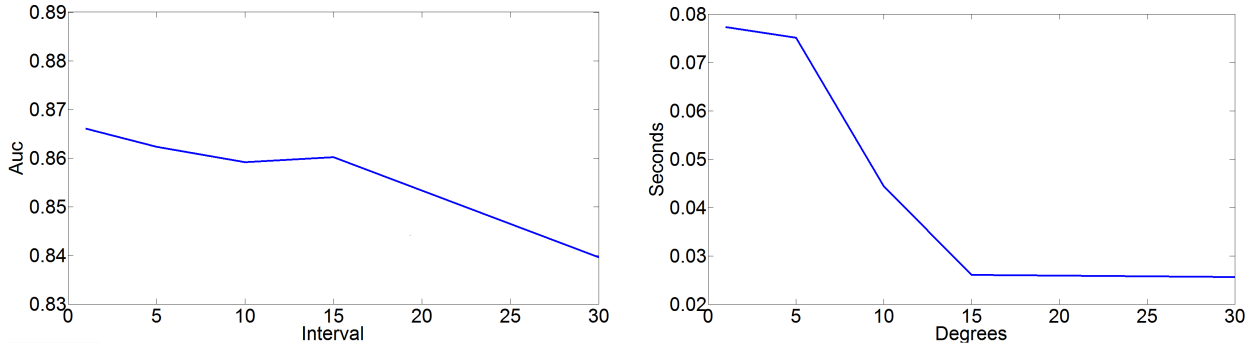
Figure 7.5: Average AUC using optimal parameters across a set of 5 varying window sizes. The LMS (the upper blue line) maintains a consistent AUC even as the window size grows very large. This is contrasted with the texture features (lower green line) graph which shows a degradation of results along with the expanding window size.

7.2.5 Experiment 3: Impact of Interval Size

The question of how the interval ϵ impacts the quality of the results is important. As a result of the Monte-Carlo sampling of the local region, we can trade some accuracy for speed, but the exact degradation of quality is important to identify. To observe the response of a reduced sampling, we perform the identical grid search as Experiment 1 except report the mean AUC across each tested ϵ tested (1, 5, 10, 15, 30) for both LMS.

Results

We present in Figure 7.6(a) the mean AUC across 50 runs using the optimal parameters for each ϵ interval and the associated time (b). From this experiment, we can see the expected behavior: as the rate of sampling is reduced the accuracy falls, as well as the time required per sample. The positive point to consider is that even as the sampling rate drops to 1/30th of the original (from 1° to 30° interval), the accuracy only falls about 3%, a favorable cost vs time ratio when we see that the time required drops 75%.



(a) (b)

Figure 7.6: Average AUC (a) using optimal parameters across a set of 5 varying ϵ intervals contrasted with the speed per sample in (b). As expected we see that as the sampling goes down, so does the time. The total degradation due to smaller sampling is only 3% in exchange for a 75% speed up.

7.2.6 Experiment 4: Combined Classifier

Given the results from Experiment 1, we decided to investigate how well a classifier would perform if it used the optimal configurations for LMS and texture features in a joint feature space. As such, we concatenate feature vectors produced by the two algorithms and use them in the same classifier, and report results across 50 runs.

Results

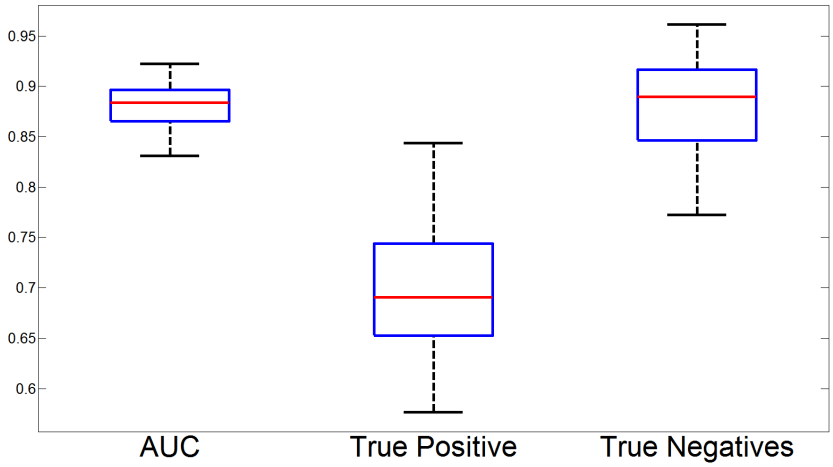


Figure 7.7: The three box plots associated with the joint classifier created by concatenating the LMS features with the texture features. We can see the combination of two of the feature sets produces better results, even using an unsophisticated classifier.

We present the AUC, true positive percent and the true negative percents across 50 runs in Figure 7.7. The true negative and average AUC approach .90, indicating a stronger classifier. When looking at this figure in comparison with Figure 7.2, Figure 7.3, and Figure 7.4, we can infer that the classifiers encapsulate different information, as their combination provides notable improvement with a decrease in variance. This result is a good indicator that LMS could be used in conjunction with other chromatic or texture based features for a more complete view of features of the query point leading to better classification performance.

7.3 Experiments in Region Identification

In this section we apply LMS to 3 other data domains: prostate cancer images stained with hematoxylin and eosin (H&E), prostate cancer images stained with eosin (E) and breast cancer images stained with H&E. Unfortunately, when these data sets were created, they were not cross-stained for identification of the lymphocytes. As a result, we slightly modify the previous TIL versus non-TIL experiment to use the stained data sets, with the laboriously created ground truth, we had available. Since we don't have lymphocyte information, which provided the points of interest in the previous experiment, we randomly pick locations in the image and "suppose" a lymphocyte was there. The classification process is still the same, if the pixel is inside of the tumor (as defined by the annotation provided by the expert) we hope to identify it as a TIL or tumor region versus not inside of the tumor being a non-TIL (i.e., stroma region).

7.3.1 Data Set Description

We can see the descriptions of the data sets used in this section in Table 7.3 with associated samples in Figure 7.8. For each data set, we randomly selected 100 points per image and followed the same procedure as Experiment 1.

7.3.2 Experiment 5: Breast and Prostate Pixel Classification

Setup

The operating parameters for LMS were identical to those in Experiment 1.

Data Type	Label	Properties	Number	Specific Challenge
Prostate TMA HE	S_1	H&E stain Appears blue	44 images 1600 x 1600	Classification of nuclear centers as tumor or stromal region
Prostate TMA H	S_2	Hematoxylin stain Appears purple	44 images 1600 x 1600	Classification of nuclear centers as tumor or stromal region
Breast	S_3	Hematoxylin stain Appears purple	51 images 1000 x 1000	Classification of nuclear centers as tumor or stromal region

Table 7.3: Description of non-lymphocyte data sets.

Data Type	Prostate HE	Prostate H	Breast
AUC $\pm$ Range	$.88 \pm .01$	$.87 \pm .02$	$.80 \pm .01$

Table 7.4: Bayesian classifier AUC in distinguishing stromal from tumoral lymphocytes for $S_1 - S_3$.

Results

Although the parameters were *not* individually tuned for each data set, we can see from the results in Table 7.4 that they are fairly consistent with the OCa results. For S_1 and S_2 we can see that the mean AUC is about $.88$ indicating excellent separation between tumoral and stromal cell nuclei. It is interesting to note that the H&E images did slightly better than the H alone images, most likely due to the greater contrast afforded by the counter-staining which allowed for the segmentation algorithm to more accurately create a binary mapping of the cellular information. Across 51 images of S_3 the breast images produced a mean AUC of .81. in spite of no domain specific tuning, the LMS still proceeded extremely well.

7.3.3 Qualitative Evaluation

In Figure 7.9 we present LMS signatures in red/green overlaid on both tumor and non-tumor regions for all domains. Consistently across tumor based regions (a-d), the heterogeneity created by the cancer cells is evident by the strong fluctuations in the LMS signature (e-h). As especially evident in Figure 7.9(a), we can see for a very complex region the LMS paths becoming increasingly tortuous as they adapt to the local heterogeneity. This is to say, we can see that as the complexity of the local region increases a noted change in the LMS occurs. This change is

as a result of rising entropy in the neighborhood structures resulting in the rays become more chaotic. Comparatively, in the stroma images (Figure 7.9(i-l)) we can see how the homogeneous regions have fewer obstructions as a result of the smaller endothelial cells having less of an impact on the signature and thus the LMS paths form straighter lines. Figure 7.9(l) illustrates a LMS signature for the point of interest is located in a stroma region, but bounded on the left and right sides by tumor. In this case, the LMS signature is able to extend unaltered in both north and south directions, while being constrained in the west and east directions. This is an example of when choosing a texture window size would be problematic as each orientation is constrained differently. Since LMS is rotation invariant, having a few training samples of this type easily extends to similar complex regions.

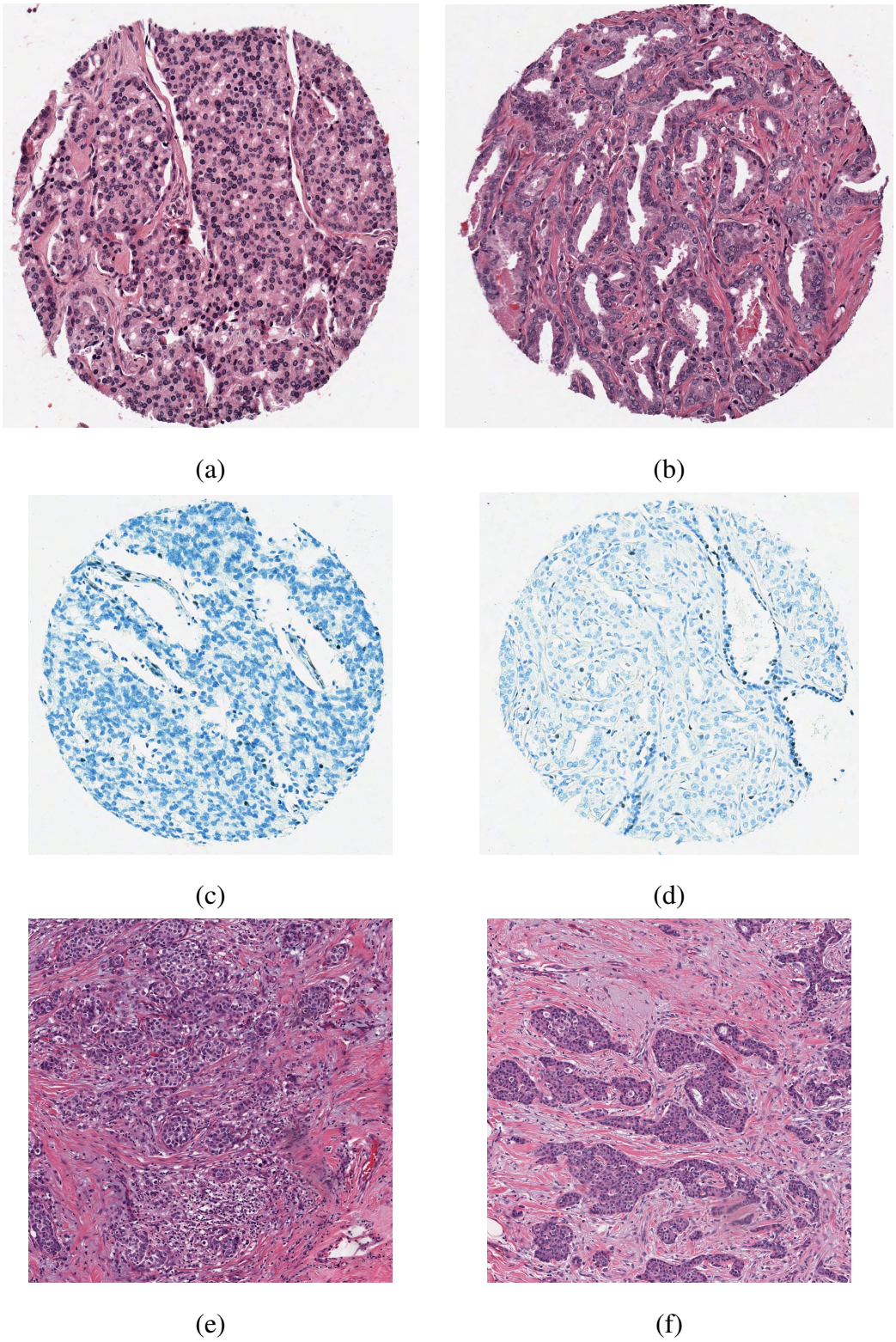


Figure 7.8: 2 sample images from each of the data sets described in Table 7.3. First row is prostate HE (S_1), second row is prostate H (S_2) and third row is Breast H (S_3). We can see that each of these images has its own unique characteristics which separate it from the OCa domain discussed earlier in this chapter.

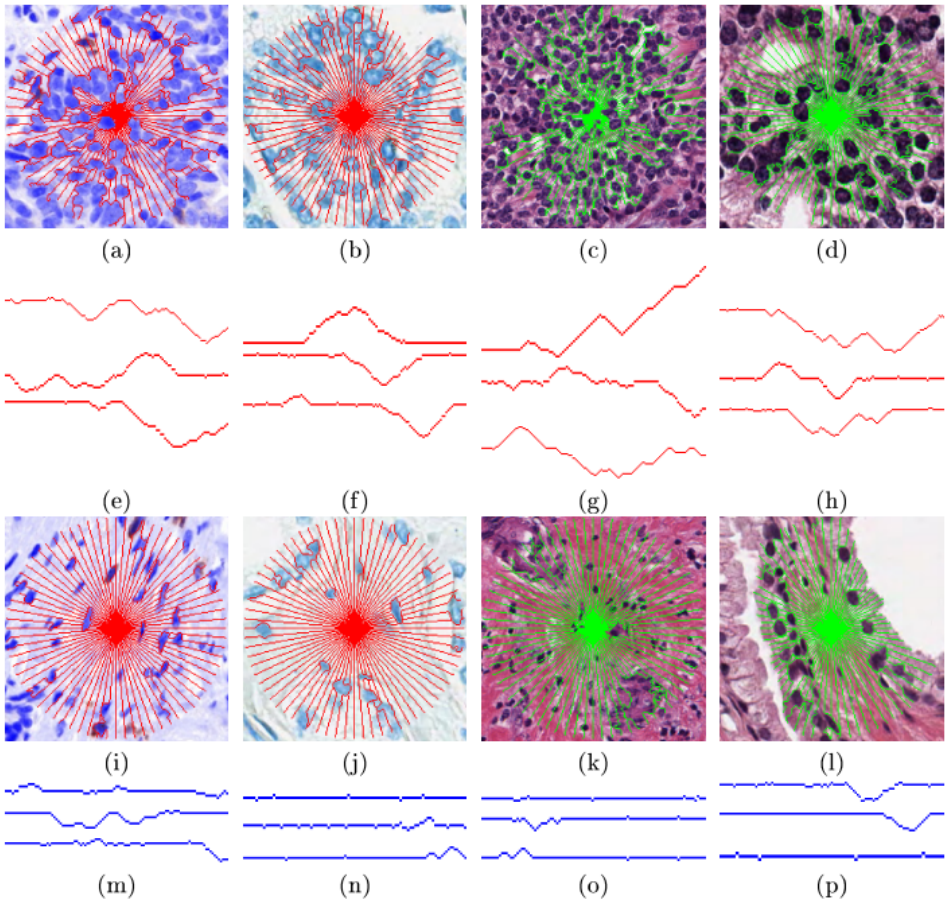


Figure 7.9: The LMS signature overlaid on a tumor regions in red/green in an (a) ovarian,(b) prostate H, (c) breast HE, and (d) prostate HE image, to be compared with the benign signatures in ((i)-(l)), respectively. Three rays from each image ((e)-(h) & (m)-(p)) are extracted and presented beneath their respective image. We can see that in the non-tumor regions ((i)-(l)) the LMS signature has fewer and smaller objects to obstruct its path, and thus the rays are less tortuous, unlike in the tumoral regions ((a)-(d)).

Chapter 8

Discussion and Future Work

In the previous chapters we have presented novel theories which have applications in both segmentation and classification. By instantiating a system which uses two derived algorithms, Hierarchical Normalized Cuts and Local Morphologic Scale, we are able to classify tumor infiltrating lymphocytes. Our high-speed solutions provide results that are comparable to state of the art approaches which are unable to surmount the huge data requirements necessary for current and future clinical usage. By showing satisfactory results across four data domains, without individual parameter tunings, our approaches await further work which will make them only more suited to broader clinical challenges and applications. As such, in the following sections we discuss possible improvements and additional applications, with the supporting motivation and high-level potential solutions. Additionally, we mention some of the challenges which will be faced during the implementation of each supplemental improvement.

Gradient Based Path Identifier

We can see that the LMS requires a binary map to compute its rays. This is a result of both a desired level of efficiency and greater simplicity of algorithm design. A notable concept which could be adopted in the future is the usage of an affinity computation function (μ) which operates directly on pixel values instead of using a binary version of the scene. The rays would then be computed via some gradient measure as they iteratively reach their destination. While this would remove the complexity of the HNCut step, it presents with it a novel set of challenges. From our brief experiments in the domain, our experience shows us that the problem is not trivial as it can be viewed as a subset of active contours. The problem, of course, is that even with a gradient measure, another constraint must be explicitly defined to determine exactly how

the curve should function, i.e., at what point should gradient affinity be scarified as a result of too steep of a slope derivative defined by the curve.

Descriptor Enhancement

From the comparison of Z_1 to Z_8 in Figure 6.1, we notice that large bodies which appear frequently can overpower the signal of the smaller bodies resulting in occasional occlusion of the needed differentiating signal. While this doesn't seem to have a large impact on the classifier for this specific domain, there is the possibility that there exists a better approach of quantifying the LMS signature. This isn't to say that Fourier descriptors are not suitable in many cases, but when examining the synthetic datasets, there does seem to be some indication that an approach which also quantifies exact location as opposed to solely signal variance may have added success. An approach such as wavelets meets this criteria, but the ability to compare wavelets, while still maintaining the generalizability due to rotational invariance, and use them as the features for a supervised classifier has yet to be seen.

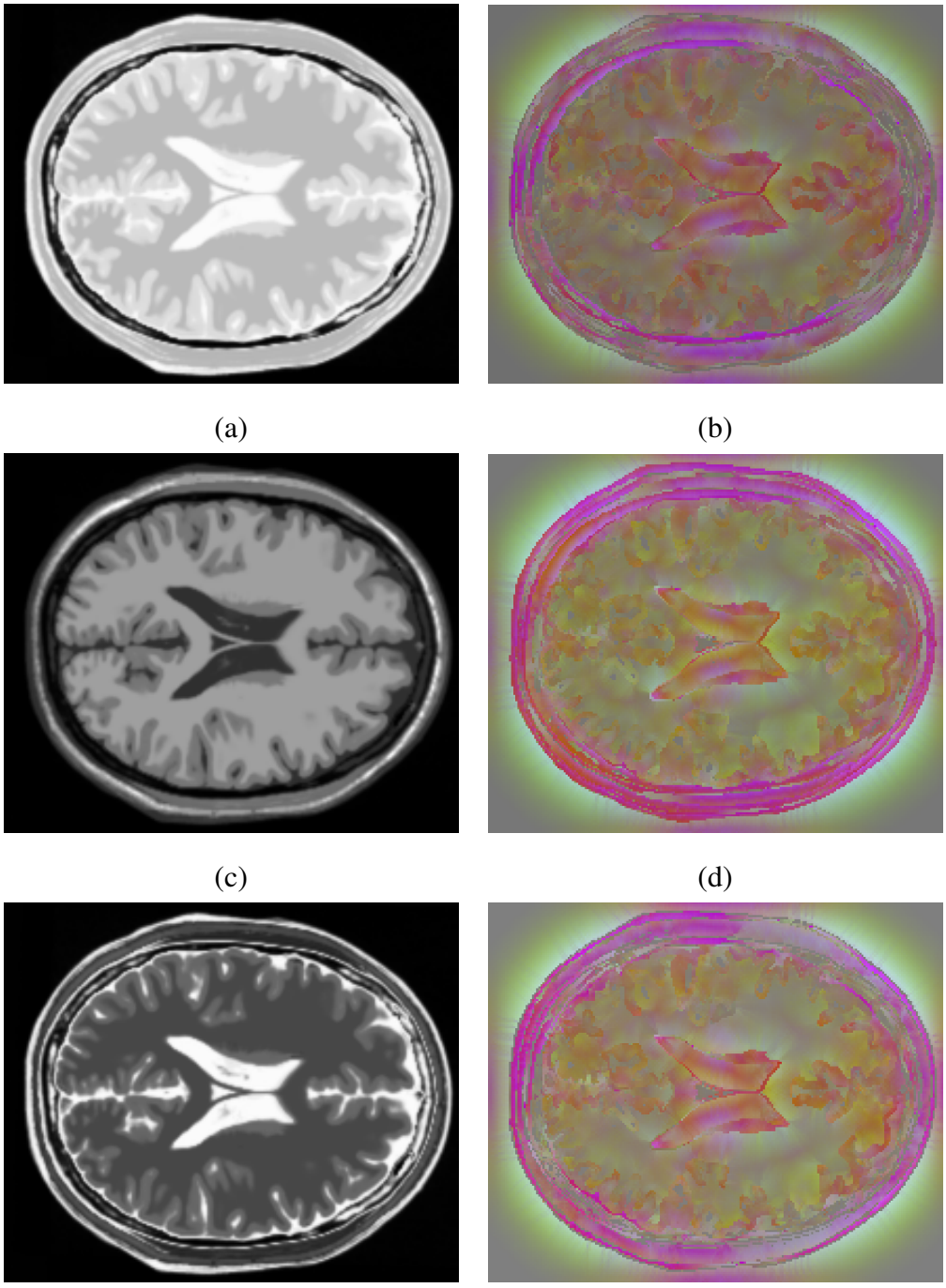
Improvement of Input Data

As noted in the Experiment 1, the pre-processing step for the binary map was to use a watershed algorithm to separate large bodies (which were too large to be a single cell) into smaller bodies. Simply by adding this pre-processing step, the accuracy gained a consistent 3%. Looking at some of the results from the pre-processing stage, the watershed often times over segments regions which is likely negatively affecting the classifier. While there are whole bodies of works, indeed a whole field of research, related to the identification and isolation of individual cells, the application and development of an approach suited to the high throughput nature of the LMS algorithm is outside the scope of this document.

Applications to Registration

Since LMS creates a unique signature, at each pixel, which encodes local morphological information, we believe that the LMS could see applications in registration. As preliminary evidence supporting this claim, we have taken a single image from each protocol (PD, T1, and T2) of a phantom brain MRI image [76] and computed the LMS signature at each pixel. Afterwards, we projected the high dimensional LMS signature into the 3 dimensional RGB space and display the results. We can see from the images presented in Figured 8.1 that even though the chromatic

values in the MRI phantom images are significantly different, when looking at the LMS representation there does seem to be a consistency across the protocols. We believe that a registration algorithm would gain performance by using this novel feature space.



(a) (b)

(c) (d)

(e) (f)

Figure 8.1: PD (a), T1 (c) and T2 (e) phantom brain MRI images. If we compute the LMS for each point and project the signature into a 3 dimensional RGB space, we can see their respective visual representations in (b), (d) and (f), respectively. We propose that since visually they appear similar, despite being vastly different in their original space, the LMS signature could have applications in registration.

Bibliography

[1] C. M. Tipton. Sushruta of India, an unrecognized contributor to the history of exercise physiology. *Journal of Applied Physiology*, 104(6):1553–1556, June 2008.

[2] R. Jain, S. Kosta, and A. Tiwari. Ayurveda and cancer. *Pharmacognosy Research*, 2(6):393–394, November 2010.

[3] R. Buckanovich, D. Sasaroli, A. O'Brien-Jenkins, et al. Tumor vascular proteins as biomarkers in ovarian cancer. *Journal Of Clinical Oncology*, 25(7):852–861, March 2007.

[4] Daniela Burgos-Ojeda, Karen McLean, Shoumei Bai, Heather Pulaski, Yusong Gong, Ines Silva, Karl Skorecki, Maty Tzukerman, and Ronald J. Buckanovich. A novel model for evaluating therapies targeting human tumor vasculature and human cancer stem-like cells. *Cancer Res*, 73(12):3555–3565, Jun 2013.

[5] Dimitra Sasaroli, Phyllis A. Gimotty, Harsh B. Pathak, Rachel Hammond, Eleni Kougioumtzidou, Dionyssios Katsaros, Ron Buckanovich, Karthik Devarajan, Raphael Sandaltzopoulos, Andrew K. Godwin, Nathalie Scholler, and George Coukos. Novel surface targets and serum biomarkers from the ovarian cancer vasculature. *Cancer Biol Ther*, 12(3):169–180, Aug 2011.

[6] A. Markowska, J. Lubin, R. Madry, and J. Markowska. Development of antiangiogenic therapies for ovarian cancer. *Eur J Gynaecol Oncol*, 34(4):303–306, 2013.

[7] Limei Wang, Xiaoyan Liu, Hong Wang, and Shuhe Wang. Correlation of the expression of vascular endothelial growth factor and its receptors with microvessel density in ovarian cancer. *Oncol Lett*, 6(1):175–180, Jul 2013.

[8] E. Sato, S. Olson, J. Ahn, et al. Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory t cell ratio are associated with favorable prognosis in ovarian cancer. *Proceedings of the National Academy of Sciences*, 102(51):18538–43, October 2005.

[9] B. Clarke, A. Tinker, C. Lee, et al. Intraepithelial T–cells and prognosis in ovarian carcinoma: novel associations with stage, tumor type, and BRCA1 loss. *Modern Pathology*, 22(3):393–402, March 2009.

[10] L. Zhang, J. Conejo-Garcia, D. Katsaros, et al. Intratumoral T–cells, recurrence, and survival in epithelial ovarian cancer. *The New England Journal of Medicine*, 348(3):203–13, January 2003.

[11] Katy Milne, Cheryl Alexander, John R. Webb, Winnie Sun, Kristy Dillon, Steve E. Kalloger, C Blake Gilks, Blaise Clarke, Martin Kbel, and Brad H. Nelson. Absolute lymphocyte count is associated with survival in ovarian cancer independent of tumor-infiltrating lymphocytes. *J Transl Med*, 10:33, 2012.

[12] John R. Webb, Katy Milne, Peter H. Watson, Ron J. Deleeuw, and Brad H. Nelson. Tumor-infiltrating lymphocytes expressing the tissue resident memory marker cd103 are associated with increased survival in high-grade serous ovarian cancer. *Clin Cancer Res*, Nov 2013.

[13] Anna Bachmayr-Heyda, Stefanie Aust, Georg Heinze, Stephan Polterauer, Christoph Grimm, Elena Ioana Braicu, Jalid Sehouli, Sandrina Lambrechts, Ignace Vergote, Sven Mahner, Dietmar Pils, Eva Schuster, Theresia Thalhammer, Reinhard Horvat, Carsten Denkert, Robert Zeillinger, and Dan Cacsire Castillo-Tong. Prognostic impact of tumor infiltrating cd8+ t cells in association with cell proliferation in ovarian cancer patients - a study of the ovcad consortium. *BMC Cancer*, 13:422, 2013.

[14] Michael J. Bradaric, Krishna Penumatsa, Animesh Barua, Seby L. Edassery, Yi Yu, Jacques S. Abramowicz, Janice M. Bahr, and Judith L. Luborsky. Immune cells in the normal ovary and spontaneous ovarian tumors in the laying hen (gallus domesticus) model of human ovarian cancer. *PLoS One*, 8(9):e74147, 2013.

[15] Ryan O. Emerson, Anna M. Sherwood, Mark J. Rieder, Jamie Guenthoer, David W. Williamson, Christopher S. Carlson, Charles W. Drescher, Muneesh Tewari, Jason H.

Bielas, and Harlan S. Robins. High-throughput sequencing of t cell receptors reveals a homogeneous repertoire of tumor-infiltrating lymphocytes in ovarian cancer. *J Pathol*, Sep 2013.

[16] Greg T. Motz and George Coukos. Deciphering and reversing tumor immune suppression. *Immunity*, 39(1):61–73, Jul 2013.

[17] Jiabo Di, Leon F A G. Massuger, Tjitske Duiveman-de Boer, Petra L M. Zusterzeel, Carl G. Figdor, and Ruurd Torensma. Functional oct4-specific cd4(+) and cd8(+) t cells in healthy controls and ovarian cancer patients. *Oncoimmunology*, 2(5):e24271, May 2013.

[18] H. Vrolijk, W. Sloos, W. Mesker, et al. Automated acquisition of stained tissue microarrays for high-throughput evaluation of molecular targets. *Journal Of Molecular Diagnostics*, 5(3):160–167, August 2003.

[19] J. Wu, J. Dong, and H. Zhou. Image quantification of high-throughput tissue microarray. *Society of Photo-Optical Instrumentation Engineers*, 6143:509–520, March 2006.

[20] A. Rabinovich, S. Krajewski, M. Krajewska, et al. Framework for parsing, visualizing and scoring tissue microarray images. *IEEE Transactions on Information Technology in Biomedicine*, 10(2):209–219, April 2006.

[21] W. Zhong, G. Altun, R. Harrison, et al. Improved k-means clustering algorithm for exploring local protein sequence motifs representing common structural property. *IEEE Transactions on NanoBioscience*, 4(3):255–265, September 2005.

[22] H. Fatakdawala, J. Xu, A. Basavanhally, et al. Expectation maximization driven geodesic active contour with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*, 57(7):1676–1689, July 2010.

[23] L. Cohen and R. Kimmel. Global minimum for active contour models: A minimal path approach. *IEEE International Journal on Computer Vision*, 24(1):57–78, August 1997.

[24] Z. Tu. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. *IEEE International Conference on Computer Vision*, 2:1589–1596, October 2005.

[25] P. Tiwari, M. Rosen, G. Reed, and et al. Spectral embedding based probabilistic boosting tree (ScEPTre): Classifying high dimensional heterogeneous biomedical data. *Medical Image Computing and Computer Assisted Intervention*, 1:844–851, 2009.

[26] B. Carneiro, G.and Georgescu, S. Good, and D. Comaniciu. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Transactions on Medical Imaging*, 27(9):1342–1355, September 2008.

[27] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, August 2000.

[28] V. Vazirani. *Approximation Algorithms*. Springer, March 2004.

[29] M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[30] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1101–1113, November 1993.

[31] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:359–374, September 2001.

[32] S. Dhillon, Y. Guan, and B. Kulis. Weighted graph cuts without eigenvectors: A multilevel approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1944–1957, November 2007.

[33] S. Chandran, S. Hebbar, V. Mamania, and A. Sawa. Improved cut-based foreground identification. In *The Indian Conference on Computer Vision, Graphics and Image Processing*, pages 447–454, 2004.

[34] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, January 1975.

[35] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002.

[36] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, August 1995.

[37] C. Yang, R. Duraiswami, N. Gumerov, and L. Davis. Improved Fast Gauss transform and efficient kernel density estimation. *IEEE International Conference on Computer Vision*, 1:664–671, October 2003.

[38] W. Tao, H. Jin, and Y. Zhang. Color image segmentation based on mean shift and normalized cuts. *IEEE Transactions on Systems, Man, and Cybernetics*, 37(5):1382–1389, 2007.

[39] X. Yuan, B. Hu, and R. He. Agglomerative mean shift clustering. *IEEE Transactions on Knowledge and Data Engineering*, 24(2):209–219, March 2012.

[40] A. Witkin. Scale-space filtering. In *International Joint Conference on Artificial Intelligence*, pages 1019–1022, 1983.

[41] P. Burt. Fast filter transforms for image processing. *Computuer Graphics Image Processing*, 16:25–51, 1981.

[42] Scott Doyle, Carlos Rodriguez, Anant Madabhushi, John Tomaszeweski, and Michael Feldman. Detecting prostatic adenocarcinoma from digitized histology using a multi-scale hierarchical classification approach. *Conf Proc IEEE Eng Med Biol Soc*, 1:4759–4762, 2006.

[43] T. Lindeberg. Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(1):225–270, 1994.

[44] S. Pizer, D. Eberly, B. Morse, and D. Fritsch. Zoom-invariant vision of figural shape: the mathematics of cores. *Computer Vision and Image Understanding*, 69(1):55–71, January 1998.

[45] A. Madabhushi and J. Udupa. New methods of MR image intensity standardization via generalized scale. *Medical Physics*, 33(9):3426–34, September 2006.

[46] A. Madabhushi, J. Udupa, and A. Souza. Generalized scale: theory, algorithms, and application to image inhomogeneity correction. *Computer Vision and Image Understanding*, 101(2):100–121, October 2006.

[47] N. László, J. Udupa, and P. Saha. Incorporating a measure of local scale in voxel-based 3D image registration. *IEEE Transactions on Medical Imaging*, 22(2):228–237, February 2003.

[48] I. Hontsch and L. Karam. Locally adaptive perceptual image coding. *IEEE Transactions on Image Processing*, 9(9):1472–1483, September 2000.

[49] P. Saha, J. Udupa, and D. Odhner. Scale-based fuzzy connected image segmentation: theory, algorithms, and validation. *Computer Vision and Image Understanding*, 77(2):145–174, February 2000.

[50] P. Saha. Tensor scale: a local morphometric parameter with applications to computer vision and image processing. *Computer Vision and Image Understanding*, 99(3):384–413, September 2005.

[51] Bernd Lahrmann, Niels Halama, Hans-Peter Sinn, Peter Schirmacher, Dirk Jaeger, and Niels Grabe. Automatic tumor-stroma separation in fluorescence tmas enables the quantitative high-throughput analysis of multiple cancer biomarkers. *PLoS ONE*, 6(12):e28048, December 2011.

[52] K. Mosaliganti, F. Janoos, O. Irfanoglu, et al. Tensor classification of N-point correlation function features for histology tissue segmentation. *Medical Image Analysis*, 13(1):156–166, February 2009.

[53] Andrew H. Beck, Ankur R. Sangoi, Samuel Leung, Robert J. Marinelli, Torsten O. Nielsen, Marc J. van de Vijver, Robert B. West, Matt van de Rijn, and Daphne Koller. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science Translational Medicine*, 3(108):108ra113, 2011.

[54] University Oxford. *Oxford Dictionary of Biochemistry and Molecular Biology*. Oxford University Press, USA, revised edition, February 1997.

[55] H. Rui and M.J. LeBaron. Creating tissue microarrays by cutting-edge matrix assembly. *Expert Review Medical Devices*, 2(6):673–680, November 2005.

[56] B. Silverman. *Density Estimation ,for Statistics and Data Analysis*. London: Chapman and Hall, 1986.

[57] A. Zomaya. *Parallel and Distributed Computing Handbook*. McGraw-Hill, Inc., New York, NY, USA, 1996.

[58] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

[59] S. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man and Cybernetics*, 21(3):660–674, May 1991.

[60] S. Lloyd. Least square quantization in PCM. *IEEE Transactions on Information Theory*, 28:129–137, March 1982.

[61] J. Johannessen and M. Sobrinho-Simoes. The origin and significance of thyroid psammoma bodies. *Laboratory Investigation*, 43:287–296, September 1980.

[62] J. Kim, G. Beets, M. Kim, et al. High-resolution MR imaging for nodal staging in rectal cancer: are there any criteria in addition to the size? *European Journal of Radiology*, 52(1):78–83, October 2004.

[63] A. Madabhushi, S. Doyle, G. Lee, et al. Integrated diagnostics: a conceptual framework with examples. *Clinical Chemistry and Laboratory Medicine*, 48(7):989–98, July 2010.

[64] A. Madabhushi, S. Agner, A. Basavanhally, et al. Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data. *Computerized Medical Imaging and Graphics*, 35:506–514, October 2011.

[65] D. Gleason. Classification of prostatic carcinomas. *Cancer Chemotherapy Reports*, 1(50):125–128, March 1966.

[66] Sahirzeeshan Ali and Anant Madabhushi. An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Trans Med Imaging*, 31(7):1448–1460, July 2012.

[67] Sahirzeeshan Ali and Anant Madabhushi. Graphical processing unit implementation of an integrated shape-based active contour: Application to digital pathology. *J Pathol Inform*, 2:S13, 2011.

[68] G. Granlund. Fourier preprocessing for hand print character recognition. *IEEE Transactions on Computers*, 21(2):195–201, February 1972.

[69] A. Janowczyk, S. Chandran, R. Singh, et al. Hierarchical normalized cuts: Unsupervised segmentation of vascular biomarkers from ovarian cancer tissue microarrays. In *Medical Image Computing and Computer Assisted Intervention Society*, pages 230–238, 2009.

[70] C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, January 2001.

[71] E. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[72] J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences of the United States of America*, 93(4):1591–1595, February 1996.

[73] G. Seber. *Multivariate Observations*. Wiley, New York, 1984.

[74] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, November 1973.

[75] F. Meyer. Topographic distance and watershed lines. *Signal Processing*, 38:113–125, July 1994.

[76] D. Collins, A. Zijdenbos, V. Kollokian, et al. Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17:463–486, June 1998.

Publications

Thesis Related

1. Andrew Janowczyk, Sharat Chandran, Rajendra Singh, Dimitra Sasaroli, George Coukos, Michael D. Feldman, and Anant Madabhushi. Hierarchical normalized cuts: Unsupervised segmentation of vascular biomarkers from ovarian cancer tissue microarrays. In MICCAI (1), pages 230-238, 2009. *Runner-Up Young Scientist Award*

2. Jun Xu, Andrew Janowczyk, Sharat Chandran, and Anant Madabhushi. A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: applications to histopathology image segmentation. In SPIE Medical Imaging, 7623:76230Y, 2010.

3. Jun Xu, Rachel Sparks, Andrew Janowczyk, John E. Tomaszeweski, Michael D. Feldman, and Anant Madabhushi. High-throughput prostate cancer gland detection, segmentation, and classification from digitized needle core biopsies. In Prostate Cancer Imaging, pages 77-88, 2010.

4. Jun Xu, Andrew Janowczyk, Sharat Chandran, and Anant Madabhushi. A high-throughput active contour scheme for segmentation of histopathological imagery. Medical Image Analysis, 15(6):851-862, 2011.

5. Andrew Janowczyk, Sharat Chandran, Michael D. Feldman, and Anant Madabhushi. Local morphologic scale: application to segmenting tumor infiltrating lymphocytes in ovarian cancer TMAs. In SPIE Medical Imaging, 7962:79622N, 2011.

6. Andrew Janowczyk, Sharat Chandran, Rajendra Singh, Dimitra Sasaroli, George Coukos, Michael D. Feldman, and Anant Madabhushi. High-throughput biomarker segmentation

on ovarian cancer tissue microarrays via hierarchical normalized cuts. In IEEE Transactions on Bio-Medical Engineering, 59(5):1240–52, 2012.

7. Andrew Janowczyk, Sharat Chandran, and Anant Madabhushi. Quantifying local heterogeneity via morphologic scale: Distinguishing tumor from stroma. HIMA Workshop MICCAI, 2012.

8. Andrew Janowczyk, Sharat Chandran, and Anant Madabhushi. Quantifying local heterogeneity via morphologic scale: Distinguishing tumoral from stromal regions. Journal of Pathology Informatics, 2013 (In press).

Non-Thesis Related

1. Andrew Janowczyk, Sharat Chandran, and Srinivas Aluru. Fast, processor cardinality agnostic PRNG with a tracking application. In ICVGIP, pages 171-178, 2008. *Best Poster Presentation Award*

Acknowledgements

In many ways, the completion of a Ph.D. is not the end of a journey, but a certification that one is qualified to embark on one. That journey for me began with my sister's diagnosis of Hodgkin's lymphoma. Looking back, choosing to embark on this path was my way of dealing with her own journey; thankfully it was never a process I had to undertake on my own. As a result, this section gives me a chance to acknowledge the people who have contributed to keeping me focused and concentrated on all of the aspects of life, both academic and human, so that I would have some type of balance. A good story follows a time line, and so does the following list:

- When I was considering endeavoring into the field of cancer imaging research, I was employed by an eccentric Scotsman named *John Cameron*. He was the first person that enabled me by encouraging me to work remotely from IIT while starting on my new academic path.

- In the first few weeks at IIT, I had to find an adviser. *Prof. Sharat Chandran*'s kind demeanor and experience made him an excellent choice. I can't thank him enough for his never ending advice, patience and benevolence towards me as I failed, regrouped and failed again.

- During the first few months, I had the chance to meet many friends at IIT who are still present in my life. *Adrien Bock, Biswarup Choudhury, Rhushabh Goradia, Aparna Majumdar, Anne Monnier, Appu Shaji* provided support and distractions, wisdom and silliness, experience and curiosity, catharsis and duress, water and wine. They made the day to day of living away from home burdenless.

- When it came time to delve into the broad field of cancer research, *Prof. Chandran* introduced me to my now co-advisor *Anant Madabhushi*. The combination of his demand of nothing but my best and the freedom to solve problems in a way that made sense to

me facilitated the creation of these two algorithms. His harsh editor's pen made sure that their ideas were presented accurately and clearly to the community.

- Occasionally in life there are unseen road blocks preventing progress towards one's goals. A conversation with *Stephane Gamard* at *Le Palace de Jade* resulted in the seed which grew into the conversion of the LMS signature to its 1D representation, motivating the usage of the Fourier descriptors. Without that conversation, there is the possibility that I may still be sitting in-front of Matlab trying to find a suitable quantification algorithm.

- My colleagues at the Center for Computational Imaging and Personalized Diagnostics, located at Case Western Reserve University, provided valuable insights, experience and a deeply shared passion in the domain of cancer research. These top caliber researchers incubated and help grow the ideas in this thesis, allowing them to reach their full potential.

- Throughout this adventure $\{Amanda, Janet, Michael\}Janowczyk$ and *Crystie Doell* supported me across thousands of miles and hundreds of hours, though my ambitions always seem to keep me at some distance from them (either physically or temporally); their unwavering encouragement sees me to completion.

Date: _____ Andrew Janowczyk