

# **Automatic Detection and Classification of Tumor Infiltrating Lymphocytes**

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

Doctor of Philosophy

by

**Andrew Janowczyk**  
**(Roll No. 08405201)**

Under the guidance of  
**Prof Sharat Chandran**



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

2013

# Introduction

The latest global statistics from 2008 show that 12.7 million new cancer cases and 7.6 million cancer deaths occurred worldwide<sup>1</sup>. 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 re-occurrence likelihood is becoming a reality.

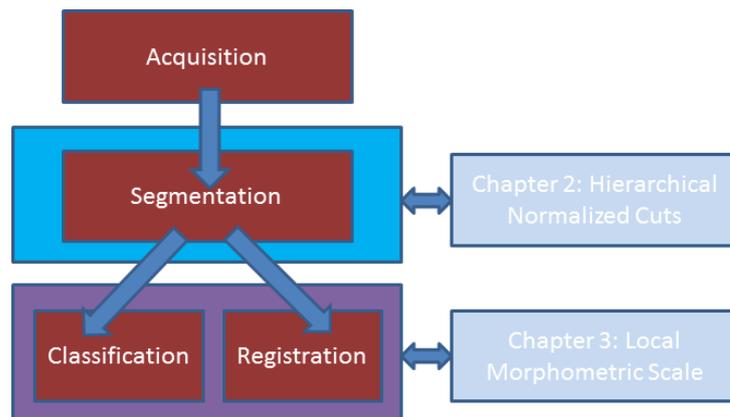


Figure 1: Three major tasks associated with computer aided diagnostic include acquisition of data, segmentation of regions of interest (ROI), and lastly the classification and registration of these ROI. The thesis makes notable contributions in segmentation (second chapter) and classification (third chapter), with a potential extension to registration .

Figure 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 are 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 understood via the desire to align the current region to an existing model, perhaps to identify anomalies in shape or location of critical organs.

<sup>1</sup>World Health Organization, GLOBOCAN 2008

## Problem Scope

As shown in Figure 1, this thesis aims to make contributions in segmentation, classification, and registration via two state of the art methods.

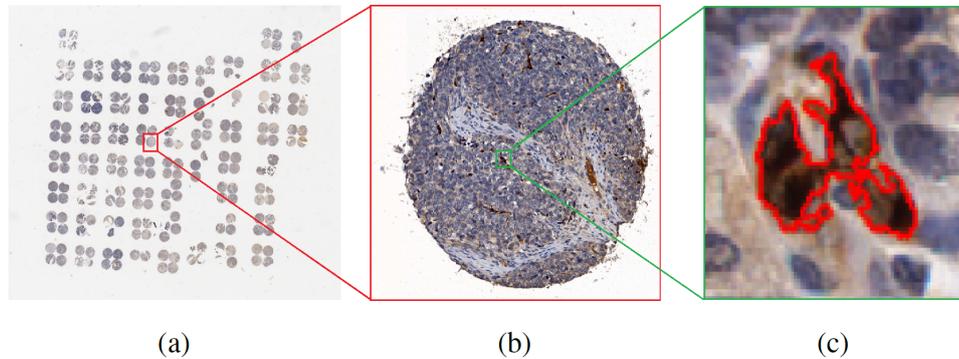


Figure 2: (a) A tissue micro array and (b) a representative magnified tissue cylinder culled out from (a) with the extracted biomarker presented in (c) delineated in red. A typical microarray could contain over 500 individual cylinders, making the biomarker detection via traditional image analysis algorithms a challenge.

In the domain of segmentation, we aim to take a histology image and extract the stained regions as demonstrated in Figure 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 environmental variables. The confidence necessary to provide an industry standard approach is non-trivial as these variances, along with a wide range of user defined inputs parameters, create a complex problem domain.

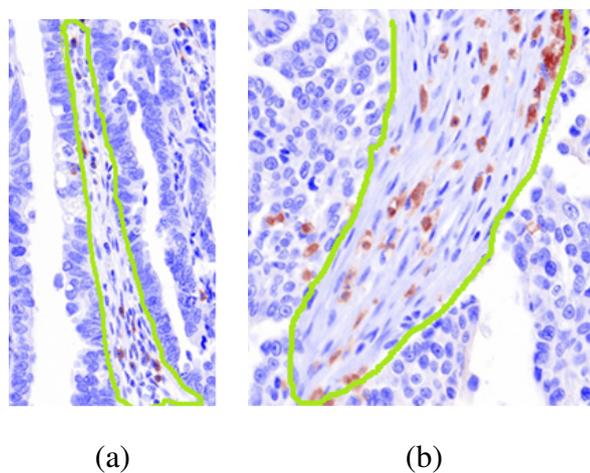


Figure 3: Stroma region manually circled in green. Although they are of the same class, notice the stark difference in size and shape between the two regions in (a) and (b). Selection of an appropriate window size or shape in a typical approach such as texture features is difficult.

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 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. We note, however, that the technique developed here is not specific to only tumor and stroma classification.

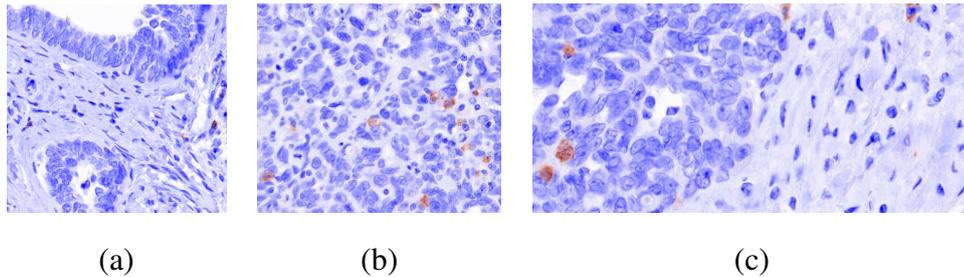


Figure 4: Lymphocytes in all panels are stained in red. All of the lymphocytes in (a) the stroma region image are non-TILs, while all of the lymphocytes in (b) the tumor region image are TILs. In (c) we see both TILs (left half) and non-TILs (right half).

The problem domain we have chosen is the identification of lymphocytes as either tumor infiltrating (TIL) or non-TIL, an example of which we can see in Figure 4. Our approach is to first use our Hierarchical Normalized Cuts (HNCut) to segment necessary information for our Local Morphometric Scale (LMS) algorithm to successfully classify pixels of interest as either tumor or stromal regions. This thesis contributes the necessary theory and implementation for an automated lymphocyte extraction and classification system via the chain of these two novel algorithms. We discuss our approaches, challenges, and contributions briefly in the following sections, followed by the layout of the thesis chapters.

## Segmentation Using Hierarchical Normalized Cuts

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 work [1] suggests 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.

## **Challenges and Novel Contributions to Segmentation**

The major contribution of this piece of the thesis is a fast, novel, hierarchical unsupervised segmentation method (HNCut), which we demonstrate with an application in Ovarian TMAs.

- 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.

## **Contributions**

- A tested hierarchical segmentation approach that marries an accelerated Mean Shift [2] method and the Normalized Cuts [3] method which we term Hierarchical Normalized Cuts (HNCut)[4] . A Matlab implementation of HNCut not only operates on large (1.5 million pixels) images in under 10 seconds, but is easily scalable to entire TMAs.
- Parameter insensitive segmentation [5] for large images and the ability of HNCut to discriminate between regions with similar color values. 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.

## Related work

Most previous computerized image analysis algorithms for TMAs have involved thresholding based schemes [6, 7, 8]. These methods are known to be highly sensitive to even slight changes in color and illumination. Clustering based approaches, including  $k$ -means [6], 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 [9]. Active contour schemes [10], 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 [11]. Supervised learning methods [12, 13] are constrained by the difficulty [14] in obtaining ground truth segmentations from experts for classifier training of the object of interest.

## Overview

Figure 5 presents a high level overview of the four stages associated with the HNCut algorithm.

We start by requiring the user to select a few sample pixels from the target class from an image. We use these pixels to guide the subsequent pixel classification process across all images in the same domain.

Next, we employ the mean-shift algorithm 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 5). Intuitively, our novel *Frequency Weighted Mean Shift* (FWMS) variant allows for identification of color values which are within some specified tolerance of each other and assigns them to the same mode. Employing the NCuts operation only on the unique values at each level of the pyramid, as opposed to all possible color values, allows for a summarization resulting in significantly fewer computations.

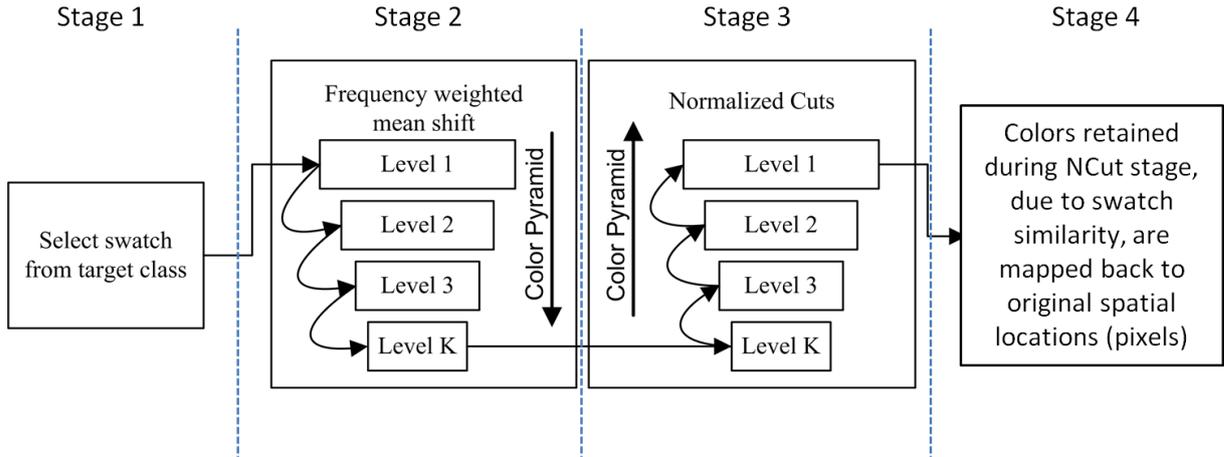


Figure 5: A flow chart of the HNCut process. Proceeding left to right, the user selects the domain swatch, followed by the Frequency Weighted Mean Shift of the image. This results in the original image being decomposed into multiple levels of color resolution, which is then followed by the application of NCut 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 NCut 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.

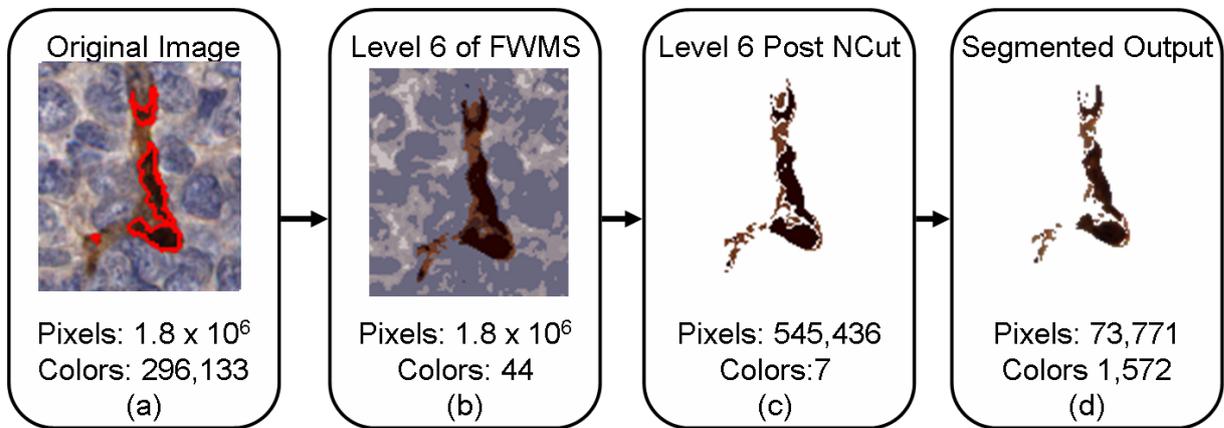


Figure 6: (a) Original image with desired TVM stain enclosed in red, (b) image at the bottom of the color pyramid during FWMS, (c) image at the bottom of the color pyramid following application of NCuts, (d) final segmentation results obtained by mapping colors not eliminated by HNCut spatially onto 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

5), which we show can be solved far more efficiently by NCut. 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 6 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 6 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.

## Results

As an example of the value of HNCut, we present in Figure 7 a comparison against a classical  $k$ -means approach and a powerful supervised classifier termed Probabilistic Boosting Trees [12]. As we can see, in both accuracy and computation time, HNCut is a superior algorithm at this segmentation task.

## Local Region Classification

Another challenging task in the field of digital pathology 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. A lymphocyte is a type of white blood cell that is sent to the proximity of objects which the body considers foreign. Recent work [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

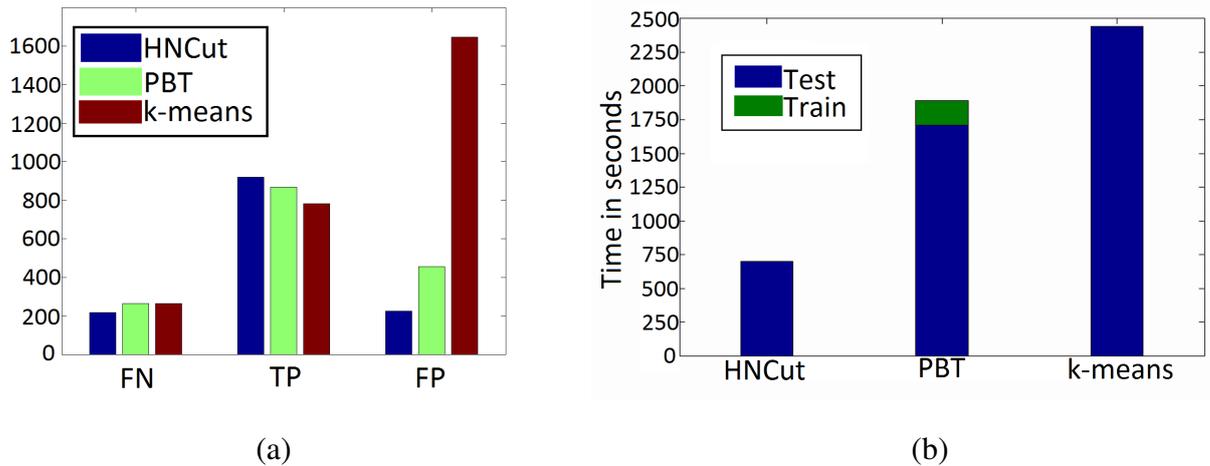


Figure 7: (a) False Negatives (FN), True Positives (TP) and False positives (FP) presented for the three different algorithms. From this visualization it is apparent that HNCut outperforms each of the other algorithms. (b) A comparison of the run times of the different algorithms. We can see that HNCut performs the fastest out of those tested, motivating its high throughput capabilities. Interestingly, the training time of PBT is long enough such that in a race HNCut would have already completed 25% of the segmentation task before PBT had begun its testing phase.

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.

## Challenges and Novel Contributions to Classification

The major contribution of this piece of the thesis is to develop a system which can accurately classify pixels as being embedded in either tumor or stromal regions, and thus extendable to classifying lymphocytes as either TILs or non-TILs.

- 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 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 the size of the associated region difficult to pre-determine.

## Contributions

- A novel signature definition, which we term Local Morphometric Scale (LMS)[18], allows for quantitatively characterizing local heterogeneity . This is especially relevant in the context of histopathology which consists of notoriously heterogeneous images.
- The LMS yields a 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.
- The algorithm is well suited for GPU computing, allowing for a very high throughput.
- We develop a novel approach to the very important problem of separating out tumoral from stromal regions via application of LMS.

## Related Work

The desire to differentiate tumor and stromal regions has been addressed recently using different approaches. In [19] an attempt at a shotgun approach using over 6000 well known features obtained only a 89% accuracy. Their novelty was defined by their successful combination of existing features. Our work presents a significantly lower dimensional novel morphometric feature set, which obtains 88% accuracy on the same task, clearly indicating a competitive approach. In [20], the authors present an approach which requires specially stained fluorescence images, while ours uses industry standard H&E, allowing broader usage in pre-existing tissue repositories. Lastly, the tensor classification using  $N$ -point correlation functions presented in [21], while notable, is a non-scalable approach which is computationally expensive.

On the theory front, the notion of locally adaptive scale is also not new and has already seen applications in a variety of image processing tasks including MRI bias field correction[22], image segmentation[23], image registration [24], and image coding [25]. Saha and Udupa introduced the notion of ball-scale[26] which at every spatial location was defined as the value

corresponding to the radius of the largest ball encompassing all locations neighboring the location under consideration and satisfying some pre-defined homogeneity criterion. Tensor-scale ( $t$ -scale) [27] was later 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 Madabhushi and Udupa with the introduction of Generalized scale ( $g$ -scale)[23].  $G$ -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. It is noteworthy to notice that all of the previous scale definitions have been inhibited by either shape constraints or homogeneity constraints. LMS is able to easily distinguish itself as a shape-free, heterogeneous modeling technique.

## Overview

Figure 8 presents an overview of the LMS creation processing as it applies to the problem of tumor versus stromal differentiation. These steps are described below in the context of lymphocyte classification.

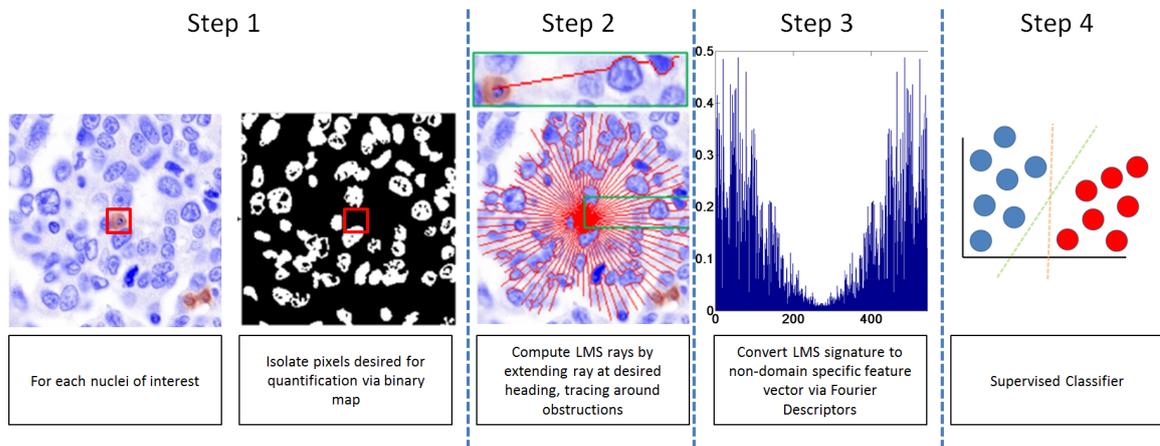


Figure 8: Overview of the LMS signature creation process. We can get the intuition that the more heterogeneous an area is the higher number of deviations from the straight line trajectory occur, on account of the rays attempting to take the path of least resistance and hence overcome obstacles along the way. On the other hand, the LMS signature will be smoother as a result of comprising fewer and smaller objects in homogeneous areas.

**Step 1:** Identify centers of interest by extracting the stained lymphocytes using HNCut. Also produce the resulting binarized images by extracting, again with HNCut, the blue stain which highlights the endothelial and tumor cells. Binarized images indicate which pixels which will be incorporated in the morphologic signature of the POI.

**Step 2:** The LMS involves projecting connected paths, radially outwards from any POI (e.g. nucleus center identified in step 1), as shown in Step 2 of Figure 8. The green box is used to illustrate the path of a single ray more clearly.

---

**Algorithm 1** LMS Signature Creation

---

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

**Output:** A set of rays  $R(q)$

- 1:  $S = C_{-q_x, -q_y}$ , a transformation such that  $q$  is located at the origin
  - 2: **for**  $\theta = 0 : \frac{2\pi}{\epsilon} : 2\pi$  **do**
  - 3:  $\delta = (\cos(\theta) * w, \sin(\theta) * w)$  Identify location of desired end point
  - 4:  $R_\theta(q) = \min_m p_{q,\delta}, m = |p_{q,\delta}|, \forall g(p_{q,\delta}) = 0$  Compute path with the least deviation
  - 5: **end for**
  - 6:  $R(q) = \{R_\theta | \forall \theta\}$  LMS is the set of all of the individual rays
  - 7: **return**  $R(q)$
- 

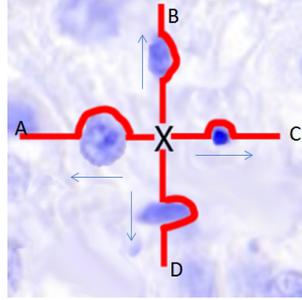


Figure 9: Point of interest is marked as  $X$  in the center of the image. We can see that 4 rays are projected from the center, and as they encounter obstructions in their path, they proceed around and return to the straight path as soon as possible.

We can see from the examples in Figure 10, the end result of our algorithm is set of connected pixels (shown in red), which travel from the POI,  $q$ , transformed to the center of the image, towards evenly spaced end points  $\delta$ . We determine the location of these end points  $\delta$  by casting them on a unit circle and multiplying by the window size to get the appropriate magnitude as seen in step 3 of 1. The path is a connected sequence of pixels,  $p_{q,\delta}$ , found by calculating the route with the least deviation from a straight line from  $q$  to  $\delta$  such that none of the pixels are foreground pixels (defined by  $g(c) = 1$ ). Essentially, this constraint results in the behavior that when an object in the foreground is encountered while traveling on a straight line from  $q$  to  $\delta$ , we avoid it and return to the straight path as soon as possible, as shown in Figure 9.

**Step 3:** The quantification of the average local topography of all these paths (via Fourier descriptors [28]) yields a measure of local heterogeneity.

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

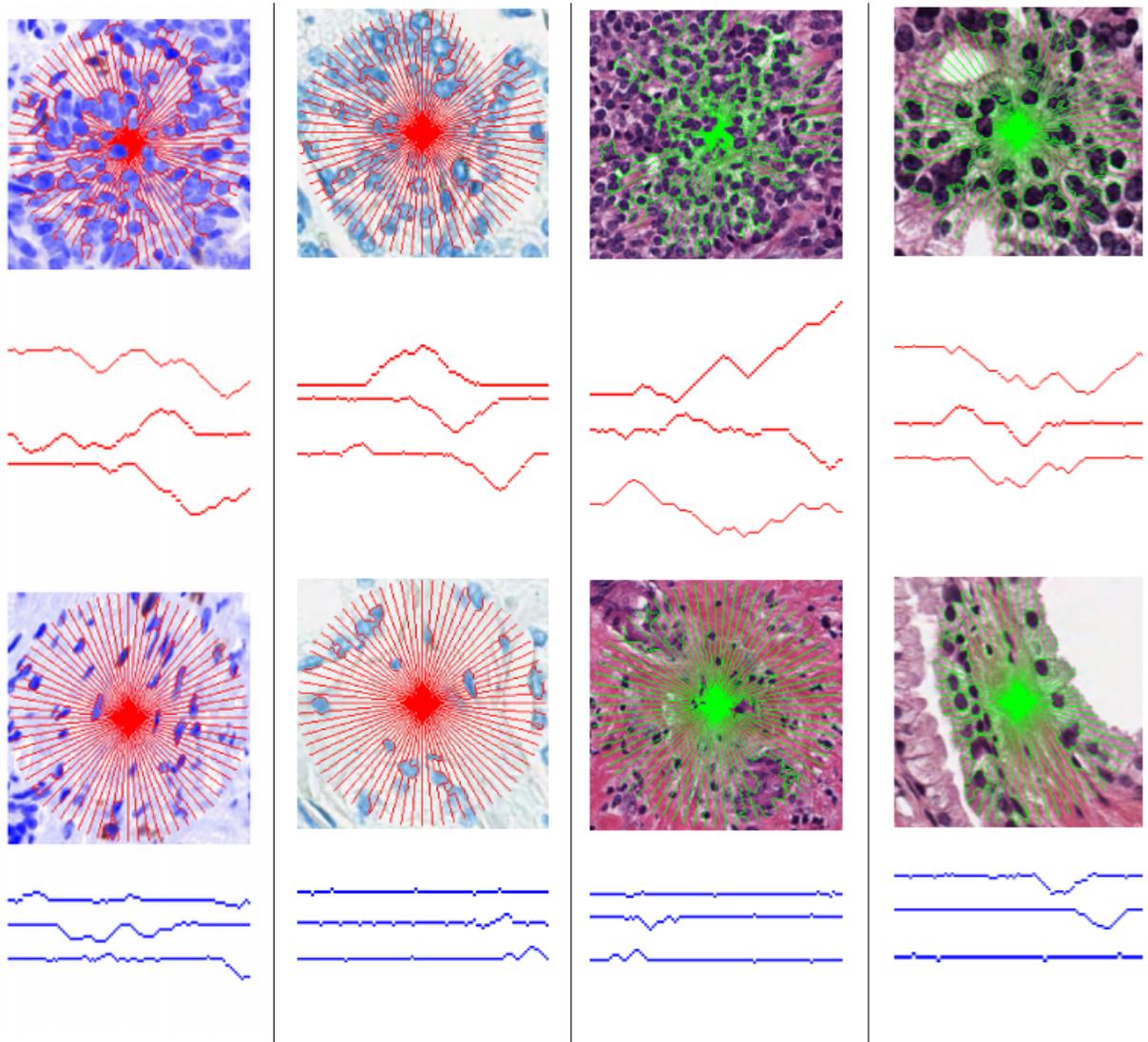


Figure 10: The LMS signature (in red or green) overlaid on a tumor regions (top) and benign regions (bottom) in an ovarian, prostate, breast, and prostate (different stain) images. Three rays are sampled from each image and presented beneath their respective images. We can see that in the non-tumor regions the LMS rays has fewer and smaller objects to obstruct its path, and thus the rays are less tortuous, unlike in the tumoral regions.

In Figure 10 homogeneous regions have few obstructions allowing the LMS rays to form straighter lines. On the other hand, as the complexity of the local region increases, a noted change in the LMS occurs. For very complex regions the LMS rays become increasingly tortuous as they adapt to the local heterogeneity. Quantifying this behavior as a non-domain specific feature set allows a supervised classifier to separate the two classes.

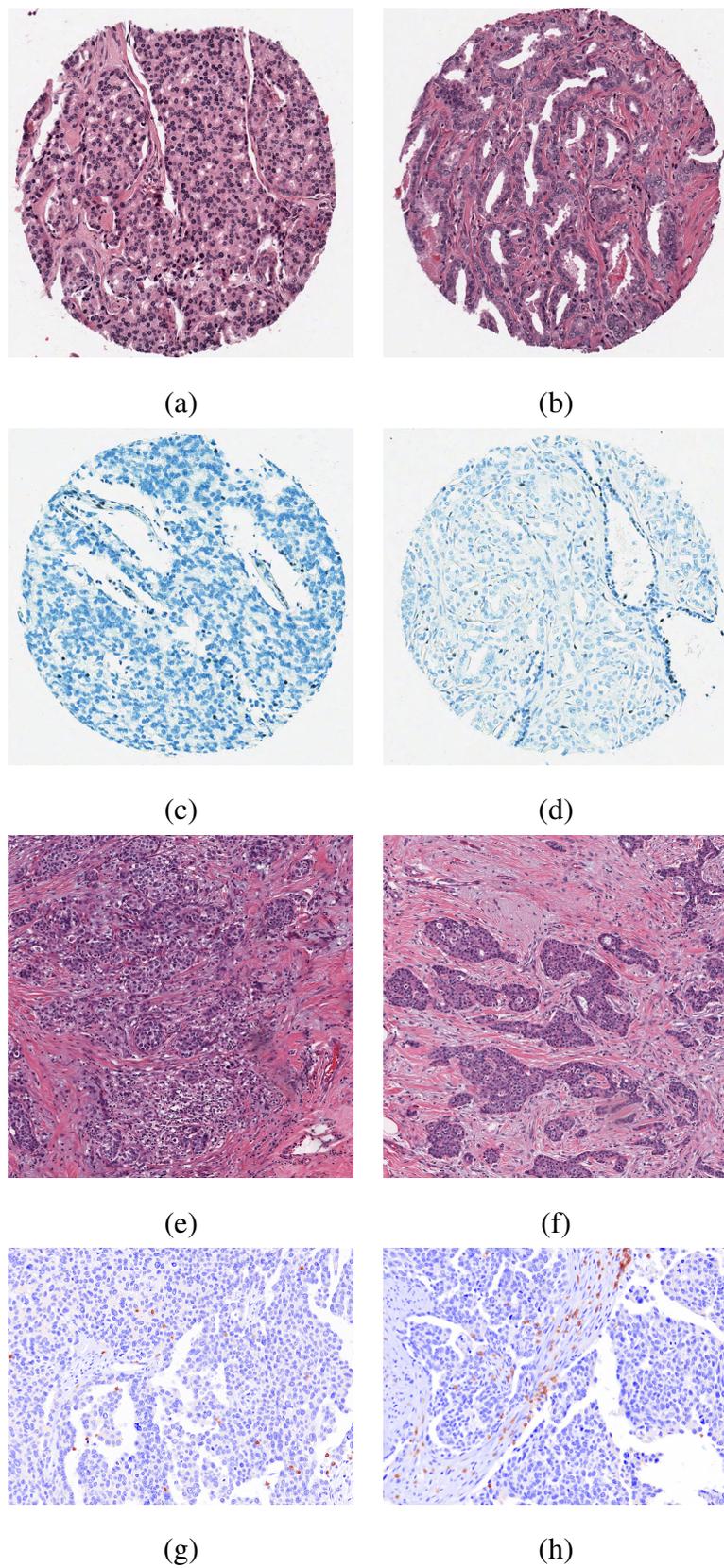


Figure 11: 2 sample images from each of the datasets listed in Table 1. First row is prostate HE, second row is prostate H and third row is Breast HE and last row is Ovarian H. We can see that each of these domains has its own unique characteristics, both of chromatic and morphological attributes, which makes a robust single approach challenging.

## Results

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

Table 1: Bayesian classifier area under the curve (AUC) in distinguishing stromal from tumoral pixels for four different domains, holding *all* parameters constant.

As a demonstration of the power of LMS, we test its ability to identify tumor versus stromal regions across 2 different types of stain, haematoxylin & eosin (HE) and solely haematoxylin(H), across 4 data domains: prostate HE, prostate H, breast HE and ovarian H. Sample images from each domain can be seen in Figure 11. For each domains, we randomly sampled pixels, separated them into non-overlapping training and test sets and used a naive Bayesian classifier to perform the classification. From the ROC curve, we computed the mean area under the curve (AUC) and associated range. We present the results in Table 1. What is especially interesting to note is that the parameters for each data domain were kept constant, which we feel indicates a strong non-domain specific feature set even though each domain appears noticeably different.

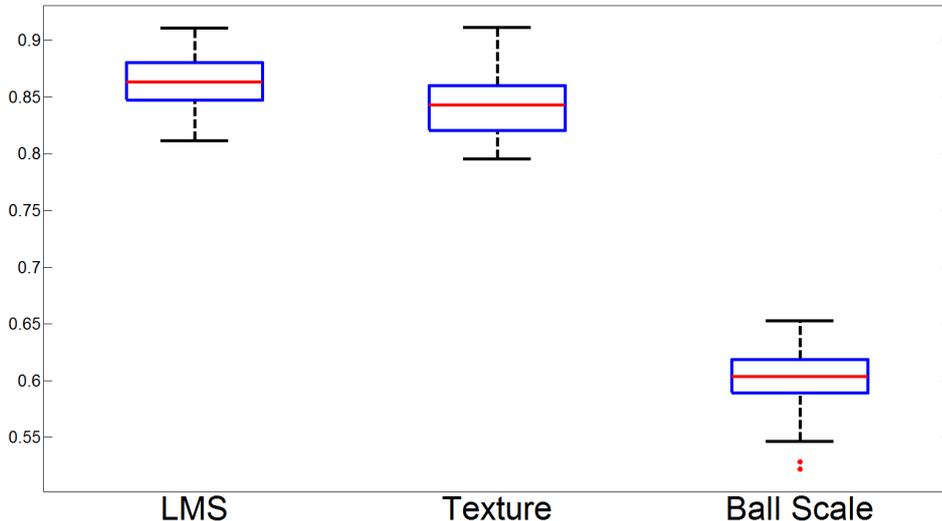


Figure 12: Box plots for the AUC across 50 runs from 3 algorithms: LMS, texture features and *b*-scale. The red line identifies the mean, the blue box encompasses 25th percentile, with the black whiskers extending to the 75th percentile. Red dots are indicative of outliers. We can see that the LMS provides a higher mean AUC than texture features with a smaller variance. On the other hand, ball scale seems to produce a poor classifier.

Also, using only the ovarian cancer domain, we compare our approach against two relevant approaches: texture features and  $b$ -scale. The box plots for the 3 approaches can be seen in Figure 12. We can see that with a mean 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 [19] with their self stated .88 accuracy. We draw attention to the significantly lower dimension of our approach (about 50) as compared to their 6,000 features. Lastly, we can see that homogeneity is not an ideal separating characteristic as  $b$ -scale fails rather poorly in this classification task, motivating the creation of our heterogeneous scale definition.

## Organization of thesis

The structure for the thesis is as follows:

- Chapter 1** In this chapter we have introduced Computer Aided Diagnostics and discuss the problem scope and challenges. We developed a high level overview of the work in this thesis, breaking it down into two parts, segmentation and classification. We explain 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 examining the weak and strong points of the various but similar approaches, our contribution become 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  $k$ -means and a supervised learning algorithm called 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, segmenta-

tion, classification or retrieval. This explanation includes the necessary theoretical background and algorithms and is further elucidated by a discussion on the various properties it posses.

**Chapter 6** The properties discussed in the previous chapter are fully vetted 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** We conclude the thesis in this chapter by summarizing the contributions made in this thesis. We also discuss the possible shortcomings of our techniques and also discuss ways to extend some of the ideas proposed in this thesis.

# 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. Tomaszewski, 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*

# Bibliography

- [1] R. Buckanovich, D. Sasaroli, and et al. Tumor vascular proteins as biomarkers in ovarian cancer. *Clinical Oncology, Journal Of*, 25(7):852–861, March 2007.
- [2] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, May 2002.
- [3] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, Aug. 2000.
- [4] A. Janowczyk, S. Chandran, R. Singh, D. Sasaroli, G. Coukos, M. Feldman, and A. Madabhushi. Hierarchical normalized cuts: Unsupervised segmentation of vascular biomarkers from ovarian cancer tissue microarrays. In *MICCAI*, pages 230–238, 2009.
- [5] A. Janowczyk, S. Chandran, R. Singh, D. Sasaroli, G. Coukos, M. Feldman, and A. Madabhushi. Hierarchical normalized cuts: Unsupervised segmentation of vascular biomarkers from ovarian cancer tissue microarrays. *Medical Image Computing and Computer Assisted Intervention*, 12(1):230–238, 2009.
- [6] H. Vrolijk, W. Sloos, W. Mesker, P. Franken, R. Fodde, H. Morreau, and H. Tanke. Automated acquisition of stained tissue microarrays for high-throughput evaluation of molecular targets. *Journal Of Molecular Diagnostics*, 5(3):160–167, Aug. 2003.
- [7] J. Wu, J. Dong, and H. Zhou. Image quantification of high-throughput tissue microarray. In A. Manduca and A. A. Amini, editors, *Society of Photo-Optical Instrumentation Engineers*, volume 6143, pages 509–520, March 2006.

- [8] A. Rabinovich, S. Krajewski, M. Krajewska, A. Shabaik, S. Hewitt, S. Belongie, J. Reed, and J. Price. Framework for parsing, visualizing and scoring tissue microarray images. *Information Technology in Biomedicine, IEEE Transactions on*, 10(2):209–219, April 2006.
- [9] W. Zhong, G. Altun, R. Harrison, P.C. Tai, and Y. Pan. Improved k-means clustering algorithm for exploring local protein sequence motifs representing common structural property. *NanoBioscience, IEEE Transactions on*, 4(3):255–265, Sept. 2005.
- [10] H. Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, M. Feldman, J. Tomaszewski, and A. Madabhushi. Expectation maximization driven geodesic active contour with overlap resolution (emagacor): Application to lymphocyte segmentation on breast cancer histopathology. *Biomedical Engineering, IEEE Transactions on*, 57(7):1676–1689, July 2010.
- [11] L. Cohen and R. Kimmel. Global minimum for active contour models: A minimal path approach. *Computer Vision, IEEE International Journal on*, 24(1):57–78, 1997.
- [12] Zhuowen Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 1589–1596, Washington, DC, USA, 2005. IEEE Computer Society.
- [13] P. Tiwari, M. Rosen, G. Reed, J. Kurhanewicz, and A. Madabhushi. Spectral embedding based probabilistic boosting tree (sceptre): Classifying high dimensional heterogeneous biomedical data. *Medical Image Computing and Computer Assisted Intervention*, 1:844–851, 2009.
- [14] 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. *Medical Imaging, IEEE Transactions on*, 27(9):1342–1355, 2008.
- [15] E. Sato, S. Olson, J. Ahn, B. Bundy, H. Nishikawa, F. Qian, A. Jungbluth, D. Frosina, S. Gnjjatic, C. Ambrosone, J. Kepner, T. Odunsi, G. Ritter, S. Lele, Y. Chen, H. Ohtani, L. Old, and K. Odunsi. Intraepithelial cd8+ tumor-infiltrating lymphocytes and a high cd8+/regulatory t cell ratio are associated with favorable prognosis in ovarian cancer. *Proc Natl Acad Sci U S A*, 102(51):18538–43, Oct. 2005.

- [16] B. Clarke, A. Tinker, C. Lee, S. Subramanian, M. van de Rijn, D. Turbin, S. Kalloger, G. Han, K. Ceballos, M. Cadungog, D. Huntsman, G. Coukos, and C. Gilks. Intraepithelial t cells and prognosis in ovarian carcinoma: novel associations with stage, tumor type, and *brcal* loss. *Mod Pathology*, 22:393–402, 2008.
- [17] L. Zhang, J. Conejo-Garcia, D. Katsaros, P. Gimotty, M. Massobrio, G. Regnani, A. Makrigiannakis, H. Gray, K. Schlienger, M. Liebman, S. Rubin, and G. Coukos. Intratumoral t cells, recurrence, and survival in epithelial ovarian cancer. *N Engl J Med*, 348(3):203–13, Jan. 2003.
- [18] A. Janowczyk, S. Chandran, and A. Madabhushi. Quantifying local heterogeneity via morphologic scale: Distinguishing tumor from stroma. In *MICCAI (1)*, page In Press, 2012.
- [19] A. Beck, A. Sangoi, S. Leung, R. Marinelli, T. Nielsen, M. Van de Vijver, R. West, M. Van de Rijn, and D. Koller. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science Translational Medicine*, 3(108):108ra113, 2011.
- [20] B. Lahrmann, N. Halama, H. Sinn, P. Schirmacher, D. Jaeger, and N. Grabe. Automatic tumor-stroma separation in fluorescence tmas enables the quantitative high-throughput analysis of multiple cancer biomarkers. *PLoS ONE*, 6(12):e28048, 12 2011.
- [21] K. Mosaliganti, F. Janoos, O. Irfanoglu, R. Ridgway, R. Machiraju, K. Huang, J. Saltz, G. Leone, and M. Ostrowski. Tensor classification of n-point correlation function features for histology tissue segmentation. *Medical Image Analysis*, 13(1):156 – 166, 2009.
- [22] A. Madabhushi and J. Udupa. New methods of MR image intensity standardization via generalized scale. *Med Phys*, 33(9):3426–34, Sept. 2006.
- [23] 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, Oct. 2006.
- [24] N. László, J. Udupa, and P. Saha. Incorporating a measure of local scale in voxel-based 3d image registration. *IEEE Trans. Med. Imaging*, 22(2):228–237, Feb. 2003.

- [25] I. Hontsch and L. Karam. Locally adaptive perceptual image coding. *IEEE Transactions on Image Processing*, 9(9):1472–1483, Sept. 2000.
- [26] 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, Feb. 2000.
- [27] 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, Sept 2005.
- [28] G. Granlund. Fourier preprocessing for hand print character recognition. *IEEE Transactions on Computers*, 21(2):195–201, Feb. 1972.