SEGMENTATION OF COLON GLANDS BY OBJECT GRAPHS

A THESIS

SUBMITTED TO THE DEPARTMENT OF COMPUTER ENGINEERING

AND THE INSTITUTE OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Melih Kandemir

July, 2008

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Çiğdem Gündüz Demir (Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Pınar Duygulu

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Volkan Atalay

Approved for the Institute of Engineering and Science:

Prof. Dr. Mehmet B. Baray
Director of the Institute

ABSTRACT

SEGMENTATION OF COLON GLANDS BY OBJECT GRAPHS

Melih Kandemir
M.S. in Computer Engineering
Supervisor: Assist. Prof. Dr. Çiğdem Gündüz Demir
July, 2008

Histopathological examination is the most frequently used technique for clinical diagnosis of a large group of diseases including cancer. In order to reduce the observer variability and the manual effort involving in this visual examination, many computational methods have been proposed. These methods represent a tissue with a set of mathematical features and use these features in further analysis of the biopsy. For the tissue types that contain glandular structures, one of these analyses is to examine the changes in these glandular structures. For such analyses, the very first step is to segment the tissue into its glands.

In this thesis, we present an object-based method for the segmentation of colon glands. In this method, we propose to decompose the image into a set of primitive objects and use the spatial distribution of these objects to determine the locations of glands. In the proposed method, pixels are first clustered into different histological structures with respect to their color intensities. Then, the clustered image is decomposed into a set of circular primitive objects (white objects for luminal regions and black objects for nuclear regions) and a graph is constructed on these primitive objects to quantify their spatial distribution. Next, the features are extracted from this graph and these features are used to determine the seed points of gland candidates. Starting from these seed points, the inner glandular regions are grown considering the locations of black objects. Finally, false glands are eliminated based on another set of features extracted from the identified inner regions and exact boundaries of the remaining true glands are determined considering the black objects that are located near the inner glandular regions.

Our experiments on the images of colon biopsies have demonstrated that our proposed method leads to high sensitivity, specificity, and accuracy rates

and that it greatly improves the performance of the previous pixel-based gland segmentation algorithms. Our experiments have also shown that the object-based structure of the method provides tolerance to artifacts resulting from variances in biopsy staining and sectioning procedures. This proposed method offers an infrastructure for further analysis of glands for the purpose of automated cancer diagnosis and grading.

Keywords: Histopathological image analysis, gland segmentation, object-based segmentation, object-graphs.

ÖZET

KALIN BAĞIRSAK BEZLERİNİN NESNE ÇİZGELERİYLE BÖLÜMLENMESİ

Melih Kandemir
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Assist. Prof. Dr. Çiğdem Gündüz Demir
Temmuz, 2008

Histopatolojik inceleme, kanseri de içeren büyük bir hastalık grubunun tanısı için en sık kullanılan klinik tanı tekniklerinden birisidir. İnsan gözüyle uygulanan bu yöntemin içerdiği gözlemci tutarsızlığını ve manüel harcanan eforu azaltmak için birçok sayısal yöntem önerilmiştir. Bu yöntemler, dokuyu bir matematiksel özellikler kümesi olarak tanımlar. Tanımlanan özellikleri sonraki biyopsi analizleri için kullanır. Bez yapıları içeren doku tiplerinde, bezlerde meydana gelen değişikliklerin incelenmesi bu analizlerden bir tanesidir. Böyle analizlerde ilk adım, dokudaki bezsel alanları bölümlemektir.

Bu tezde, bağırsak bezlerinin bölümlenmesi için nesne tabanlı bir yöntem önerilmektedir. Bu yöntemde, görüntüden bir basit nesneler kümesi çıkartılması ve bu nesnelerin uzamsal dağılımları kullanılarak bezlerin konumlarının tespit edilmesi önerilmiştir. Önerilen bu yöntemde, pikseller öncelikle çeşitli histolojik yapılara karşılık gelecek şekilde, renk yoğunluklarına göre gruplanır. Ardından, kümelenmiş görüntüden bir basit dairesel nesneler kümesi (lümen bölgeleri için beyaz nesneler ve hücresel bölgeler için siyah nesneler) elde edilir ve bu nesnelerden uzamsal dağılımlarını nicelemek amacıyla bir çizge oluşturulur. Sonrasında, bu çizgeden bir dizi özellik çıkartılır ve çıkartılan bu özellikler bez adaylarının baslangıç çekirdek noktalarını belirlemek için kullanılır. Bu çekirdek noktalarından başlayarak ve siyah nesnelerin konumları gözetilerek, bezlerin iç bölgeleri alan büyütme yöntemiyle tespit edilir. Son olarak, bezlere ait olmayan alanlar, tespit edilen iç bölgelerden çıkartılmış bir diğer özellik kümesi sayesinde elenir ve geriye kalan gerçek bezlerin sınırları, bu bezlerin yakınındaki siyah nesnelerin konumları gözetilerek belirlenir.

Kalın bağırsak biyopsi görüntüleri üzerinde yaptığımız deneyler, önerdiğimiz bu yöntemin yüksek hassasiyet, belirlilik ve doğruluk oranları sağlayabildiğini

göstermiştir. Ayrıca, deneylerimizden elde ettiğimiz sonuçlar bu yöntemin, daha önceki çalışmalarda önerilmiş piksel tabanlı bez bölümleme algoritmalarının performanslarını istatistiksel anlamlı bir şekilde iyileştirdiğini ortaya koymuştur. Yapılan deneyler, yöntemin nesne tabanlı yapısının, boyama ve kesme yöntemindeki farklılıkların yan etkilerine karşı tolerans sağladığını da göstermiştir. Önerilen bu yöntem, otomatik kanser tanısı ve derecelendirmesi amacıyla, bezlerin daha ileri düzey analizi için bir altyapı da sunmaktadır.

Anahtar sözcükler: Histopatolojik görüntü analizi, bez bölümlenmesi, nesne tabanlı bölümleme, nesne çizgeleri.

Acknowledgement

I would like to thank to Assist. Prof. Dr. Çiğdem Gündüz Demir who gave me the chance to work with her on this thesis. With her great supervision, intimate encouragement, and close interest this work has become possible. I would express my gratitudes to my dear mother, father, and sister for supporting me with their love every minute during my M.S. education. I thank to Prof. Dr. Cenk Sökmensüer for his consultancy on medical knowledge and to Hacettepe University, Department of Pathology, for providing us the dataset, and to Akif Burak TOSUN for preparing the dataset. I also express my pleasure to TÜBİTAK (The Scientific and Technological Research Council of Turkey) for supporting me financially.

Contents

List of Figures

List of Tables

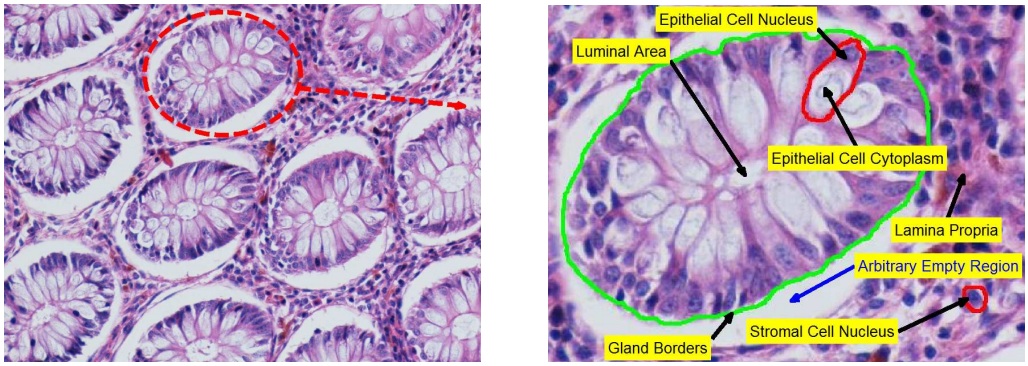
Chapter 1

Introduction

1.1 Overview & Motivation

In histopathological examination of a biopsy tissue, the pathologist visually examines the tissue under a microscope to identify tissue changes related to disease of the interest. This histopathological examination is the most important tool for routine clinical diagnosis of a large group of diseases including cancer. However, this examination may lead to considerable amount of intra- and inter-observer variability as it mainly relies on the visual examination [1, 2]. To reduce the observer variability, computational methods that provide objective measures have been proposed [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. These computational methods quantify a tissue image and the tissue changes related to disease by extracting different types of mathematical features from the tissue and make decisions based on the extracted features. In literature, different types of features have been used (e.g., morphological [3, 4], textural [5, 6, 7, 8, 9], fractal [10, 11], and structural [6, 12, 13, 14, 15] features) as the tissue structure shows differences from one tissue type to another. Several types of tissues such as prostate, colon, breast, and thyroid include glandular structures. To quantify such tissues, and hence to identify the related diseases, the very first step is to segment these issues into their gland structures.

In literature, there are only few studies that focus on the problem of the automatic gland segmentation for tissues that contain gland structures [16, 17, 18, 19]. These studies make use of the fact that glands are characterized by their luminal areas surrounded by the epithelial cells; an example of the histopathological image of a colon tissue is given in Figure 1.1. In order to capture this characterization, these studies first identify the pixels of different histological structures in a tissue (e.g., nucleus, stroma, cytoplasm, and lumen). Then, they detect the seed regions at the locations that contain significant amount of lumen pixels and grow these regions until nucleus pixels are encountered. Finally, they eliminate false gland regions based on their areas and/or the color properties of their pixels.



(a) (b)

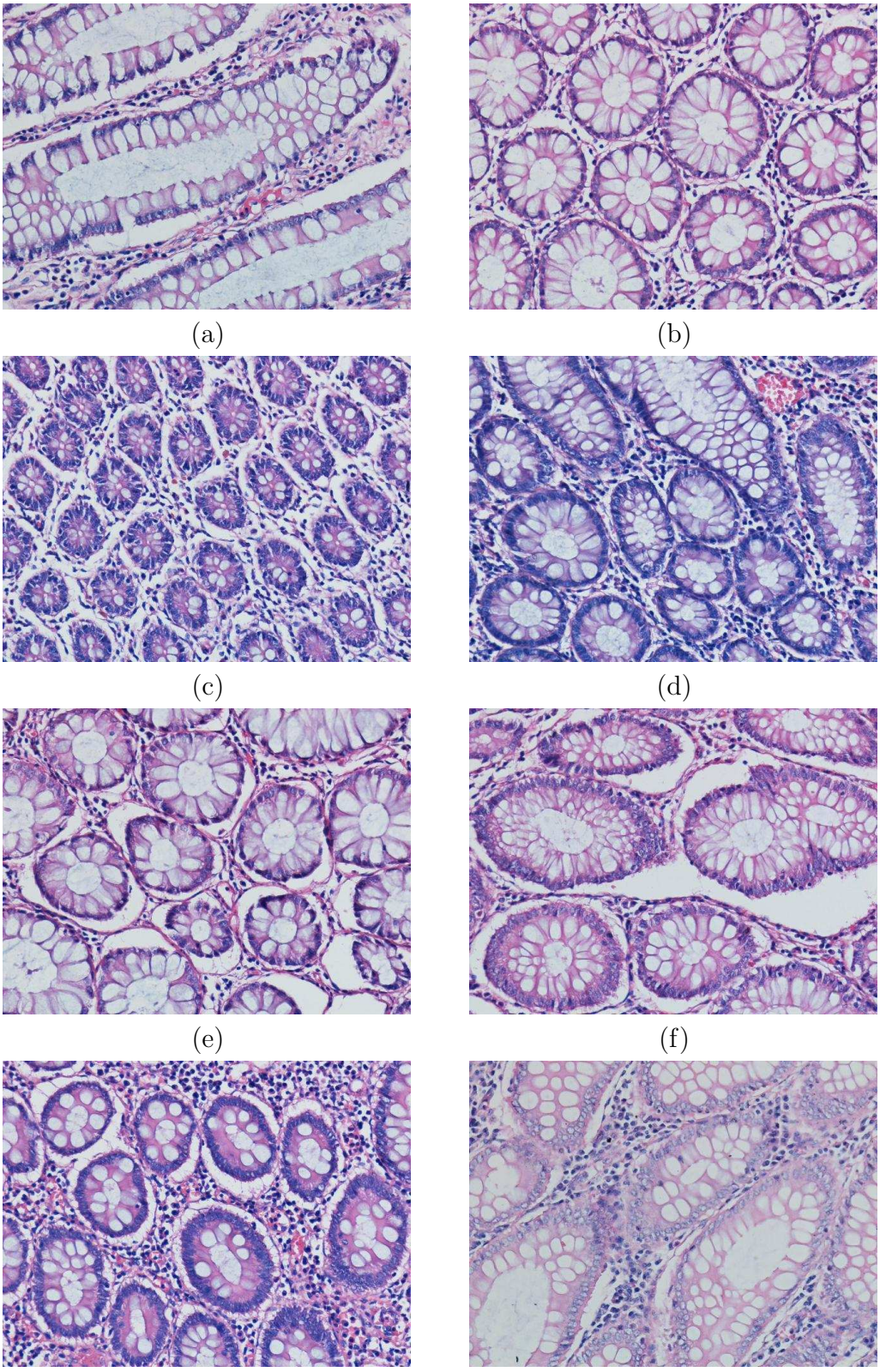
Figure 1.1: (a) An example image of a colon tissue sample that is stained with hematoxylin-and-eosin and (b) an individual gland of a tissue.

These previous studies yield promising results for tissues where the gland appearance has relatively more regular structure, the gland boundaries are more prominent, and the tissue has fewer amounts of noise and artifacts. However, many tissue sections commonly consist of considerable amount of noise and artifacts due to the staining and/or sectioning procedures. Moreover, the variations in these procedures may result in huge variances in gland appearances. First, glands could have different sizes depending on the orientation of a tissue at the time of sectioning. For example, the gland sizes are different in the tissue images shown in Figures 1.2(a-c) although all of these images are taken with the same magnification; glands with different sizes could even exist within the same image (Figure 1.2(d)). Therefore, false gland elimination based on area might lead to misleading results; no single area threshold could be found for all these images.

Second, because of the density difference between the glandular and connective tissue structures, the sectioning procedure may result in large white artifacts on the boundaries of the glands (Figures 1.2(e and f)). These areas are not luminal areas and do not belong to gland structures; thus, they should be eliminated. However, it is much more difficult to distinguish these white regions and the true luminal areas by using only the pixel-based information. Third, due to the staining procedures, it is rare to find continuous nucleus pixels around the luminal area. Thus, the growing process of the seed region could not be stopped and flooding occurs. For example, although it is possible to find such continuous nucleus pixels in the tissue shown in Figure 1.2(g), it is much more difficult to find those in a tissue shown in Figure 1.2(h). Because of all these issues, using only the pixel-based information leads to incorrect gland segmentations for especially tissues with artifacts and variations.

The contribution of this thesis is as follows: it presents a new gland segmentation algorithm that relies on decomposing the image into a set of primitive objects and employs the spatial relations between these objects instead of directly using the pixel-based information. This object-based algorithm suggests constructing a graph on all of the primitive objects and determining the gland seeds based on the features extracted from this object graph. Then, it constructs another graph on the nucleus objects, and uses this second graph to grow the gland seeds. At the last step, it eliminates false glands based on another set of features, which are extracted from these glands, and determines the final boundaries of the true glands regarding the locations of nucleus objects. As opposed to the previous approaches that use only the pixel-based information, this thesis proposes to use object-based information for the automatic segmentation of colon glands.

This thesis is organized as follows: In Chapter 2, we introduce the medical terminology that we refer in our work and give summary about the related studies. In Chapter 3, we describe our object-based algorithm in detail. In Chapter 4, we present our experimental results. Finally, we provide a summary of our work and a future perspective for our research in Chapter 5.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 1.2: Histopathological images of colon tissues, which are stained with the routinely used hematoxylin-and-eosin technique. All of the images are taken with the same magnification and the same lightning conditions.

Chapter 2

Background

In this thesis, we focus on the segmentation of colon glands stained with the hematoxylin-and-eosin technique. In this chapter, we first introduce the medical terminology that we refer in the text. We then provide a survey of related studies in the context of both general image segmentation and gland segmentation.

2.1 Terminology

In this thesis, we focus on the images stained with the hematoxylin-and-eosin technique. This staining technique is the one that is routinely used in hospitals. In this technique, the basic dye hematoxylin color basophilic structures with blue-purple hue, and alcohol-based acidic eosin color eosinophilic structures with bright pink [20]. Therefore, the color spectra of the images of tissues stained by this technique are commonly rich of blue-purple, pink, and white pixels.

In the context of gland segmentation, histological structures in the colon tissue and the spatial relations between these structures are the primary concern. The histological structures that we refer in this work are marked on the colon tissue image in Figure 2.1.

A *gland* is defined as a specialized group of cells that secrete a substance for use

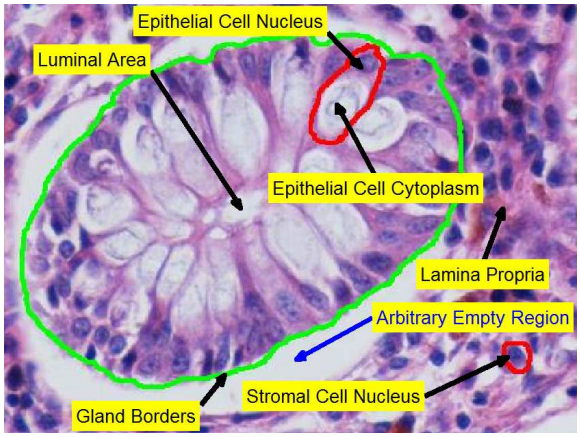


Figure 2.1: The histological structures in a colon tissue.

in the body [21]. Several types of tissues such as prostate, small intestine, colon, thyroid, and breast include glandular structures. The appearance of glands vary with the type of the tissue. In this thesis, we focus on the automatic detection of the borders of glands in colon tissue images; in the image given in Figure 2.1, the border of a gland is marked with green.

There are two types of cells in a colon tissue: epithelial cells and stromal cells. The gland is formed by a chain of *epithelial cells*. In the image, the borders of an epithelial cell is marked with red, the dark purple region inside these borders is the *nucleus* of the epithelial cell, and the large white region below the nucleus is its *cytoplasm*. Epithelial cells appear side by side around an oval vacant region called *luminal area*. As shown in Figure 2.1, the luminal area locates at the center of the glandular region and is surrounded by the epithelial cells.

We refer all other type of cells in the connective tissue as *stromal cells*. They appear as scattered around the area outside the gland bodies. In the context of gland segmentation problem, they should not be included into the gland region since they are not a part of the gland structure.

There exists a pink area between all of the cells in the tissue. This is a non-cellular material called *lamina propria*. This material is the connective tissue that holds all the entities in the tissue together. Since the connective tissue exists among every type of entity, it does not provide any discriminative information

that is useful for gland segmentation.

In a typical colon tissue, there may also exist empty white regions outside the gland bodies. These regions do not include any epithelial cells, stromal cells, and lamina propria. They are the artifacts that arise from the sectioning procedure (e.g., the one shown with blue arrow in Figure 2.1). For the gland segmentation problem, they are considered as noisy content.

2.2 Related Work

In this section, we first discuss the general image segmentation approach. We then discuss the previous gland segmentation approaches for different tissue types.

2.2.1 Image Segmentation

Image segmentation is defined as the process of partitioning an image into non-overlapping and connected pixel groups (regions) that are semantically coherent in a particular context. It is a special case of pixel classification, in which pixels do not have to be connected and instead of semantic coherence, similarity of low-level features (i.e., intensity) is the primary concern [22].

There are numerous approaches to image segmentation in literature. These approaches can mainly be grouped into three categories [23]:

1. Feature-space based approaches
2. Image-domain based approaches
3. Physics based approaches

In feature-space based approach, the units of data in the image, often the pixels (or voxels), are grouped into several classes regarding their values in a certain

feature space. The definition of the feature space is application-dependent. To this extent, the feature-space based approaches, when used alone, offer solutions to the pixel classification phases of segmentation problems.

One of the oldest methods among feature-space based pixel classification approaches is histogram thresholding. Although being more trivial, these are generally most efficient methods in terms of computational requirements. Otsu [24] has a seminal work in which he proposed a statistical threshold determination method for grayscale images. The primary disadvantage of thresholding is that it can only be applied to grayscale images. However, some recent statistical techniques are developed which can learn the best parameters for transforming a multichannel image to grayscale. For example, Mao et al. [25] propose a method for cell segmentation that maps the color image into grayscale by a transform whose parameters are determined through supervised-learning over a training set.

Different learning algorithms have also been used after the feature-space has been defined. In the case of unsupervised pixel classification, k-means clustering [26] is widely preferred. For example, Park et al. [27] use the k-means algorithm by defining RGB channels as the feature-space. They extract a number of initial seeds from the difference-of-Gaussian (DoG) smoothed 3-D color histogram. Wu et al. [28] use the c-means algorithm, which is the fuzzy extension of k-means.

Apart from the clustering schemes, statistical pattern recognition methods are also used in image segmentation. Belongie et al. [29]. define a transformation from raw pixel data to image regions that are coherent in color and texture space, which they call *Blobworld*. Segmentation is made using an expectation-maximization algorithm on combined color and texture features. The Blobworld representation is used for query evaluation in a content-based image retrieval (CBIR) system.

In image-domain based approaches, in addition to the classification of pixels according to some criteria, the spatial relationships of the pixels are also taken into consideration. These methods are frequently used together with feature-space based methods. One of the ways of combining the two approaches is the

split-and-merge scheme, in which the image is first split into regions by a feature-space based pixel classification technique (i.e, k-means [30]), then the regions are merged according to an application-specific homogeneity metric. For example, Deng and Manjunath [31] use the *J-value*, which is a function of the variance of color intensities of pixels belonging to the same class, as a homogeneity metric.

Region growing scheme is an alternative combiner of homogeneity and compactness concepts. In this scheme, initial seeds are determined, then the neighbors of these seeds are assigned to the seeds according to a homogeneity measure. When determining the seeds, selecting the local minima of pixel intensities [32] is one of the most accepted seed determination strategies. As in k-means clustering, region growing algorithms also have fuzzy counterparts [33, 34].

There are several alternative techniques that approach segmentation problem as a special case of edge-detection problem. There has been developed advanced mathematical models for dynamic contour detection. Kass et al. [35] propose active contour models (snakes) that are deformed with respect to an energy function along a vector field defined on the image.

The physics-based techniques approach the segmentation problem in the inverse direction. Given an image, considering the physical illumination models, these methods try to infer by what kind of objects the light should have interacted when forming that image. Healey [36, 37] proposes several segmentation algorithms based on this principle that work when there is one illumination source in the environment. Maxwell and Shafer [38] introduce a more general algorithm that can work for arbitrary number of illumination sources.

2.2.2 Gland Segmentation in Histopathological Images

In this section, a survey of methods related to gland segmentation in various types of tissues is presented.

In [16], Wu et al. propose a method that adopts seed detection and region-growing paradigm for segmenting glands of small intestine tissues stained with

hematoxylin-and-eosin. In the proposed method, the color image is first converted to grayscale and then manually thresholded in order to mark the nuclear pixels. The method suggests detecting the seed regions from where region growing will start. These regions correspond to vacant regions in which there is no nuclear pixel. This is done by detecting the pixel coordinates in which a large round window of non-nuclear pixels can fit. The pixels in these windows are regarded as gland seeds. Then an iterative region growing starts from each seed by dilating the image with another round window. The size of this window should be chosen to be larger than the largest gap between two epithelial cells in the image, otherwise flooding occurs. After region growing, false glands are eliminated based on two assumptions. First, it is assumed that if a seed corresponds to a gland, growing should stop after a small number iterations. Otherwise, it is regarded as a *false gland* and the entire region is eliminated. Second, it is assumed that the epithelial cell nuclei around a true gland form a thick dam. After eliminating false glands based on these assumptions, the epithelial cell nuclei that belong to each true gland seed are detected using a dilation belt.

In [17], same authors introduce another method that works on the grayscale images of small intestine tissues stained with hematoxylin-and-eosin. The method suggests enhancing the image so that the chains of epithelial cell nuclei that surround the luminal area become more apparent. Enhancement is done by lowering the intensities of the dark pixel groups that have a specific orientation and smoothing the others. They produce four intermediary images using directional 2-dimensional linear filters with four different orientations; each of these filters enhance the pixels of epithelial cells in a particular orientation. Then, these four images are combined into a single image by taking the minimum intensity for each pixel. The resultant image is expected to contain apparently dark and thick regions at the places of epithelial cell nuclei since it is produced by collection of enhancement in all directions. The dark dam around the gland body is then extracted by manual thresholding. After several morphological operations such as dilation and region filling are applied, the gland borders are obtained. In [18], Wu et al. use the same idea with four median filters biased along four different orientations. The idea is identical to the one in [17] except that instead of using

directional Gaussian filters, biased median filters are applied over the input image for enhancing the epithelial cells. The filtered images are combined into a single image just in the same way as in [17].

Naik et al. [19] propose a Bayesian approach for automatic segmentation of prostate glands. The algorithm works on color images of the tissue samples stained with hematoxylin-and-eosin. A Bayesian classifier is trained by manually labeling the pixels of the images in the training set with the histological structures they belong to (luminal region, epithelial cytoplasm, and epithelial nucleus). For a query image, luminal regions are first detected by collecting the pixels with highest posterior conditional probability for the lumen class. From the detected regions, ones with very small and very large sizes are considered as false detections and they are eliminated. The boundaries of the detected luminal regions are used to initialize a level-set curve which evolves until it reaches a region which most likely belongs to nuclei. Another elimination of false detections is performed considering the final sizes of the resulting level-set curves. By augmenting the nuclear regions at the neighborhood of the level-set curves to the regions surrounded by the curves, the borders of the glands are determined.

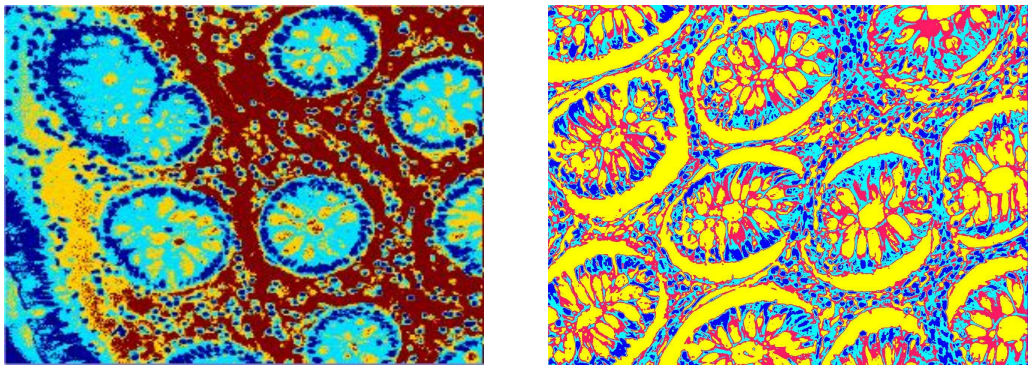
In [39], Farjam et al. propose a textural approach for the segmentation of glands in prostate tissues stained with hematoxylin-and-eosin. After clustering the pixels according to their textural properties, they obtain prostate glands by excluding the regions that contain nucleus pixels from those that contain stroma and lumen pixels.

In [40], Fernandez-Gonzalez et al. define an algorithm for automatic segmentation of ductal regions in mammary gland tissues and reconstruction of surface geometries of the glands in 3-D. The images are obtained from the tissue blocks stained with hematoxylin-and-eosin. The image is first enhanced by background-correction. Then the approximate contours of the ductal regions are extracted using the Fast-Marching algorithm. The borders are then refine using the Level-Set method. Finally, the 3-D surface geometries of the ducts are generated by combining the 2-D results obtained at the earlier steps.

The main motivation of segmenting glands is to detect malignancies. One

alternative approach to malignancy detection in histopathological images suggests segmenting the cells and analyzing their morphological features instead of detecting the gland borders. The seminal survey paper published by Fernandez-Gonzalez et al. [41] presents a concise summary of recent approaches to quantitative analysis of mammary gland images. Various morphological tissue analysis methods are briefly described for different imaging technologies. In the paper, several techniques for the segmentation of mammary cell nuclei are also described. There exist less automated primitive methods [42, 43] that require user intervention. Thresholding based methods [44, 45, 46, 47, 48] rely on the assumption that an intensity threshold determines whether or not a pixel belongs to a nuclear region. These methods are generally easy to implement, but they are quite sensitive to noise. Pixel classification based approaches [49, 50] make use of the supervised and unsupervised machine learning techniques when determining the nuclei pixels in an image. Model fitting based approaches [51, 52, 53, 54, 55] utilize a priori information such as nuclei size and shape over regions in the image. Although these methods provide accurate and smooth results, they are computationally expensive and sensitive to noise. Another cell segmentation scheme is region growing using active contours [56, 57, 58, 59, 60, 61]. This scheme also provides accurate and smooth results. However, high computational cost and need of user intervention for initial seeding of the nuclei are among the disadvantages of this scheme. The cell segmentation problem becomes more difficult for the images of biopsy tissues that are prepared by routinely used staining and sectioning procedures as these procedures lead to a significant amount of noise. Hence, the studies focus on analyzing hyperspectral data, rather than intensity based images taken from optical microscopes. Although they provide more information, hyperspectral images can be taken by electron microscopes which are not very prevalent due to their high cost. In one of these studies that use hyperspectral data, Rajpoot et al. [62] propose an algorithm that segments the malignant regions in a hyperspectral colon image by classifying the nuclei as normal and malignant regarding their morphological properties. For this purpose, as the first step, they use the k-means clustering algorithm to segment the image into four constituent parts of a gland tissue (nuclei, cytoplasm, lamina propria, and lumen) over the hyperspectral data. Thanks to the high-dimensional spectral data, the

segmentation step gives quite accurate results. The borders of the glands can be straightforwardly obtained from this segmented image. However, such a successful segmentation can not be obtained when the k-means clustering algorithm is applied over color intensities of an image taken from an optical microscope, which is our concern. The segmentation results of two colon tissue images obtained by applying k-means over hyperspectral data and over the color intensities of one of the images in our dataset are given in Figure 2.2 (a and b), respectively. As it can be seen in Figure 2.2 (b), none of the four clusters provide any hint about the gland borders in our case.



(a) (b)

Figure 2.2: (a) The cyan cluster provides direct information about gland borders when k-means is applied to hyperspectral data [63]. (b) None of the clusters provide information about gland borders when k-means is applied to color intensities of optical microscopic images.

Chapter 3

Segmentation of Colon Glands by Object Graphs

In this thesis, we propose an object-based method for the segmentation of colon glands. To this end, we define objects to represent the tissue components and use the spatial relations between these objects for the segmentation of gland structures.

The proposed object-based method comprises a series of analysis steps. In the first step, the pixels of the input image are clustered into three groups (each of which correspond to a biologically meaningful color) based on their color information. Then the primitive objects (white objects and nuclear objects) are defined from the clustered pixels using the *Circle-Fit Transform* algorithm. In the next step, a graph is constructed from the primitive objects by making use of the white and nucleus objects and gland seeds are detected with respect to a set of features extracted from this graph. Then, another graph is constructed by connecting the nuclear objects and the inner gland regions (*gland candidates*) are determined by region growing starting from the gland seeds and ending at the edges of this nuclear graph. Subsequently, in the false gland elimination step, gland candidates are classified as *false glands* and *true glands* with respect to a set of features. Finally, in the last step, the exact borders of the true glands are

detected by merging the corresponding epithelial cells to the inner gland regions. The summary of the object-based algorithm is given in Figure 3.1. The details of each step will be explained in the following subsections.

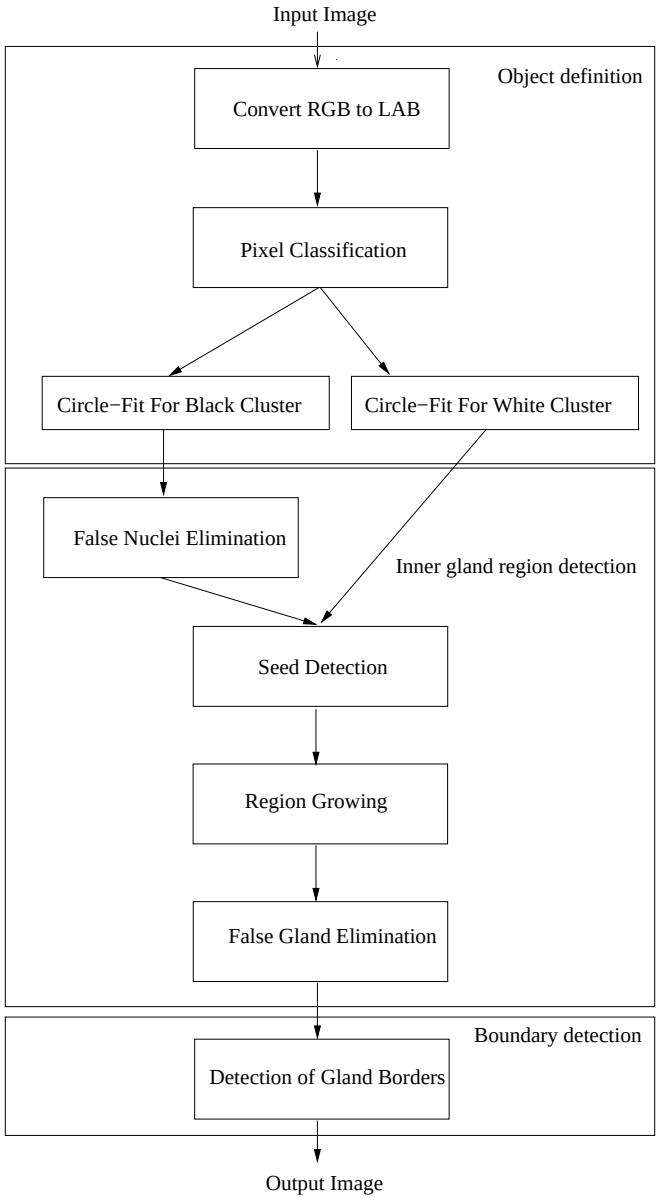


Figure 3.1: The block diagram of the proposed system.

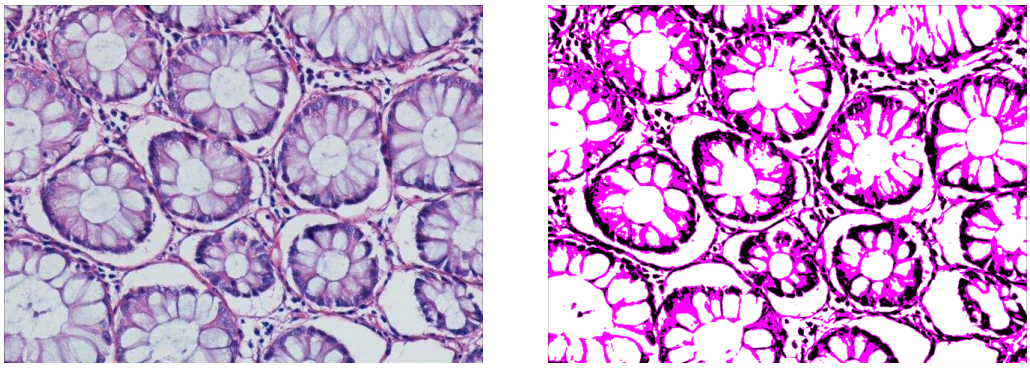
3.1 Pixel Classification

There are three main color groups in the image of a tissue stained with hematoxylin-and-eosin (pink, blue-purple, and white). The chemical structure of hematoxylin-and-eosin staining technique provides us some information for matching these color groups to histological structures [20]. Since the cell nuclei are commonly chromatin-rich, it is most likely that a purple pixel belongs to a nuclear region. On the other hand, the cytoplasmic regions of stromal cells and connective tissues are known to be eosinophilic. Thus, a pink pixel most likely belongs to a stromal cell cytoplasm or connective tissue. The epithelial cell cytoplasms and luminal area occur in very light pink or white, as opposed to dark pink stromal cell cytoplasms, since they include secretion material which is affected from neither hematoxylin nor eosin. The remaining arbitrary empty regions that do not include any histological structures also remain colorless, and produce white pixels.

We have exploited the a-priori semantic information that comes from the staining technique for pixel classification. We assume that the pixels form three disjoint clusters in the color space. Hence, we apply the k-means clustering algorithm [26] with k=3 over raw pixel data represented in the Lab color space. The color vectors are attached to the clusters by considering the Euclidean distances of their centroids to those of absolute white, pink, and dark purple. In the text, these clusters are referred to as the *white cluster*, the *pink cluster*, and the *black cluster*, respectively. The classification result of a sample image is given in Figure 3.2.

3.2 The Circle-Fit Transform

Although classification of pixels provides semantics at the lowest level, this information is not specific enough for our purposes. As an example, although we know the nuclear regions, we are not able to identify individual nuclei. Applying connected component analysis over the pixels that are labeled as *nucleus* does not



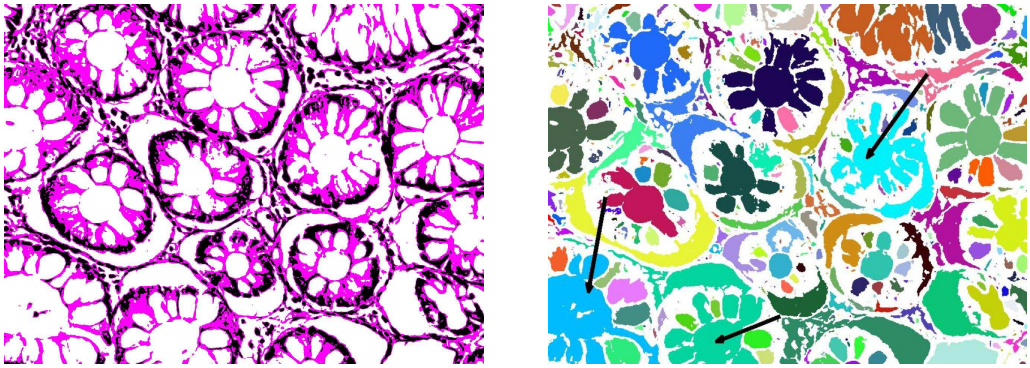
(a) (b)

Figure 3.2: (a) The original colon tissue image and (b) the quantized tissue image obtained using k-means with k=3 over the color intensities of the pixels.

provide sufficiently reliable results since epithelial nuclei appear side by side and form a single connected component consisting of multiple nuclei. Thus, higher level analysis is necessary to represent the nuclei in the scene. Moreover, a purple pixel could belong to an epithelial cell nucleus or a stromal cell nucleus. Similarly, a white pixel may correspond to an epithelial cytoplasm, a luminal area, or an artifactual empty region. These ambiguities can not be resolved when the pixel-based information is used alone. Thus, we propose to define objects for each of the histopathological structures and use the information extracted from these objects instead of using the pixel-based information alone.

The most ideal way for object definition is to segment the primitive histological entities directly and to define the connected components in the resultant segmentation as objects. In other words, such an approach suggests dividing the gland segmentation problem into smaller subproblems by treating nucleus segmentation, lumen area segmentation, and epithelial cytoplasm segmentation as separate problems, and then combining the results in order to determine the gland borders. However it gives rise to more difficult segmentation problems than the original one. In literature, there are a considerable amount of studies on cell segmentation; however, most of them require high-magnification images and/or high-dimensional hyperspectral data [63, 62, 64]. There do not exist any previous studies that focus on cytoplasm and lumen segmentation.

An alternative way for object definition is to identify each connected component in the image as a separate object. In the common practice, this scheme is applied after a series of morphological operations to eliminate noise. However, our experiments on histopathological images have shown that this scheme is not applicable for our problem since the objects in the scene are not well separated. For instance, it is quite likely that there are white empty regions surrounding the glands. When the nucleus pixels do not form a closed dam, the white regions inside the glands may appear connected to the empty regions outside. Such a primitive object would be meaningless and inappropriate for higher level analysis. The problem is illustrated on an exemplary image shown in Figure 3.3 (b). In this image, the connected components of the white pixels are shown in arbitrary colors, and three artifactual components are marked with black arrows.



(a) (b)

Figure 3.3: Inappropriate primitive object definition when connected component analysis is used. (a) The clustered image and (b) the connected components of the white pixels.

To circumvent the aforementioned difficulties, we propose to approximately represent the histological structures in a tissue. For this purpose, we implement a new method, the *Circle-Fit Transform*, that transforms the raw pixel data into a set of circular objects. Then, we use these objects to represent the histological structures.

The *Circle-Fit Transform* algorithm is given in Algorithm 1. This algorithm inputs a clustered image (I_i) and a label of interest (loi). It first converts the clustered image into a binary image I (line 4) by marking the pixels that are equal

Algorithm 1 Circle Fit Transform Algorithm

function CIRCLEFITTRANSFORM(I_i, loi)
    //I_i: Clustered image
    //loi: Label of interest
    $I \leftarrow (I_i = loi)$
    $I_{cc} \leftarrow$ConnectedComponentAnalysis(I)
    //Initialize the image that holds radius of the largest circle for each pixel
    $I_{cf} \leftarrow$EmptyImageOfSize(sizeof(I_i))
    $InitializeImage(I_{cf}, 0)$
    //Initialize the image that holds circular objects labeled by ids
    $I_{cid} \leftarrow$EmptyImageOfSize(sizeof(I_i))
    $InitializeImage(I_{cid}, 0)$
    //Id number that will be given to the next circle
    $cid \leftarrow 1$
    for all $(p, k) \in I_{cc}$ **do**
        //For each nonzero pixel in I_{cc}
        if $I_{cc}(p, k) > 0$ **then**
            //Find the radius r of the largest circle
            $r \leftarrow argmax_r\{\forall$ i,j Icc(i,j)=1 & $\sqrt{(i-p)^2 + (j-k)^2} < r\ \}$
        end if
        //If a larger circle is found, override the previous ones
        for all $(i, j) \in I_{cc}$ **do**
            if $\sqrt{(i-p)^2 + (j-k)^2} <= r$ & $r > I_{cf}(p, k)$ **then**
                $I_{cf}(i, j) \leftarrow$r
                $I_{cid}(i, j) \leftarrow$cid
            end if
        end for
        //Increment the id number
        cid++
    end for
    return I_{cid}
end function

to *loi* as true, and marking else as false. It then finds the connected components of this binary image (line 5). Next, for each particular pixel it finds the radius of the largest circle that contains this particular pixel in I_{cf} and the id of this largest circle in I_{cid} (lines 8-26). Finally, it returns I_{cid}.

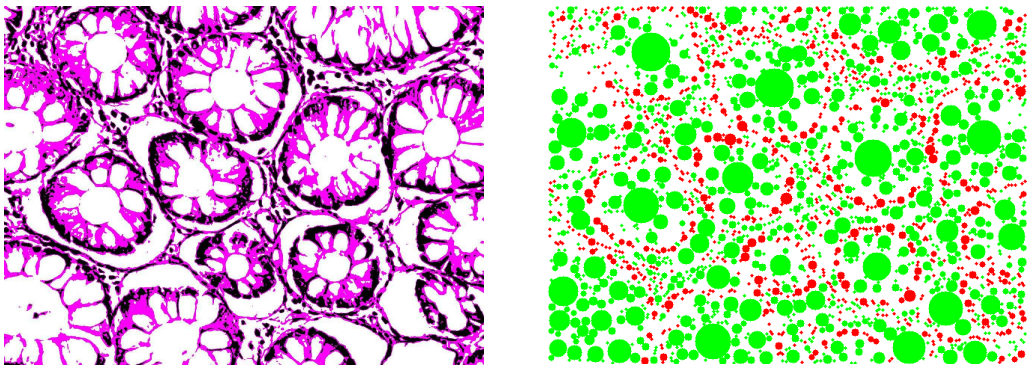
The output of circle-fit transform contains a number of objects, but not all of them are circular. There may exist crescent-like objects next to circles. Such objects occur when a circle with a larger radius partially overrides a previously generated circle. These regions may be sometimes small enough to be ignored, or they may be as large as to be represented by a set of circles. In order to handle such cases, we define an algorithm, *Iterative Double Circle-Fit Transform*, that utilizes the circle-fit transform iteratively and only outputs a set of circular objects. The algorithm consists of two loops in which the circle-fit transform is applied to the binary image I_{in} iteratively until it converges to a state in which no further change occurs between subsequent iterations. During each iteration, after the circle-fit transform is called, some postprocessing is applied to I_{in} to eliminate non-circular and small regions. $EliminateSmallComponents$ function eliminates the regions in I_{in} that correspond to objects in I_{cid} with areas less than the area threshold τ. $EliminateNonCircularRegions$ function computes circularity of each object in I_{cid} using a roundness measure and eliminates the regions in I_{in} that correspond to non-circular objects. The pseudocode of iterative double circle-fit transform is given in Algorithm 2.

The motivation to define circular primitives is that the borders of all of the histological entities of our interest are circular. Since the cell nuclei are in round shape, they are generally represented by a single circle. In the cases when they are not perfectly round, the remaining region out of the circle produces very tiny circles in the succeeding iterations, hence they are all eliminated. Lumen area is represented by one or few large circles in the middle of the gland area, and the epithelial cytoplasms are often represented by uniformly distributed singular middle-scale circles around the lumen circles. The iterative double circle-fit transform of an example image is given in Figure 3.4. In the image, the circles fit into the black cluster are in *red*, and the ones that fit into the white cluster are in *green*. The transform is not applied to pixels in the pink cluster since they do

Algorithm 2 Iterative Double Circle Fit Transform Algorithm

function IterativeDoubleCircleFitTransform(I_i, loi, τ)
    //I_i: Clustered image
    //loi: Label of interest
    //τ: Area threshold
    I_{first} ←EmptyImageOfSize(sizeof(I_i))
    $InitializeImage(I_{first}, 0)$
    $I_{in} \leftarrow I_i$
    while true do
        I_{cid} ←CircleFitTransform(I_{in}, loi)
        I_{cs} ←EliminateSmallComponents(I_{cid}, τ)
        I_{cr} ←EliminateNonCircularRegions(I_{cs})
        I_{in} ←($I_{cr} > 0$)
        if $\|I_{in} - I_{prev}\| = 0$ **then**
            $I_{first} \leftarrow I_{cr}$
            break
        else if
            then $I_{prev} \leftarrow I_{in}$
        end if
    end while
    while true do
        I_{cid} ←CircleFitTransform(I_{in}, loi)
        I_{cs} ←EliminateSmallComponents(I_{cid}, τ)
        I_{second} ←EliminateNonCircularRegions(I_{cs})
        I_{in} ←($I_{second} > 0$)
        if $\|I_{in} - I_{prev}\| = 0$ **then**
            break
        else if
            then $I_{prev} \leftarrow I_{in}$
        end if
    end while
    I_{res} ←Merge(I_{first}, I_{second})
    return I_{cr}
end function

not provide any valuable information.



(a) (b)

Figure 3.4: (a) The clustered image and (b) the iterative double circle-fit transforms of black and white clusters. The circles of black cluster are given in red and the circles of white cluster are given in green.

There have been studies on detecting the circular shapes in an image [65]. There also exist methods that utilize circle fitting for region segmentation in arbitrary scenes, as in [66]. Yet, there exist no previous studies which treat circle fitting as a means of transformation that produces a set of primitive objects which can then be employed by higher-level analysis.

The circle-fit transform provides an acceptable means for primitive object definition. It facilitates the definition of discriminative features for ambiguity resolution. It also enables representation of higher level objects (such as nucleus dams and glands). The circle-fit transform also prevents flooding of connected components. The utilization of these benefits in gland segmentation problem will be clearer in the following steps.

3.3 Detection of Gland Candidates

In this step, we apply seed detection and region growing approaches to detect the regions that are likely to correspond to gland bodies. In the seed detection phase, the aim is to find a set of seed pixels each of which is most likely to be inside a gland body. Then in the region growing step, we find the initial inner borders of

the gland candidates by starting the region growing iterations from the detected seeds.

In this approach, first, the circles in the iterative double circle-fit transform of the white (lumen) cluster are classified as *gland* and *non-gland* circles. Then, close gland circles are combined into a single seed component and the centroid of each combined circle group is considered as the seed of a gland candidate. Region growing starts from these centroids and forms the *gland candidates*, since these regions may correspond to a gland or a false gland. The actual true glands are determined after *False Gland Elimination* step, which is described in the next section.

3.3.1 False Nucleus Elimination

After k-means clustering, the white and black connected components in the clustered image are input to circle-fit transform. Two output images (one for the black cluster and one for the white cluster) that consist of circles are obtained. Herein the circles in the output of the black cluster are referred to as *black circles* and those of the white cluster to as *white circles*.

In the original image, the white circles may correspond to a luminal area, an epithelial cell cytoplasm, or an arbitrary empty region. Among these histological structures, the first two reside in the gland body. Hence, we will not deal with discriminating the first from the second. Instead, we will discriminate the third one from the first two in *Seed Detection* and *False Gland Elimination* steps as described in the following sections.

The black circles may correspond to cell nuclei or noisy dark regions. In the context of *false nucleus elimination*, we discriminate the first from the second and then eliminate the circles that correspond to noisy dark regions. Elimination of noise affects the gland border detection; existence of false nucleus circles in gland bodies results in partial detection of their borders.

To detect the false nuclei, we use the following heuristic: if a black circle

is *isolated* from the other black circles, it most likely corresponds to noise. The epithelial cell nuclei around a gland body locate close to each other, hence it is not likely that an isolated black circle corresponds to an epithelial cell nucleus. On the other hand, stromal cell nuclei appear at arbitrary non-glandular regions in the scene. There may occur isolated stromal cells in a scene since the heuristic defined above does not hold in this case. Although this may cause misclassification of the isolated stromal cells as noise, it does not affect the performance since we are only interested in epithelial cell nuclei in this context.

The false nucleus elimination algorithm is given in Algorithm 3. In this algorithm, the mean and the standard deviation of the distances of circles to their closest neighbors and the sum of their distances to first two closest neighbors are calculated separately (lines 1-16). The circles whose distances to their closest neighbors are greater than the image average plus K_1 times the standard deviation and whose sum of distances to first two closest neighbors are greater than the image average plus K_{12} times the standard deviation are considered as *False Nucleus* and they are eliminated from the image (lines 17-24). We define K_1 and K_{12} as *isolation factors* since these coefficients adjust the degree of isolation that discriminates the false nuclei.

The iterative double circle-fit transform of the black cluster of a sample image is shown in Figure 3.5. The eliminated circles are given in red. Note that the isolated circles in the image are the eliminated ones.

3.3.2 Seed Detection

In this step, we detect a set of seed pixels from which region growing will start. We define a *seed* as a pixel that is most likely to locate at a coordinate close to the centroid of the actual gland body.

Having an object-based representation of the image, we proposed a novel seed detection scheme that exploits this representation. White circles in a given image

Algorithm 3 False Nucleus Elimination Algorithm

procedure EliminateFalseNuclei(I, K_1, K_{12})
  //I: Iterative double circle-fit transform of black cluster
  //K_1 and K_{12}: Isolation factors
  for all $c \in I$ **do**
    //For each circle in the image
    for all $c_i \in I$ & $c_i \neq c$ **do**
      $dlist \leftarrow \|c_i - c\|$
    end for
    $neighOrder \leftarrow Sort$(dlist)
    //Distances from the nearest circle and 2-nearest circles
    $dist_1[c] \leftarrow neighOrder[0]$
    $dist_{12}[c] \leftarrow neighOrder[0] + neighOrder[1]$
  end for
  $mean_1 \leftarrow Mean(dist_1)$
  $mean_{12} \leftarrow Mean(dist_{12})$
  $std_1 \leftarrow StdDeviation(dist_1, mean_1)$
  $std_{12} \leftarrow StdDeviation(dist_{12}, mean_{12})$
  for all $c \in I$ **do**
    //For each circle in I
    if $dist_1[c] > mean_1 + K_1 \times std_1$&
      $dist_{12}[c] > mean_{12} + K_{12} \times std_{12}$ **then**
      //Remove circle c from I
      $Remove$(c,I)
    end if
  end for
end procedure

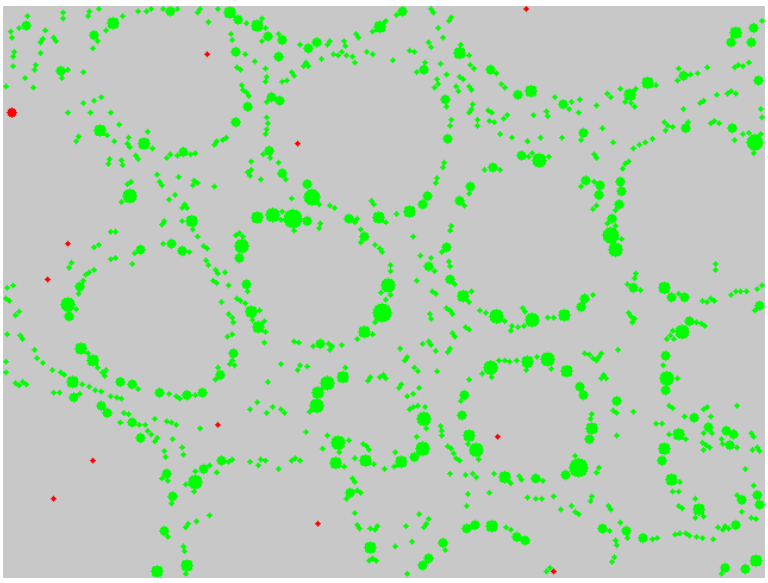


Figure 3.5: False nucleus elimination for a sample image; eliminated circles are shown in red.

are clustered into two with respect to a set of features using the k-means algorithm. One cluster corresponds to a luminal area or an epithelial cell cytoplasm and the other corresponds to arbitrary empty regions. For a particular white circle c, these features are listed below:

1. Distances between the centroid of the white circle c and the centroids of its K-nearest black (nucleus) circles.
2. Distances between the centroid of the white circle c and the centroids of its K-nearest white circles.
3. Polar angles between the line segments that start from the centroid of the white circle c and ends at each of the centroids of its K-nearest black circles.
4. Polar angles between the line segments that start from the centroid of the white circle c and ends at each of the centroids of its K-nearest white circles.
5. Areas of its K-nearest black circles.
6. Areas of its K-nearest white circles.

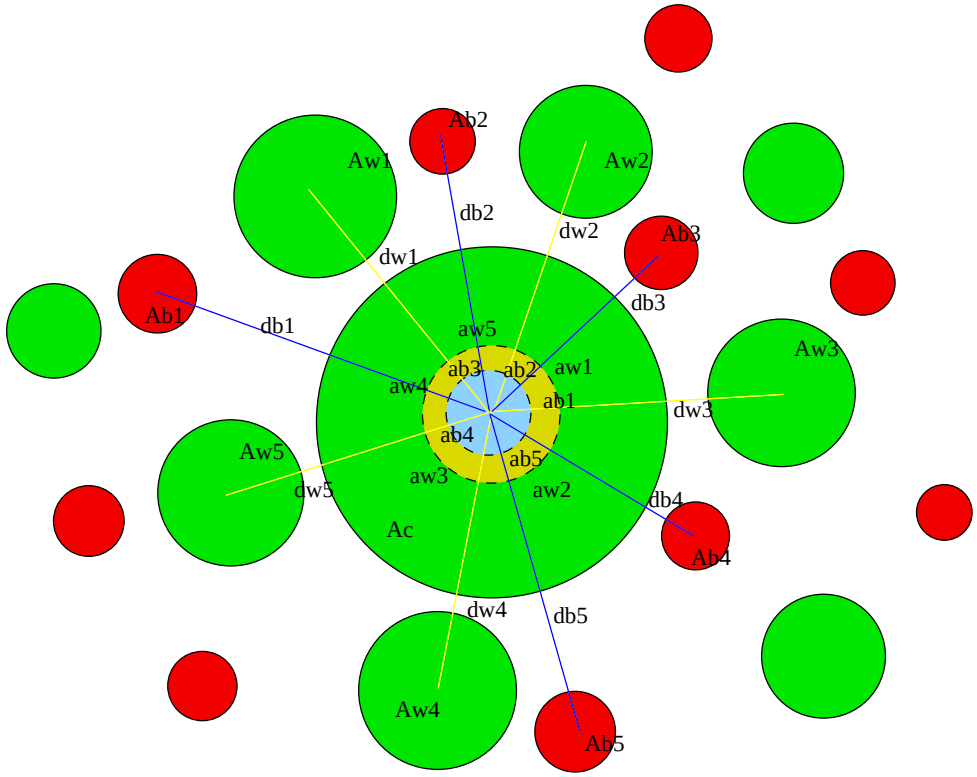
7. Area of the white circle c.

These features are visually illustrated in Figure 3.6. The intuition behind the definition of these features is that the relative spatial ordering of the closest circles to a white circle differs inside and outside the glandular region. Nearest black (nuclear) circles to a white circle in the glandular region are expected to correspond to the epithelial cell nuclei, which have a characteristic ordering around the gland body. In particular, while the nearest black circles generally locate on one side of a gland circle, they are more homogeneously spread around a non-gland circle. This ordering affects the polar angles they make with the white circles inside the glandular region. Moreover, the epithelial cell nuclei are generally larger than the stromal cell nuclei. In order to utilize this information, area of the closest black circles are also considered. The same information is also employed for the closest white circles. Similarly, the white circles inside the gland body also have an ordering and size pattern, as opposed to the ones outside.

After clustering the white circles into two with respect to these features, we automatically determine the *gland circles* and *non-gland circles* using the following heuristic. For each of the clusters, we compute the average radius of the white circles assigned to this cluster. Then, we label the cluster with greater average radius as gland circles and the remaining as non-gland circles.

The results obtained by our seed detection step are illustrated on three sample images in Figure 3.7. As shown in these images, the appearance of a gland significantly changes when the sectioning angle changes; note that in all these images the magnifications and the lighting conditions remain the same. In the image in Figure 3.7 (a), the glands are circular and the epithelial cell nuclei, the epithelial cell cytoplasms, and the luminal area are conspicious. Unlike, in Figure 3.7 (c), glands appear relatively small. The epithelial cell cytoplasms and the luminal area inside the glandular regions are less apparent. In Figure 3.7 (e) due to the sectioning angle, taller and larger luminal areas appear.

In Figures 3.7(b), 3.7(d), and 3.7(f), the circles classified as *gland* are given in red and the remaining non-glandular circles are given in green. These results show



Ac: Area of the circle c
Aw1−5: Areas of 5−nearest white circles
Ab1−5: Areas of 5−nearest black circles
aw1−5: Polar angles between 5−nearest white circles
ab1−5: Polar angles between 5−nearest black circles
dw1−5: Distances between the circle and its 5−nearest white circles
db1−5: Distances between the circle and its 5−nearest black circles

Figure 3.6: 5-nearest neighbors of a white circle and the features extracted from these neighbors. Here, white and black circles are shown in green and in red, respectively.

that the proposed seed detection algorithm is considerably adaptive to changes in sectioning procedure. Note that here, there are some gland circles in some-nonglandular regions. Such regions (and the circles) are to be eliminated in the false gland elimination step.

After determining the white gland circles, we put these circles into the same seed group if the distance between their boundary pixels is smaller than a threshold (i.e., if their boundary pixels are close enough). Then we consider the centroids of each group as a seed pixel.

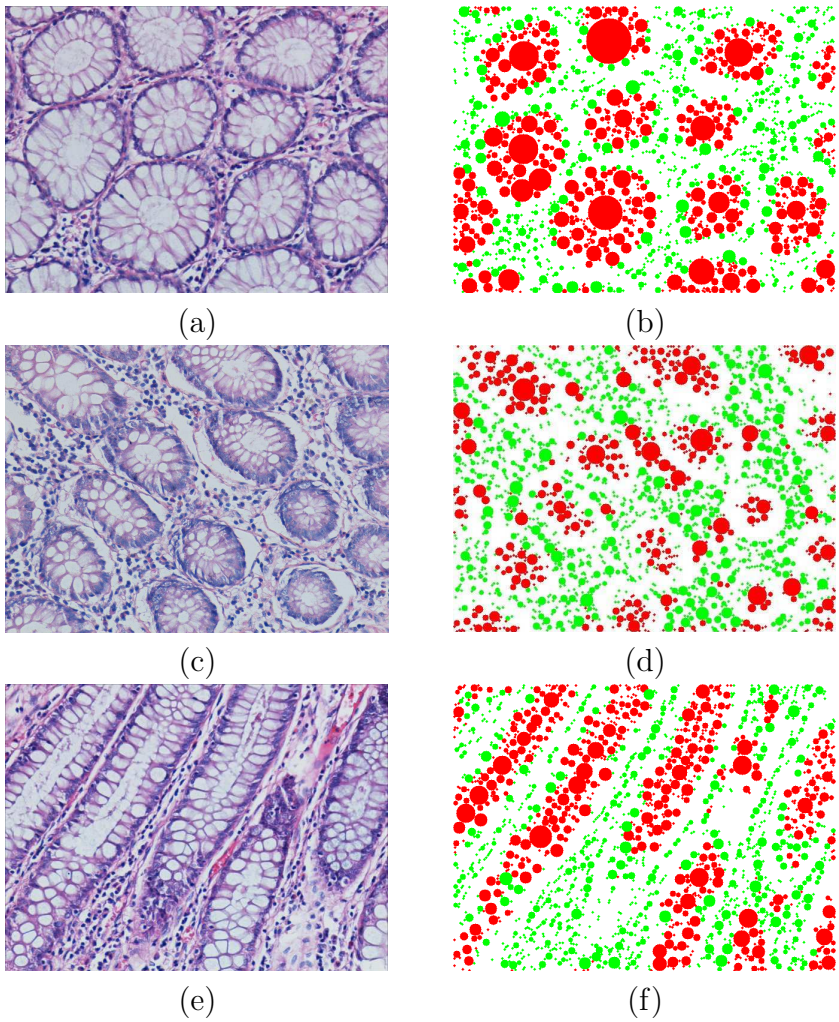


Figure 3.7: Three sample colon tissue images (a, c, e) and the classification results of the corresponding white circles (b, d, f). The red circles are classified as *gland* and the green ones are classified as *non-gland*.

3.3.3 Region Growing

Once we have found a seed pixel for each gland candidate, the next step is to find their initial inner borders. Our approach depends on the following characteristics of gland regions:

- A gland region is surrounded by a chain of epithelial cells. In the iterative double circle-fit transform of black cluster, there should be a chain of circles at the border of the gland body. These circles should be close to their

nearest neighbors.

- There is not any cell nucleus inside the gland body.After false nucleus elimination, this region is expected to be empty.

We exploit these properties to find the correct places to set the dams where the region growing is supposed to stop. We generate a graph of black circles where the nodes are circle centroids and edges are one-pixel wide straight lines that start from a circle centroid and end at another one. For each black circle c, the edges are assigned as follows: First the image is separated into four quadrants (see Figure 3.8). The circle c is connected to its K-nearest neighbor circles at each quadrant if these circles are in its N-nearest neighborhood. We define K as *quadrant neighborhood cardinality* and N as *total neighborhood cardinality*. In Algorithm 4, the algorithm that generates the nucleus network on a binary image (I_o) is given. The output image I_o is used in accordance with the seed pixels that are found in the previous step for finding the initial inner borders of gland candidates.

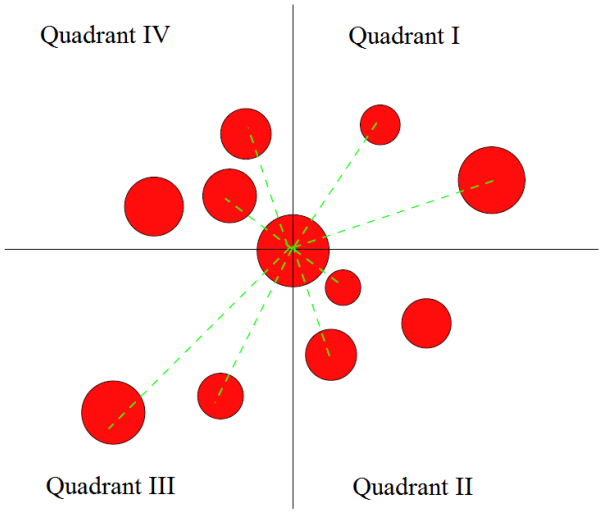


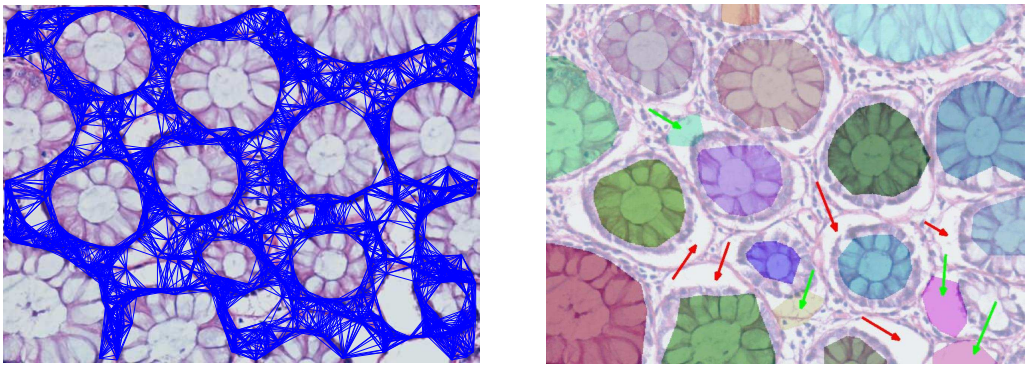
Figure 3.8: A circle centroid and its four quadrants.

The result of this algorithm on a sample image is shown in Figure 3.9. As shown in this image, typically, there are not any edges passing through these regions. These regions can be obtained as connected components by starting

Algorithm 4 Nucleus Network Generation Algorithm

```
1: function GenerateNucleusNetwork(I, K, N)
2:     //I: False nucleus elimination output
3:     //K: Quadrant neighborhood cardinality
4:     //N: Total neighborhood cardinality
5:     ImgSize ← SizeOf(I)
6:     //Create an empty image of the same size as I and initialize to 0
7:     I_o ← CreateEmptyImage(ImgSize)
8:     I_o ← 0
9:     for all c ∈ I do                          ▷ For each circle in the image
10:        for all c_i ∈ I & c_i ≠ c do
11:            dlist ← ‖c_i − c‖
12:        end for
13:        neighOrder ← Sort(dlist)
14:        numOfCirclesInQuad1 ← 0
15:        numOfCirclesInQuad2 ← 0
16:        numOfCirclesInQuad3 ← 0
17:        numOfCirclesInQuad4 ← 0
18:        for i ← 1, N do
19:            c_i ← neighOrder[i]
20:            //Create a line segment from the centroid of c to that of c_i
21:            l ← CreateLineSegment(CentroidOf(c), CentroidOf(c_i))
22:            //Calculate the slope angle of line l w.r. to positive x-axis
23:            A ← CalculateSlopeAngleOfLine(l)
24:            if A <= 90 & numOfCirclesInQuad1 < K then
25:                // Put one pixel wide line between the centroids of c and c_i
26:                DrawLine(I_o, CentroidOf(c), CentroidOf(c_i))
27:                numOfCirclesInQuad1++
28:            else if A > 90 & A <= 180 &
                       numOfCirclesInQuad2 < K then
29:                DrawLine(I_o, CentroidOf(c), CentroidOf(c_i))
30:                numOfCirclesInQuad2++
31:            else if A > 180 & A <= 270 &
                       numOfCirclesInQuad3 < K then
32:                DrawLine(I_o, CentroidOf(c), CentroidOf(c_i))
33:                numOfCirclesInQuad3++
34:            else if A < 270 & A < 360 &
                       numOfCirclesInQuad4 < K then
35:                DrawLine(I_o, CentroidOf(c), CentroidOf(c_i))
36:                numOfCirclesInQuad4++
37:            end if
38:        end for
39:    end for
40:    return I_o
41: end function
```

region growing from the seed pixels that are detected in the previous step and stopping at the edges embedded into the image I_o. The resultant initial inner borders obtained after region growing are shown as overlaid on the tissue image in Figure 3.9 (b). Note that here there are also some empty regions (e.g., the one shown with red arrow in Figure 3.9 (a)) for which no gland region is found since the white circles corresponding to these regions are labeled as non-gland.



(a) (b)

Figure 3.9: An example image, its nucleus network, and its gland candidates.

Among these gland candidate regions, some of them correspond to arbitrary empty regions. These non-glandular regions will be detected as false glands by means of a rule set that is generated in a supervised manner. The details are described in the next section.

3.4 False Gland Elimination

Among the gland candidates obtained in the previous step, some of them do not correspond to true glands (e.g., the ones shown in green arrows in Figure 3.9 (b)). These false glands can be considered as *false alarms*, hence they should be eliminated.

We approach false gland elimination as a supervised learning problem. After the region growing step, we have a number of initial gland regions. Let:

1. R_i be a gland region and r be the radius of R_i

2. C_i be the convex hull of the pixels inside the gland region R_i
3. K_i be the K pixels-wide dilation belt around the convex hull C_i
4. B_i be the set of pixels at the boundary of the gland region R_i.

We extract the following features from these regions:

1. Number of white pixels inside the gland region R_i
2. *Number of pink pixels inside the gland region* R_i
3. *Number of black pixels inside the gland region* R_i
4. *Percentage of white pixels inside the gland region* R_i
5. Percentage of pink pixels inside the gland region R_i
6. Percentage of black pixels inside the gland region R_i
7. Area of R_i
8. Number of white pixels inside the convex hull C_i
9. Number of pink pixels inside the convex hull C_i
10. *Number of black pixels inside the convex hull* C_i
11. Percentage of white pixels inside the convex hull C_i
12. Percentage of pink pixels inside the convex hull C_i
13. Percentage of black pixels inside the convex hull C_i
14. Area of C_i
15. *Number of white pixels inside the K-pixels wide dilation belt* K_i
16. Number of pink pixels inside the K-pixels wide dilation belt K_i
17. Number of black pixels inside the K-pixels wide dilation belt K_i

18. *Percentage of white pixels inside the K-pixels wide dilation belt* K_i
19. *Percentage of pink pixels inside the K-pixels wide dilation belt* K_i
20. *Percentage of black pixels inside the K-pixels wide dilation belt* K_i
21. Area of K_i
22. The standard deviation of the set:

$$S = \{\forall p \ \in \ B_i \ \|p - r\|\} \tag{3.1}$$

A training set is generated from the feature vectors of the gland candidates in a set of training images. Each gland candidate in the training set is manually labeled as *gland* and *non-gland*. Then this labeled set is used to train a *decision-tree* using c4.5 algorithm to obtain the corresponding rule set. This rule set is used to eliminate the false glands. The rule set that we use in our work consists of the following rules:

1. **Rule 1**: If number of pink pixels inside $R_i \leq 559$ and percentage of pink pixels inside $K_i \leq 34.098$, then the candidate is a false gland.
2. **Rule 2**: If percentage of white pixels inside $R_i > 6.872$ and number of black pixels inside $C_i \leq 1862$ and number of white pixels inside $K_i > 630$ and percentage of pink pixels inside $K_i \leq 35.189$, then the candidate is a false gland.
3. **Rule 3**: If number of pink pixels inside $R_i \leq 559$ and percentage of white pixels inside $K_i > 28.24$, and percentage of pink pixels inside $K_i \leq 42.789$, then the candidate is a false gland.
4. **Rule 4**: If number of pink pixels inside $R_i \leq 1509$ and percentage of pink pixels inside $K_i \leq 26.489$, then the candidate is a false gland.
5. **Rule 5**: If number of pink pixels inside $R_i \leq 2016$, and number of black pixels inside $R_i > 674$, and percentage of white pixels inside $R_i > 2.746$, and percentage of pink pixels inside $K_i > 26.469$, and percentage of black pixels inside $K_i > 26.761$, then the candidate is a false gland.

Note that in this rule set only eight features are used although we extract 22 features. These features are shown in italic fonts in the list. The decision tree selects the most discriminant features at the time of decision tree construction.

3.5 Detection of Gland Borders

Once the false glands are eliminated, the final step is to determine exact gland borders. For that, we define an algorithm (see Algorithm 5) that inputs 1) the regions of initial inner gland borders, 2) the iterative double circle-fit transform of the black cluster, 3) radius R of the dilation belt in which epithelial cell nuclei belonging to an inner gland region will be searched, and 4) a curve simplification factor N.

First, each region is dilated by a circular structural element. Then the black circles inside the dilated regions are found. These circles are sorted with respect to polar angles between the line segment from the region centroid to their centroid and the positive x-axis. This gives an ordered set of points (circle centroids), hence a simple polygon. This polygon is then simplified by connecting its vertices in their N-neighborhood in the vertex order (S_{io}) (lines 30-35) and filling the inner region (line 36). The value of N adjusts the level of simplification. When N is set to 1, the output will be the image of the particular simple polygon. When N is set to the number of nuclei found in the region, the output will be the convex hull of the polygon. In the algorithm, I_{res} is the image that contains the glandular regions.

The result of this algorithm is visually illustrated in Figure 3.10. In this image, the segmented glands are embedded in the tissue image and each segmented gland is shown with a different color.

Algorithm 5 Exact Gland Border Detection Algorithm

function DETECTEXACTGLANDBORDERS(I_{fgo}, I_{nc}, R, N)
    //I_{fgo}: False gland elimination output
    //I_{nc}: The circle-fit transform of black cluster
    //R: Radius of the dilation belt
    //N: Simplification factor
    //For each connected component label r in the image
    for all $r \in I_{fgo}$ **do**
        $I_r \leftarrow (I_{fgo} = r)$
        $se \leftarrow$CreateCircularStructuralElementOfRadius(R)
        $I_d \leftarrow$Dilate(I_r, se)
        //Set of black circles inside the region
        $S_i \leftarrow \emptyset$
        //Set of black circles ordered with respect to polar angles
        $S_{io} \leftarrow \emptyset$
        $ImgSize \leftarrow SizeOf(I_{fgo})$
        $I_{res} \leftarrow CreateEmptyImage(ImgSize)$
        for all $c \in I_{nc}$ **do**
            //For each circle c
            if $c \cap r \neq \emptyset$ **then** $S_i \leftarrow S_i \cup c$
            end if
        end for
        for all $c_i \in S_i$ **do**
            //For each circle c_i
            //Find polar angle of c_i about the centroid of c
            $dlist[c_i] \leftarrow$CalculatePolarAngleAboutPoint(c_i, c)
        end for
        //Sort the circles in S_i with respect to polar angles
        $S_{io} \leftarrow$Sort($dlist$)
        $I_{nh} \leftarrow CreateEmptyImage(ImgSize)$
        for all $c_i \in S_{io}$ **do**
            //For each circle c_i
            for $k = 1 : N$ **do**
                $DrawLine(I_{nh}, c_i, c_{(i+k \bmod N)})$
            end for
        end for
        $FillRegion(I_{nh})$
        $I_{res} \leftarrow I_{res} \cup I_{nh}$
    end for
    return I_{res}
end function

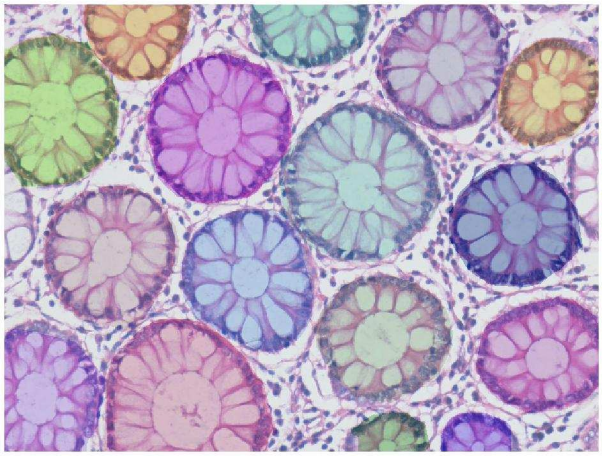


Figure 3.10: An example image and the exact borders of detected glands.

Chapter 4

Experiments

In this chapter, we explain our experiments on the histopathological images of colon biopsies and discuss their results.

4.1 Experimental Setup

The dataset that we use in the experiments consists of 72 colon tissue images taken from the biopsy samples of 36 patients. The images are acquired using a 20× microscope objective lens. The hematoxylin-and-eosin staining technique is used. As stated in Section 3.4, false glands are eliminated with respect to a rule set generated by training a decision tree classifier. From the images in the dataset, 24 images of 12 patients are used as the *training set* to train the decision tree classifier. In order to represent the largest possible input space, sample images with different visual characteristics are selected. The 48 images of the remaining patients are used as the *test set*.

The experimental system is developed in ANSI C programming language, then deployed on a UNIX-based server with two quad-core CPUs and 4 GB of main memory. The processing of an image of resolution 480×640 lasts about 180 seconds on the server.

4.2 Results

4.2.1 Parameter Selection

Our proposed method involves a set of input parameters. Those parameters need to be set to optimum values in order to achieve the best performance. In this work, the actual parameter values are selected by experimentation and visual examination on the training samples. The descriptions of the input parameters, their chosen values, and their effects to the segmentation accuracy are explained below:

1. *Circle area threshold*: Minimum area of a circle that is produced by the iterative double circle-fit transform. The circles with areas under this threshold are eliminated. In the experiments, this threshold is set to 10 pixels for both black and white clusters. As circle area threshold increases, iterative double circle-fit transform produces less and greater circles. When applied to black cluster, a greater threshold filters a large number of black circles. This reduces the performance since some epithelial nuclei are eliminated. A smaller threshold produces additional tiny circles, which behave as noise and reduce the performance. On the other hand, when applied to white cluster, increasing the threshold does not affect the performance until a significantly large value is reached. A smaller threshold again produces additional tiny circles, which distorts the pattern of white circles in a glandular region, thus reduces performance.

2. *Neighbor count*: In the seed detection step, a white circle is classified as gland or non-gland based on the features extracted from the spatial relationships between this white circle and its N-nearest white and N-nearest black circles. Neighbor count N gives the number of the nearest neighbors. In our experiments, 5-nearest neighbors of both black and white circles are considered for each white circle. Extracting the features of a larger number of nearest white and black circles does not increase the overall performance.

When the number of features becomes significantly large, curse of dimensionality occurs, hence the performance decreases.

3. *Gland circle grouping proximity threshold*: After white circles are classified as *gland* and *non-gland* circles, they are grouped into components with respect to an Euclidean proximity measure. If the smallest distance between the boundary pixels of a pair of gland circles is smaller than this threshold, these circles are assigned to the same component. In the experimental configuration, a threshold of 2 pixels is chosen. A larger threshold would connect farther white circles, hence produces larger groups. This reduces the overall pixel-wise segmentation performance, since the connectivities of white circles would flood out of actual border of small glands. A smaller threshold, such as 1 pixel, may produce more than one connected component inside a gland, hence lead to detection of two glands instead of one.

4. *Isolation factors*: During false nucleus elimination, the black circles that are isolated from other black circles are considered as noise and eliminated. Degree of isolation of a circle is determined by the distance to its nearest neighbor and the total distance to its two-nearest neighbors. These distances are compared to a linear combination of the average and standard deviation of distances of circles to their one-nearest and two-nearest neighbors and if these distances are larger than the corresponding values for a black circle c, the circle is eliminated. The coefficients of this combination (K_1 and K_{12}) are defined as *isolation factors*. As these factors increase, the black circles that are more significantly isolated from their neighbors will be considered as false nuclei and will not be eliminated. In the opposite case, more black circles will be eliminated, since a looser isolation criterion is configured. In the experiments, K_1 is set to 2.3 and K_{12} is set to 2.5.

5. *Dilation belt radius*: The dilation belt around an initial inner gland region is used during both false gland elimination and detection of gland borders. During the false gland elimination step, a set of features is extracted from the dilation belts around the glands. These features identify the structure of the dams surrounding the glands. As the radius of the structural element

increases, the dilation belt will include the stromal region, hence the corresponding features will no longer be helpful for identification of the dam structure. The same will be true when a significantly small radius is selected, since in this case the features will represent only a portion of the dam.

In detection of gland borders, the epithelial cell nuclei belonging to an inner gland region are searched in the dilation belt with radius R around the region. When a large radius is selected, some stromal cell nuclei may be considered as the epithelial cell nuclei in the dam. This causes flooding of the detected borders into stromal region, hence reduces pixel-wise segmentation performance. A small radius entails misdetection of the cell nuclei in the dam, which reduces performance. In our experiments, considering the resolution of images that we work on, we select this radius as 7 pixels.

6. *Quadrant and total neighborhood cardinalities*: During nucleus graph preparation in the region-growing step, a black circle c is connected to its K-nearest neighbor circles at each quadrant if these circles are in its N-nearest neighborhood. The integers K and N are defined as *quadrand neighborhood cardinality* and *total neighborhood cardinality*, respectively. As K and N increase, connectivity in the nucleus graph increases, thus smaller initial inner gland regions occur at the end of region growing. Smaller K and N values cause the occurrence of a less connected nucleus graph, hence region growing produces larger initial inner gland regions. In our experiments, we set K to 2 circles and N to 15 circles.

7. *Simplification factor*: At the gland border detection step, the black circles around a gland are identified and ordered with respect to their polar angles from the centroid of the gland. Such an ordering gives a simple polygon, which does not correspond to exact gland borders. A simplified form of this polygon gives the gland borders more accurately. The polygon is simplified by connecting its each vertex to its K-nearest neighbors by straight lines and filling the resultant region. The factor K gives the degree of simplification, hence referred to as *simplification factor*. If K is selected to be the number of black circles around the gland, it will give the convex hull of the simple

polygon, which may not provide the correct gland borders since a glandular region is not necessarily convex. A small K will lead to a small degree of simplification, which is also insufficient for detection of exact borders. In our experiments, we select K to be 6.

8. *Border threshold*: The edges of the nucleus graph are expected to form closed dams around the gland borders. Region growing iterations stop at these edges, thus the inner regions of the glands are detected. An exceptional case occurs for the glands at the image boundaries. Polar ordering of the epithelial cell nuclei around the gland centroid does not give the inner region since only a subset of the nuclei at the dam are included in the image. In order to handle this exception, the black circles with vertical or horizontal pixel distance to the image border closer than this threshold are connected to the borders by vertical or horizontal edges, respectively. This threshold is set to 30 pixels in the experiments, considering the image resolution. A larger threshold connects larger number of inner circles to image borders. This causes separation of glands close to image borders. This would reduce both pixel-based segmentation and detection performance, since two small glands will be detected instead of a single gland. A smaller threshold will connect less circles to borders. This time, the exception may remain for some glands at the borders.

4.2.2 Segmentation Results

In Figures 4.1 to 4.8, segmentation results of 8 sample images are given. First two of these samples (Figures 4.1 and 4.2) are taken from the training set and the remaining 6 samples (Figures 4.3 to 4.8) are taken from the test set. In each sample, the manually segmented glands (*gold standard*) are given in (a), the result of our proposed method is given in (b), the result of the region-growing based method proposed by Wu et al. [16] is given in (c) and the result of the directional filtering based method proposed by Wu et al. [17] is given in (d). Both of these methods use pixel-based information in the segmentation of glands. In contrast, our method uses object-based information in gland segmentation. For both the

region-growing based method and the directional filtering based method, some parameters should be manually selected. For these methods, we conduct the experiments over a large range of these parameters and for each image, we use the parameters that yield the best segmentation performance and report these best performance values. Note that although we optimize the parameters for each image for these two methods, we use the same set of parameters for all images in our proposed object-based method.

In the image shown in Figure 4.1, the glands appear in relatively oval shape. However, the connective tissue between the glands is not apparent due to staining artifacts. Our method detects all the glands but the ones at the very boundary of the image (marked with red arrows in Figure 4.1 (b)). The region-growing based method detects most of the glands (marked with green arrows in Figure 4.1 (c)). Nevertheless, it is unable to eliminate the false glands on the connective tissue (red arrows), all of which are eliminated in our method. Moreover, although the nuclei boundaries are found for some of the glands (blue arrows), the entire gland could not be detected as these boundaries do not form a closed dam (e.g., for the ones shown in Figure 4.1 (c), the existence of gaps on the gland boundaries cause not to find the entire glands). The directional filtering based method does not produce such false glands, but in some cases one gland is detected instead of two (e.g., for the ones shown with yellow arrows in Figure 4.1 (d), the algorithm detects two individual glands as one). The borders of none of the glands are detected as accurately as our method.

In Figure 4.2, another example with staining artifacts is demonstrated. Besides, in this image, there are also glands with different sizes. As in the previous sample, the connective tissue is not apparent. Our method detects the gland located at the top left corner (marked with the yellow arrow) as two separate glands and misses another gland (marked with the red arrow in Figure 4.2 (b)). The region-growing based method detects three glands successfully (green arrows), including the one at the top left corner, however there remains many false glandular regions (red arrows) as in the previous sample. Similarly, although the borders of some glands are partially detected, they could not be entirely segmented as there are gaps in the detected borders (marked with blue arrow in Figure 4.2 (c)). One

gland is detected with a flooded part (marked with the yellow arrow in Figure 4.2 (c)) which leaks out to the border of the neighboring misdetected gland. In the directional filtering based method, two glands are detected successfully. However there exists a very large region containing six glands (marked with the yellow arrow in Figure 4.2 (d)). Considering these samples, it can be observed that our method is less sensitive to staining artifacts than its pixel-based counterparts.

As opposed to the previous images, in the next image (Figure 4.3) the connective tissue is apparent and glands have proper shapes and sizes. Our method detects all the glands but the one located at the top right corner and the one located at the very bottom of the image (marked with red arrows in Figure 4.3 (b)). However, the region-growing based method detects two of the glands with entirety (marked with green arrows in Figure 4.3 (c)). It also detects one gland with a part flooded to the neighboring gland as in the previous sample (marked with the red arrow in Figure 4.3 (c)). The directional filtering based method does not detect any glands with proper borders (Figure 4.3 (d)). Similar to the image given in Figure 4.2 (d), there exists a very large region spreading around the image, which corresponds to segmentation of none of the glands. Here, this large spread is attributed to the difficulty of determining a threshold value that can be applied to all of the glands. Note that although the directional filtering based method finds partial glands, this gland segmentation could not be used for further analysis as it is not able to separate the different glands. For further analysis (for example the diagnosis and grading based on the features extracted for a gland), the correct separation of individual glands is very important.

The image given in Figure 4.4 is taken in similar sectioning angle and lighting conditions as the previous one. Our method is able to detect most of the glands with almost exact borders, except two misdetected glands (marked with red arrows Figure 4.4(b)) at the bottom and top right corners; again these glands are located on boundaries and most of them are outside of the image. Additionally, there is one gland at the bottom left of the image. Although this gland is successfully detected, there are small problems at its boundaries. For this image, the region-growing based method can only detect one gland with exact borders (marked with green arrow in Figure 4.4 (c)) and detects another gland for which

there are some problems in the detection of its surrounding nuclei; here the growing process of the nuclei does not stop properly. The directional filtering based method is not able to detect any true glands; it detects one of the glands entirely, but the boundaries are not identified successfully (marked with yellow arrow in Figure 4.4 (d)). As shown in the last two examples, our method exhibits a significantly superior performance regardless of the difference in gland sizes and the staining conditions.

For the images in Figures 4.5 and Figure 4.6, our method detects most of the glands with correct borders. While the region-growing based method misses most of the glands, the directional filtering based method produces many partial illegitimate segmentations and some other segmentations that include the regions of more than one gland.

In the image shown in Figure 4.7, the glands appear smaller and the epithelial cytoplasms are less apparent compared to the previous ones. As shown in Figure 4.7 (b), our method detects almost all of the glands in such an image. However, the region-growing based method detects only one gland with correct borders (marked with green arrow in Figure 4.7 (c)). Here the incorrect segmentation of the region-growing based method is attributed to the small gland sizes. In the region-growing based method, the growing process starts with the large circular white region. For this image, such white regions are not found in many of the glands. The directional filtering based method detects only a small subset of glands producing many additional regions that do not correspond to a gland.

In the image given in Figure 4.8, the connective tissue and the nuclei are apparent, but there exist empty white regions surrounding the glands due to the sectioning procedure. Our method detects all the glands except the one on the top left corner only a very small portion of which is in the image scene. The region-growing based method also detects most of the glands for this image since the region-growing process could stop properly as the surrounding nuclei are relatively more apparent; however the directional filtering based method does not detect any glands with correct borders.

To obtain a quantitative evaluation of the performance of our proposed

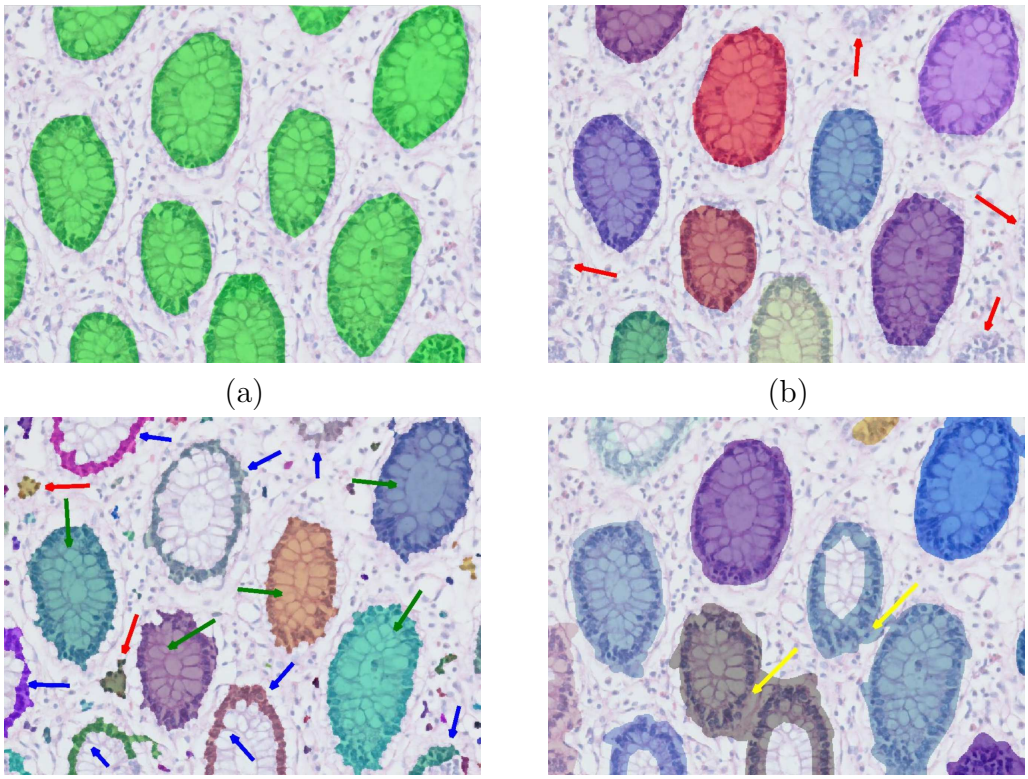


Figure 4.1: On an exemplary tissue image: (a) gold standard which is obtained by manual segmentation, (b) the segmentation result of our proposed method, (c) the segmentation result of the region growing based method proposed in [16], and (d) the segmentation result of the directional filtering based method proposed in [17].

method, we calculate several measures. We also calculate the same measures for two pixel-based methods, which are proposed by Wu et al. [16, 17], and compare the success of our proposed method with those of these methods.

In our experiments, for each pixel, the segmentation result of the method is compared with the gold standard. A pixel that corresponds to a glandular region in both the segmentation result and the gold standard is defined as true positive (TP). A pixel that corresponds to a glandular region in the gold standard but not in the segmentation result is defined as false negative (FN). A pixel that corresponds to a glandular region in the segmentation result but not in the gold standard is defined as false positive (FP). A pixel that does not correspond to a glandular region in both the segmentation result and the gold standard is

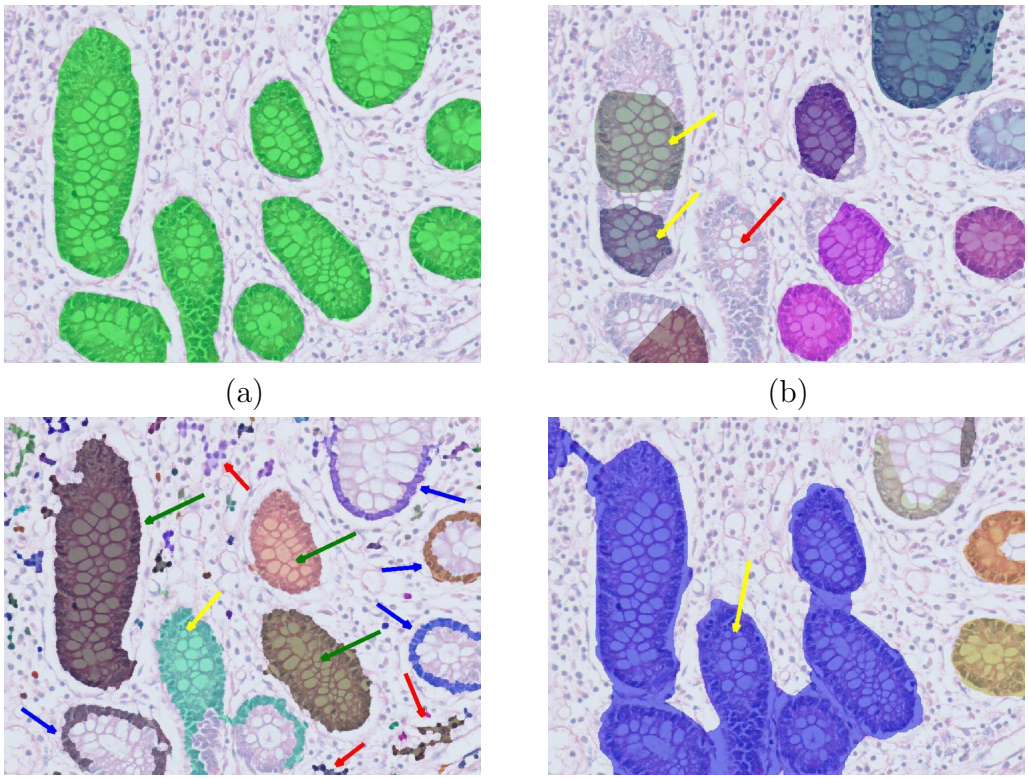


Figure 4.2: On an exemplary tissue image: (a) gold standard which is obtained by manual segmentation, (b) the segmentation result of our proposed method, (c) the segmentation result of the region growing based method proposed in [16], and (d) the segmentation result of the directional filtering based method proposed in [17].

defined as true negative (TN). When evaluating the performance, we refer to the following three metrics:

- *Sensitivity*: The probability that a test will produce a true positive result on the pixels that belong to a glandular region determined by the gold standard ($\frac{TP}{TP+FN}$).
- *Specificity*: The probability that a test will produce a true negative result on the pixels that belong to a non-glandular region determined by the gold standard ($\frac{TN}{TN+FP}$).
- *Accuracy*: The probability that the system will produce a correct classification as compared to the gold standard ($\frac{TP+TN}{TP+FN+FP+TN}$).

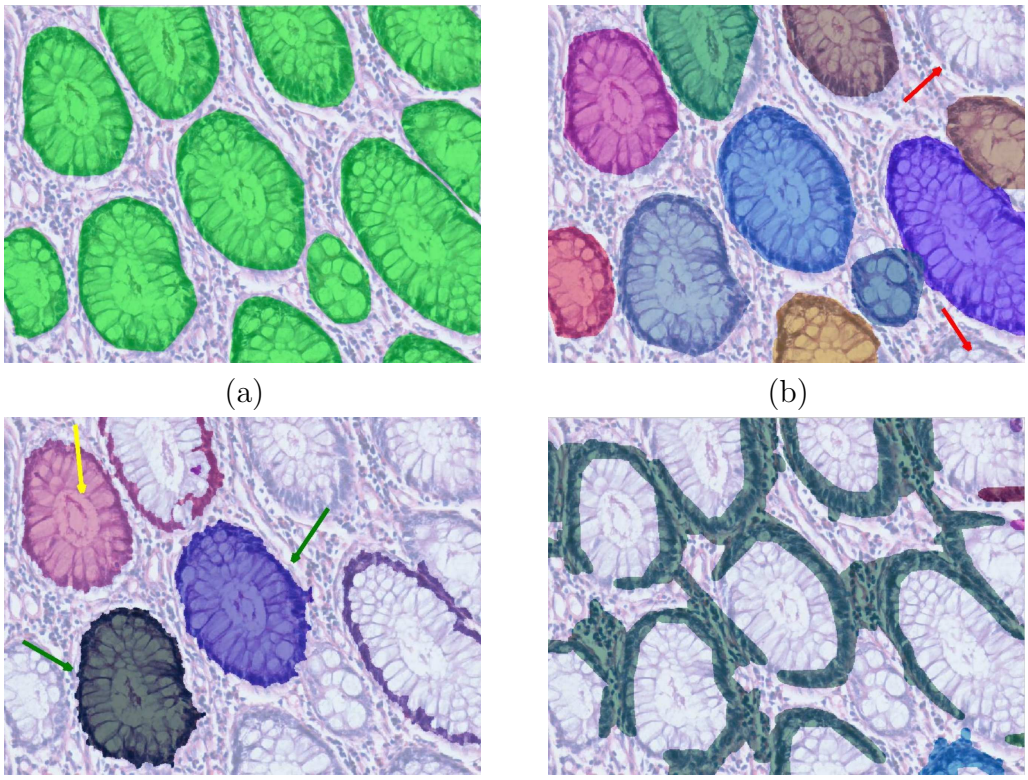
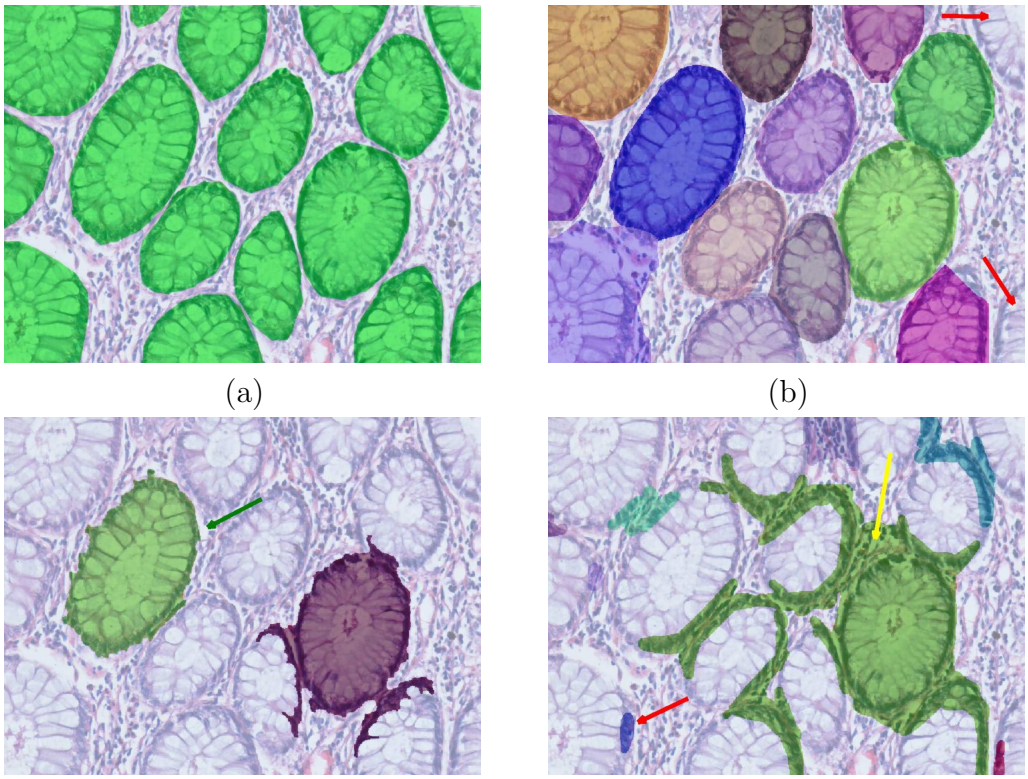


Figure 4.3: On an exemplary tissue image: (a) gold standard which is obtained by manual segmentation, (b) the segmentation result of our proposed method, (c) the segmentation result of the region growing based method proposed in [16], and (d) the segmentation result of the directional filtering based method proposed in [17].

The average performances of our object-based method and the previous pixel-based methods and their standard deviations are reported for training and test sets in Tables 4.1 and 4.2, respectively. As shown in these tables, our method provides performances above 80 percent with reasonable standard deviation in all metrics. The sensitivity of the region-growing based method is as low as 22 percent where its specificity is 95.3 percent, which means that the method eliminates most of the regions as false glands and rarely detects a true gland. The accuracy is still 52.5 percent thanks to the highest true negative rate. The directional filtering based method exhibits performances about 50 percent for sensitivity, specificity, and accuracy. This implies that it detects more true glands but also more false glands compared with the region-growing based method. The statistics given



(a) (b)

(c) (d)

Figure 4.4: On an exemplary tissue image: (a) gold standard which is obtained by manual segmentation, (b) the segmentation result of our proposed method, (c) the segmentation result of the region growing based method proposed in [16], and (d) the segmentation result of the directional filtering based method proposed in [17].

in the tables reveal that our method greatly improves the performances of the pixel-based methods. In order to understand whether or not this improvement is significant, we use a statistical test. Using the Wilcoxon test with a significance level of 0.1, the results show that the segmentation accuracy of our method are significantly better than those of the other two methods. The performance result of each image in the training set and the test set is given in Table A.1 and Table A.2 in the appendix, respectively.

The fact that the sensitivity of the directional filtering based method is higher than 50 percent (see Tables 4.1 and 4.2) is inconsistent with visual observation, considering the visual examples given in Figures 4.1 (d)-4.8 (d) where only a few

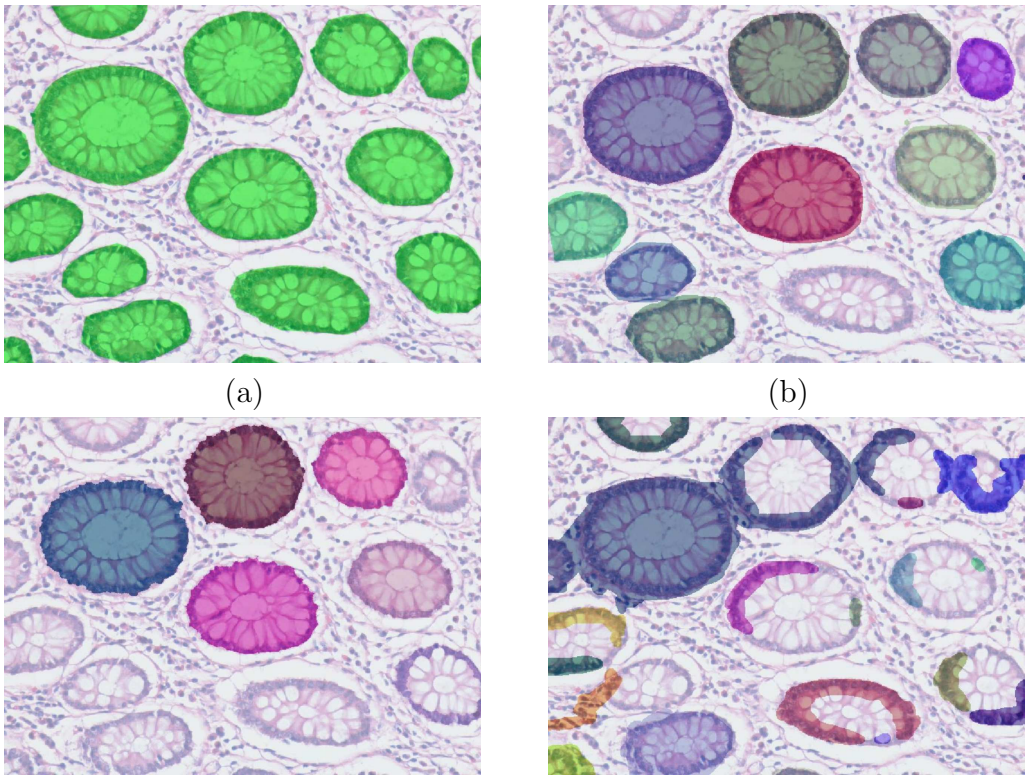


Figure 4.5: On an exemplary tissue image: (a) gold standard which is obtained by manual segmentation, (b) the segmentation result of our proposed method, (c) the segmentation result of the region growing based method proposed in [16], and (d) the segmentation result of the directional filtering based method proposed in [17].

glands are successfully detected. This is due to the nature of pixel-based performance evaluation scheme, in which gland-level identification is not considered. This causes problems in performance evaluation of results as given in Figure 4.2 (d) where a very large component (the one shown with yellow arrow in the figure) spans multiple number of glands. Since all of the pixels corresponding to a gland in the gold standard are considered as true positive, such large components severely increase performance although none of the glands are properly segmented. In order to obtain a more reliable evaluation of performance at pixel level, we make two modifications to the pixel-based evaluation method:

1. Each connected component in the resulting image corresponds to only one

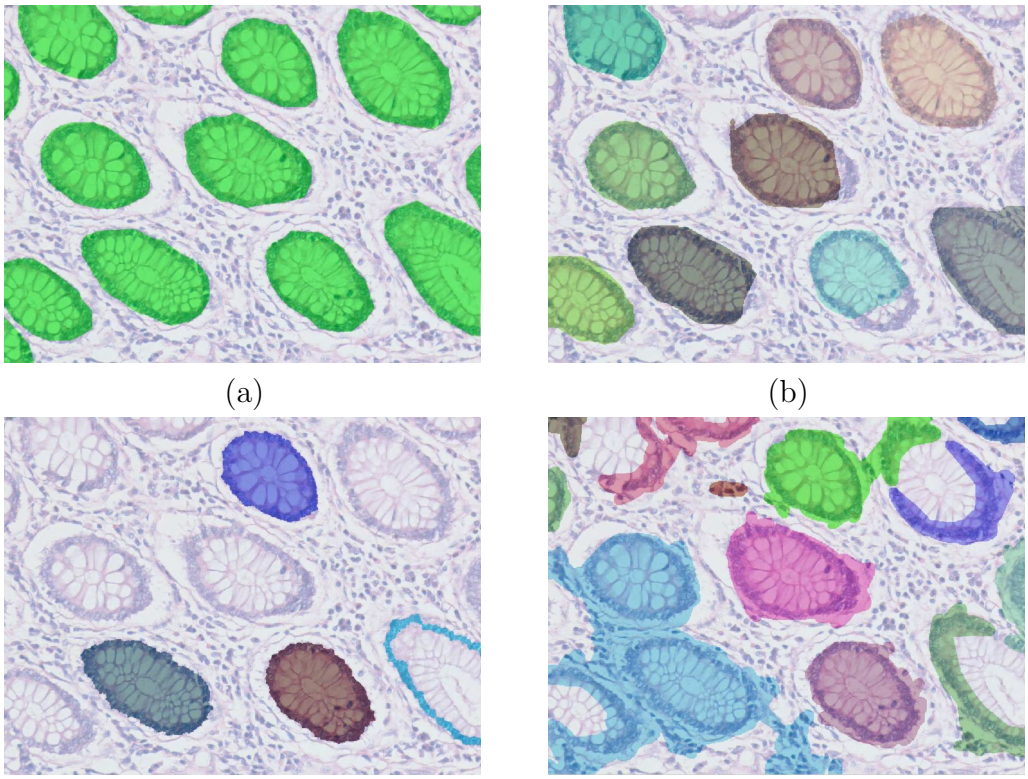


Figure 4.6: On an exemplary tissue image: (a) gold standard which is obtained by manual segmentation, (b) the segmentation result of our proposed method, (c) the segmentation result of the region growing based method proposed in [16], and (d) the segmentation result of the directional filtering based method proposed in [17].

gland, and thus, the remaining pixels belonging to other glands contained by the connected component will be considered as false positive.

2. A gland is considered to be detected if its centroid is contained by a connected component. The components that intersect with the gland region but does not contain its centroid are considered as false positive. For example, for the component shown with blue arrow in Figure 4.8 (d), the pixels are not considered to be correctly segmented.

The average performances of the methods and their standard deviations calculated with this modified pixel-based scheme are reported for training and test sets in Tables 4.3 and 4.4, respectively. The tables show that the performance of our

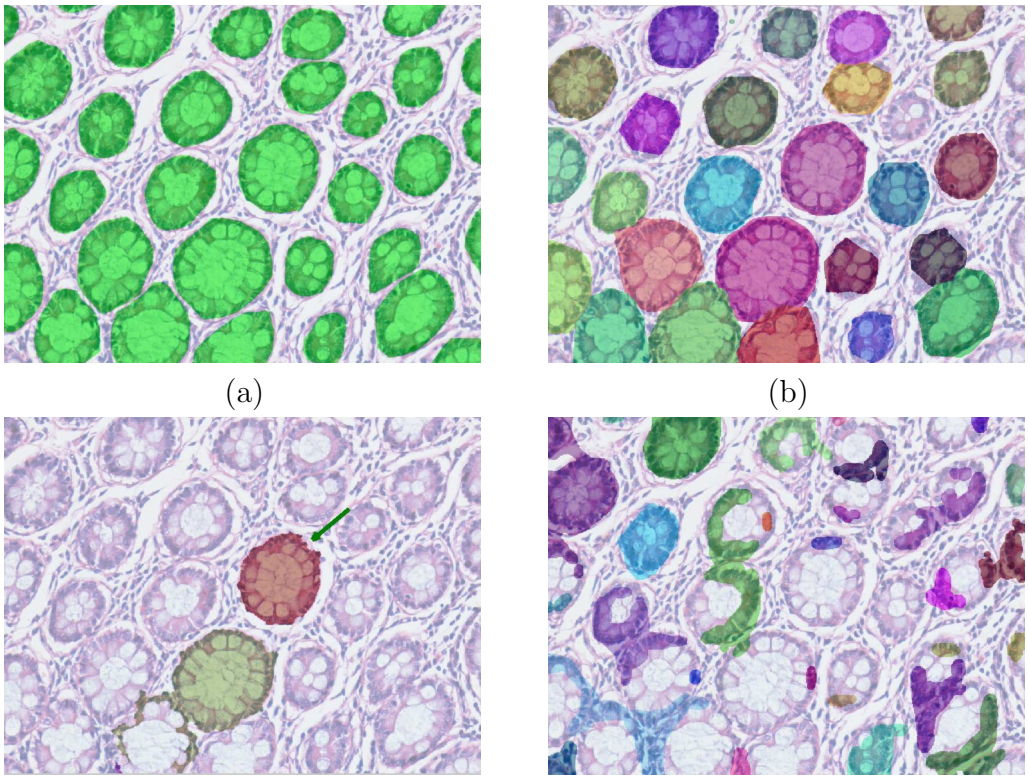


Figure 4.7: On an exemplary tissue image: (a) gold standard which is obtained by manual segmentation, (b) the segmentation result of our proposed method, (c) the segmentation result of the region growing based method proposed in [16], and (d) the segmentation result of the directional filtering based method proposed in [17].

method almost remains the same, while the performances of alternative methods, especially the filtering based method, severely decrease. When the Wilcoxon test with a significance level of 0.1 is applied, the superiority of the segmentation accuracy of our method to both alternatives remains statistically significant for this performance evaluation scheme. The performance of each image in the training and the test sets is given in Table A.3 and A.4, respectively.

The gland segmentation performance of a method can be calculated considering the percentages of glands that the method identifies with legitimate borders, those that it misses, and those it identifies with unreasonable borders. This provides a higher level means for performance evaluation than the pixel-based approach. The gland-based segmentation results for the training and test sets

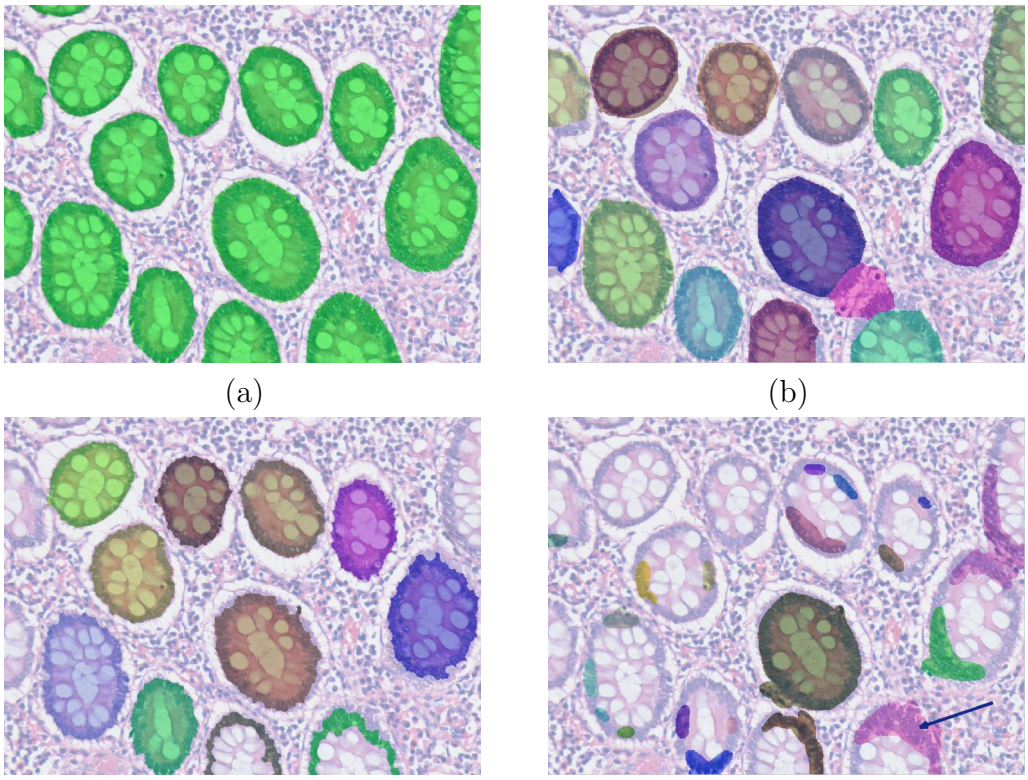


Figure 4.8: On an exemplary tissue image: (a) gold standard which is obtained by manual segmentation, (b) the segmentation result of our proposed method, (c) the segmentation result of the region growing based method proposed in [16], and (d) the segmentation result of the directional filtering based method proposed in [17].

are reported in Tables 4.5 and 4.6, respectively. In these tables, the average of the sensitivity, specificity, and accuracy values and their standard deviations are given. The corresponding performances for each image are also reported in Appendix; in Table A.5, for the images in the training set, and in Table A.6, for the images in the test set. In Tables 4.5 and 4.6, glands that are detected with legitimate borders are considered as truly identified glands and the corresponding percentages for our method and the other two methods are reported in the columns with header *True Gland Percentage*. The glands for which illegitimate borders, such as ones that flood out to other glands (e.g., the ones shown with yellow arrows in Figure 4.2 (c)) or multiple tiny glands inside one gland (e.g., the ones shown with blue arrows in Figure 4.1 (c)), are defined to be glands with

	Sensitivity	**Specificity**	**Accuracy**
Our Method	0.839 ± 0.126	0.880 ± 0.067	0.871 ± 0.048
Region-Growing	0.244 ± 0.221	0.953 ± 0.051	0.544 ± 0.138
Directional Filtering	0.517 ± 0.328	0.646 ± 0.351	0.557 ± 0.188

Table 4.1: The pixel-based performance of our method and the methods proposed by Wu et al. [16, 17]. Here, we report the average sensitivity, specificity, and accuracy values obtained on the training set and their standard deviation.

	Sensitivity	**Specificity**	**Accuracy**
Our Method	0.863 ± 0.083	0.805 ± 0.105	0.844 ± 0.069
Region-Growing	0.220 ± 0.184	0.953 ± 0.060	0.525 ± 0.124
Directional Filtering	0.552 ± 0.335	0.522 ± 0.387	0.554 ± 0.118

Table 4.2: The pixel-based performance of our method and the methods proposed by Wu et al. [16, 17]. Here, we report the average sensitivity, specificity, and accuracy values obtained on the test set and their standard deviation.

misdetected borders. Their percentage of occurrence for the three methods are reported in the columns with header *Illegitimate Border Detection Percentage*. The glands with no regions are detected or the glands only a small portion of which are detected are considered as misdetected glands (e.g., the ones shown with red arrows in Figure 4.4 (d)). The corresponding percentages are reported in the columns with header *Misdetected Gland Percentage*. The ratio of false alarms to all alarms are defined as false positive percentage and reported in the columns with header *False Alarm Percentage*.

The results in Tables 4.5 and 4.6 demonstrate that the average true gland detection performance of our method is significantly higher than the previous pixel-based methods; 81.0 percent for the training set and 73.1 percent for the test set, while the other two methods perform lower than 16 percent for both training and the test sets. The percentage of gland detections with illegitimate borders is as low as 3.6 percent in our method, whereas in the directional-filtering method this percentage is larger than 78 percent. Our method misses 23.3 percent of the glands, while this value is 71.3 percent in the region-growing based method. This implies that the region-growing based method misses a very large percent of the glands. The filtering-based method detects a very large percent of the glands with wrong borders. In contrast, our method truly detects most of the glands.

	Sensitivity	Specificity	Accuracy
Our Method	0.833 ± 0.132	0.864 ± 0.071	0.862 ± 0.058
Region-Growing	0.172 ± 0.181	0.894 ± 0.115	0.493 ± 0.132
Directional Filtering	0.168 ± 0.246	0.537 ± 0.295	0.351 ± 0.247

Table 4.3: The modified pixel-based performance of our method and the methods proposed by Wu et al. [16, 17]. Here, we report the average sensitivity, specificity, and accuracy values obtained on the training set and their standard deviation.

	Sensitivity	Specificity	Accuracy
Our Method	0.844 ± 0.097	0.800 ± 0.101	0.829 ± 0.084
Region-Growing	0.179 ± 0.166	0.918 ± 0.091	0.494 ± 0.126
Directional Filtering	0.094 ± 0.102	0.433 ± 0.330	0.274 ± 0.177

Table 4.4: The modified pixel-based performance of our method and the methods proposed by Wu et al. [16, 17]. Here, we report the average sensitivity, specificity, and accuracy values obtained on the test set and their standard deviation.

According to the Wilcoxon test with significance level of 0.1, our method yields significantly more accurate results than its pixel-based counterparts.

	True Gland Percentage	**Illegitimate Border Detection Percentage**	**Misdetected Gland Percentage**	**False Alarm Percentage**
Our Method	0.810 ± 0.171	0.011 ± 0.056	0.179 ± 0.147	0.046 ± 0.101
Reg. Grow.	0.158 ± 0.266	0.177 ± 0.200	0.665 ± 0.347	0.495 ± 0.416
Dir. Filt.	0.143 ± 0.219	0.754 ± 0.229	0.104 ± 0.121	0.593 ± 0.378

Table 4.5: The gland-based performance of our method and the methods proposed by Wu et al. [16, 17]. The average true gland detection percentage, the percentage of glands detected with illegitimate borders, the percentage of misdetected glands, and the percentage of false alarms and their standard deviation are reported for the training set.

In the visual examples given in Figures 4.1 to 4.8, it can be visually observed that some of the glands that are close to the image borders and that are partially in the scene are misdetected. This is as expected since the assumptions are made for the glands completely inside the scene, and these assumptions may not be valid for partially apparent ones. As an example, the luminal area in the middle of the gland may be out of the scene (e.g., the glands shown in red arrows in Figure 4.1 (b)). Furthermore, for only such glands, only some of the

	True Gland Percentage	**Illegitimate Border Detection Percentage**	**Misdetected Gland Percentage**	**False Alarm Percentage**
Our Method	0.731 ± 0.196	0.036 ± 0.106	0.233 ± 0.125	0.086 ± 0.109
Reg. Grow.	0.100 ± 0.150	0.187 ± 0.248	0.713 ± 0.253	0.455 ± 0.405
Dir. Filt.	0.061 ± 0.095	0.782 ± 0.218	0.157 ± 0.201	0.522 ± 0.393

Table 4.6: The gland-based performance of our method and the methods proposed by Wu et al. [16, 17]. The average true gland detection percentage, the percentage of glands detected with illegitimate borders, the percentage of misdetected glands, and the percentage of false alarms and their standard deviation are reported for the test set.

epithelial cell nuclei are apparent in the scene. In order to evaluate the effect of misdetection of such glands to overall performance, we re-evaluate the pixel-based performances by excluding the glands with centroids closer to the image borders than 40 pixels and reported the average performances and the standard deviations in Tables 4.7 and 4.8 for the training and test sets, respectively. The corresponding performances for each image are also given in Tables A.7 and A.8, respectively. The tables show that when the glands close to borders are excluded, the performances of all methods increase. However, the difference in the results of our method and those of the others still remain significant with respect to the Wilcoxon test with significance level of 0.1.

	Sensitivity	**Specificity**	**Accuracy**
Our Method	0.847 ± 0.137	0.871 ± 0.070	0.873 ± 0.058
Region-Growing	0.176 ± 0.184	0.900 ± 0.111	0.509 ± 0.136
Directional Filtering	0.156 ± 0.246	0.702 ± 0.251	0.420 ± 0.211

Table 4.7: The modified pixel-based performance of our method and the methods proposed by Wu et al. [16, 17]; the glands with centroids closer to the image boundaries than 40 pixels are excluded. Here, we report the average sensitivity, specificity, and accuracy values obtained on the training set and their standard deviation.

	Sensitivity	Specificity	Accuracy
Our Method	0.865 ± 0.102	0.814 ± 0.095	0.844 ± 0.083
Region-Growing	0.184 ± 0.171	0.926 ± 0.084	0.511 ± 0.129
Directional Filtering	0.059 ± 0.095	0.737 ± 0.283	0.375 ± 0.133

Table 4.8: The modified pixel-based performance of our method and the methods proposed by Wu et al. [16, 17]; the glands with centroids closer to the image boundaries than 40 pixels are excluded. Here, we report the average sensitivity, specificity, and accuracy values obtained on the test set and their standard deviation

4.2.3 Discussion

The proposed method can be used as an infrastructure for further biopsy analysis. For example, such an infrastructure could be used for detecting cancerous glandular structures. In colon tissues with low-grade cancer, the structures of glands are distorted. Their shapes turn into arbitrary convex curves. The amount of epithelial cell cytoplasms decreases, while the number of epithelial cells increases. Hence, the nuclear dams around glands get thicker. These facts can be used to determine if such a malignant gland exists in the image, and in case of the existence of such glands, the tissue can be classified as cancerous. A promising solution could be obtained for detection of such malignant glands by extracting a set of features for a larger dilation belt (as the nuclei dam around glands get thicker) and using this feature set for classification. To illustrate the use of such a system, we obtain the gland borders by our proposed algorithm and extract the same set of seven features for a larger dilation belt L_i with a radius of 41 pixels. Subsequently, training a decision tree classifier on these seven features, we classify the glands as *cancerous* and *other*. The rule set R_2 generated by the decision tree classifier is given below:

1. **Rule 1**: If number of black pixels inside $R_i > 1975$ and percentage of black pixels inside $R_i \leq 4.688$, then the gland is not cancerous.

2. **Rule 2**: If percentage of black pixels inside $R_i > 53.893$ and percentage of pink pixels inside $K_i > 43.966$ and percentage of pink pixels inside $L_i \leq 51.551$, then the gland is not cancerous.

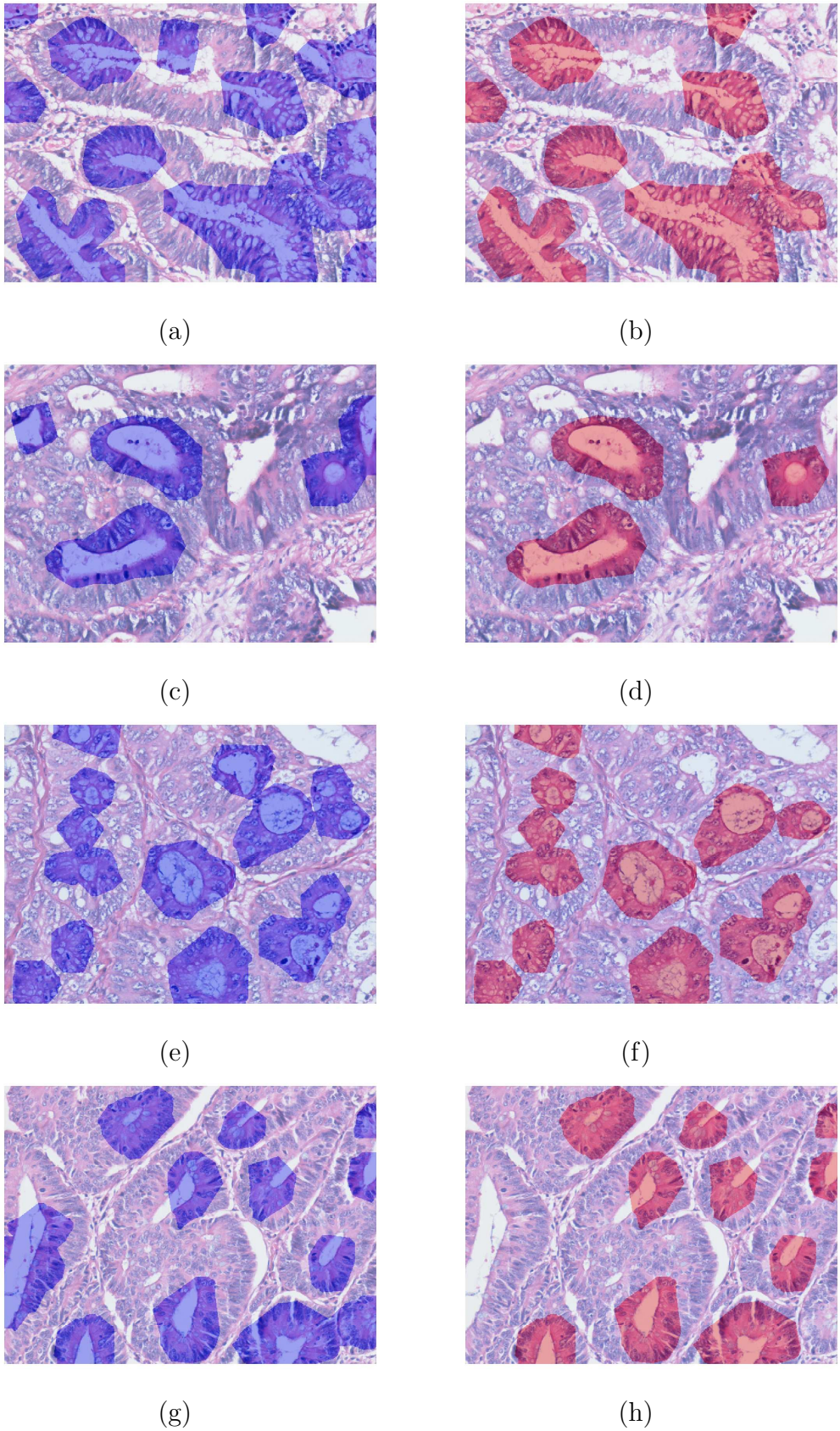
3. **Rule 3**: If number of black pixels inside $K_i \leq 996$ and percentage of white pixels inside $L_i \leq 29.615$ and percentage of pink pixels inside $L_i \leq 41.044$, then the gland is not cancerous.

4. **Rule 4**: If percentage of white pixels inside $L_i > 37.598$, then the gland is not cancerous.

5. **Rule 5**: If percentage of pink pixels inside $K_i \leq 43.966$, then the gland is not cancerous.

6. **Otherwise**: The gland is cancerous.

We apply this rule set to four cancerous and four non-cancerous tissues. The results for the images of these tissues are given in Figures 4.9 and 4.10. In the images given in Figures 4.9 (a,c,e,g), the blue components correspond to the output of our proposed method. In the images shown in Figure 4.9 (b,d,f,h), the red components are identified as cancerous by the rule set R_2. When the same method is applied to healthy images, all healthy glands are eliminated by the rule set R_2, hence no cancerous glands are identified (see Figure 4.10). For a future study, the proposed method can further be developed so that more glands with low-grade cancer are detected with more accurate borders. For colon tissues with high-grade cancer, an object-based texture analysis approach can be used to capture the global structure of primitive objects, rather than detecting the borders of these glands, since the gland structure disappears in such higher grades. Such an infrastructure could also be used for detecting other types of anomalies rather than cancer. For such anomalies, the morphological or structural features could be extracted for each identified gland and decisions could be made based on these features.



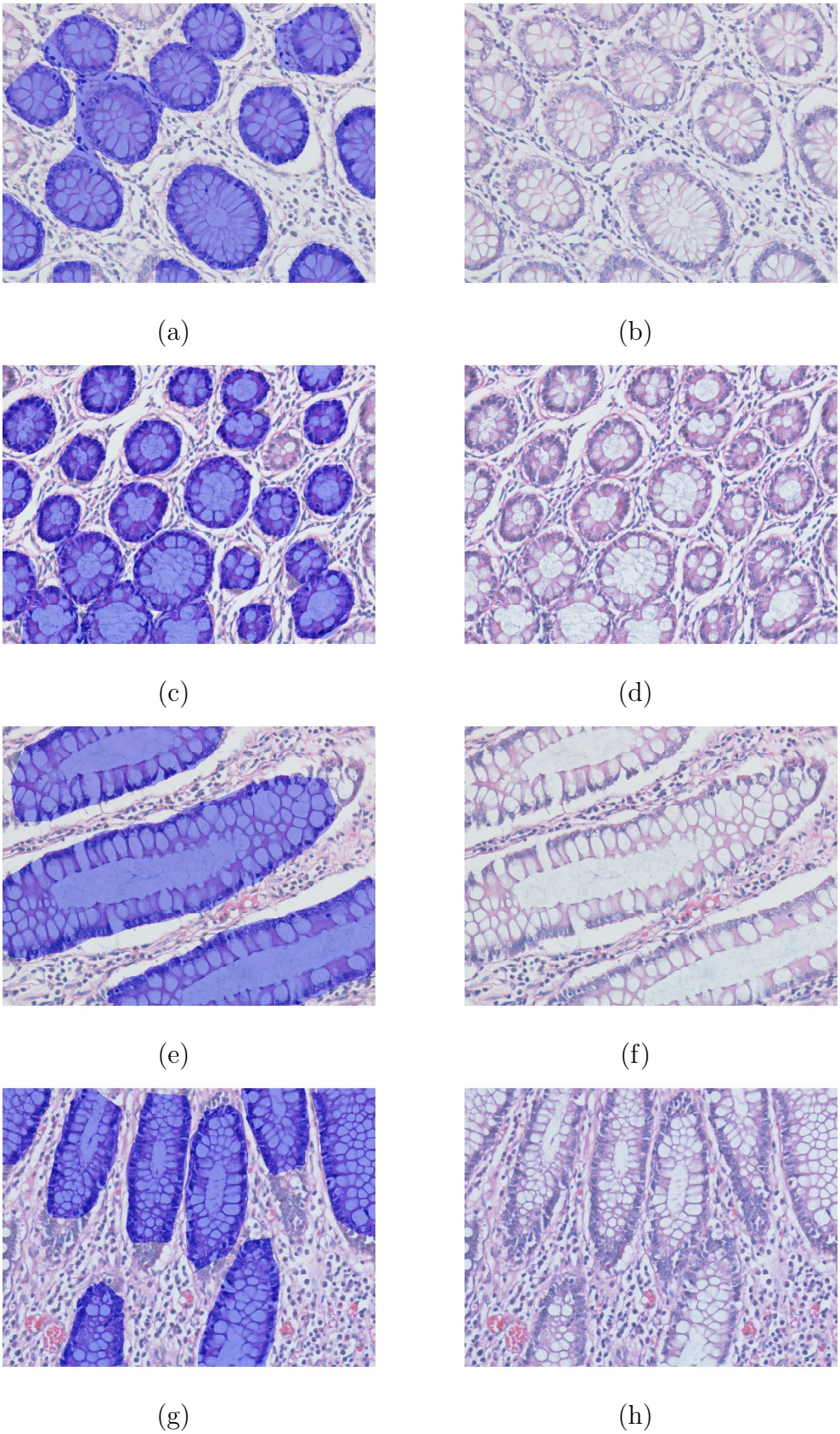
(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 4.9: (a,c,e,g): Glands segmented by our proposed system are shown in blue, and (b,d,f,h): cancerous glands detected by the second-level decision tree classifier R_2 are shown in red.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 4.10: (a,c,e,g): Glands segmented by our proposed system are shown in blue, and (b,d,f,h): cancerous glands detected by the second-level decision tree classifier R_2 are shown in red. Note that the decision tree classifier R_2 does not detect any cancerous glands for these images.

Chapter 5

Conclusion

In this thesis, we presented an object-based method for segmentation of colon glands in microscopic biopsy images. This method greatly improves the performance of its pixel-based counterparts in terms of sensitivity, specificity, and accuracy as it utilizes the spatial distributions of pixel groups corresponding to different histological structures. The object-based nature of the method also provides tolerance to noise and artifacts that arise from sectioning procedures, considering its high success rate on a dataset comprising images with a considerable amount of noise and artifacts.

Our proposed method involves decomposing the image into a set of circular objects. Such a representation of an image defines a transformation, which we refer to as *the circle-fit transform*. This transform provides a circular representation of histological structures in the tissue. Such a representation of luminal and nuclear objects enables extraction of discriminative features for location of the glands. The circles corresponding to nuclear objects are used to locate the dam of the epithelial cells surrounding a gland, and hence, they lead to detection of exact gland borders.

This object-based method provides an infrastructure for further biopsy analysis. It could be extended for detection of cancerous glands for tissues with low-grade cancer. The morphological or structural features of the segmented

glands could be used for automated detection, identification, and grading of other anomalies as well. One of the future research directions is to implement such automated systems. Another future work is to apply texture analysis techniques that use the proposed object-based representation for detection of anomalies (such as high-grade cancer) in which the gland structure disappears.

Appendix A

Performance Results of Images

The performance results obtained on each individual image are calculated using various performance evaluation schemes, which are explained in Subsection 4.2.2. These results are reported in Tables A.1, A.2, A.3, A.4, A.5, A.6, A.7, and A.8. The average and standard deviation of performances for training and test sets with respect to these schemes are given in Tables 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8, respectively.

	Our Method			Region Growing			Directional Filtering		
Image	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**
1	0.952	0.784	0.892	0.369	0.831	0.534	0.162	0.826	0.399
2	0.871	0.892	0.880	0.465	0.967	0.673	0.232	0.869	0.496
3	0.849	0.815	0.831	0.000	1.000	0.526	0.956	0.001	0.453
4	0.457	0.962	0.757	0.000	1.000	0.594	0.877	0.850	0.861
5	0.935	0.841	0.904	0.100	0.994	0.396	0.111	0.657	0.292
6	0.899	0.949	0.929	0.727	0.899	0.830	0.890	0.778	0.823
7	0.850	0.919	0.875	0.238	0.964	0.503	0.183	0.902	0.446
8	0.894	0.867	0.886	0.307	0.926	0.506	0.132	0.953	0.396
9	0.641	0.960	0.837	0.651	0.921	0.817	0.902	0.853	0.872
10	0.891	0.884	0.888	0.341	0.887	0.574	0.239	0.942	0.540
11	0.829	0.908	0.852	0.379	0.937	0.540	0.162	0.851	0.361
12	0.873	0.937	0.904	0.000	1.000	0.488	0.968	0.000	0.495
13	0.878	0.784	0.846	0.167	0.975	0.436	0.242	0.517	0.334
14	0.715	0.811	0.754	0.000	1.000	0.401	0.223	0.901	0.494
15	0.757	0.994	0.905	0.492	0.911	0.753	0.816	0.928	0.886
16	0.952	0.751	0.903	0.000	1.000	0.242	0.400	0.047	0.315
17	0.938	0.890	0.920	0.250	0.984	0.526	0.259	0.869	0.488
18	0.629	0.952	0.832	0.000	1.000	0.627	0.659	0.920	0.823
19	0.714	0.956	0.826	0.000	1.000	0.462	0.727	0.768	0.746
20	0.851	0.866	0.856	0.428	0.934	0.605	0.661	0.408	0.573
21	0.891	0.838	0.868	0.125	0.997	0.506	1.000	0.000	0.562
22	0.950	0.839	0.916	0.067	0.974	0.340	0.305	0.769	0.444
23	0.978	0.806	0.928	0.445	0.847	0.560	0.870	0.062	0.638
24	0.932	0.907	0.919	0.307	0.923	0.625	0.421	0.843	0.638
Average	0.839	0.880	0.871	0.244	0.953	0.544	0.517	0.646	0.557
Std.Dev	0.126	0.067	0.048	0.221	0.051	0.138	0.328	0.351	0.188

[Pixel-based performances of the images in the training set]

Table A.1: Pixel-based performances of the images in the training set. For each individual image; sensitivity, specificity, and accuracy values are reported.

	Our Method			Region Growing			Directional Filtering		
Image	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**
1	0.860	0.511	0.724	0.000	0.998	0.387	1.000	0.000	0.612
2	0.764	0.779	0.770	0.295	0.739	0.459	0.717	0.102	0.490
3	0.939	0.811	0.899	0.375	0.886	0.534	0.077	0.941	0.346
4	0.913	0.880	0.902	0.233	0.983	0.487	0.124	0.801	0.353
5	0.856	0.694	0.800	0.126	0.963	0.416	0.116	0.828	0.362
6	0.816	0.891	0.851	0.048	0.968	0.476	0.216	0.900	0.535
7	0.947	0.802	0.898	0.439	0.870	0.585	0.501	0.380	0.460
8	0.834	0.947	0.900	0.155	0.999	0.648	0.844	0.678	0.747
9	0.863	0.773	0.808	0.167	0.996	0.680	0.741	0.760	0.753
10	0.903	0.858	0.878	0.382	0.992	0.711	0.821	0.533	0.666
11	0.887	0.939	0.917	0.324	0.993	0.715	0.779	0.664	0.712
12	0.828	0.651	0.777	0.390	0.795	0.506	0.841	0.002	0.601
13	0.969	0.740	0.897	0.403	0.860	0.547	0.944	0.000	0.647
14	0.885	0.724	0.817	0.000	1.000	0.422	1.000	0.000	0.578
15	0.743	0.889	0.811	0.000	1.000	0.464	0.999	0.000	0.536
16	0.852	0.683	0.793	0.167	0.961	0.443	0.067	0.945	0.372
17	0.934	0.629	0.839	0.420	0.855	0.556	0.879	0.003	0.607
18	0.879	0.860	0.872	0.431	0.960	0.622	0.299	0.555	0.391
19	0.938	0.724	0.868	0.275	0.964	0.502	0.284	0.741	0.435
20	0.908	0.846	0.879	0.154	0.984	0.547	0.410	0.861	0.623
21	0.914	0.777	0.854	0.160	0.981	0.521	0.312	0.815	0.534
22	0.813	0.749	0.780	0.130	0.919	0.534	0.328	0.877	0.609
23	0.865	0.820	0.841	0.000	1.000	0.536	0.537	0.526	0.531
24	0.799	0.891	0.845	0.168	0.953	0.559	0.481	0.719	0.600
25	0.951	0.921	0.939	0.002	0.991	0.385	1.000	0.000	0.613
26	0.907	0.807	0.870	0.000	1.000	0.369	0.855	0.001	0.540
27	0.848	0.961	0.901	0.051	0.990	0.489	0.579	0.767	0.667
28	0.782	0.861	0.822	0.043	0.999	0.525	0.269	0.929	0.602
29	0.792	0.731	0.767	0.011	0.987	0.420	0.011	0.996	0.423
30	0.829	0.936	0.886	0.573	0.995	0.798	0.548	0.898	0.734
31	0.939	0.746	0.876	0.204	0.965	0.451	1.000	0.000	0.676
32	0.884	0.913	0.896	0.301	0.955	0.560	0.292	0.831	0.505
33	0.859	0.804	0.835	0.226	0.967	0.556	0.279	0.915	0.562
34	0.941	0.803	0.901	0.351	0.933	0.519	1.000	0.000	0.712
35	0.926	0.656	0.839	0.000	1.000	0.321	0.998	0.001	0.678
36	0.980	0.759	0.909	0.324	0.917	0.515	0.214	0.765	0.391
37	0.884	0.676	0.826	0.448	0.876	0.567	0.259	0.455	0.314
38	0.783	0.826	0.800	0.000	1.000	0.391	0.989	0.001	0.603
39	0.887	0.836	0.867	0.101	0.987	0.450	0.259	0.857	0.494
40	0.703	0.916	0.758	0.030	0.980	0.274	0.984	0.000	0.731
41	0.739	0.926	0.793	0.000	1.000	0.290	0.204	0.821	0.382
42	0.906	0.796	0.863	0.402	0.926	0.605	0.568	0.754	0.640
43	0.879	0.867	0.874	0.053	0.994	0.423	0.832	0.013	0.510
44	0.792	0.908	0.850	0.386	0.943	0.662	0.992	0.051	0.525
45	0.870	0.822	0.846	0.511	0.965	0.736	0.217	0.841	0.526
46	0.944	0.865	0.907	0.349	0.954	0.636	0.155	0.942	0.528
47	0.954	0.923	0.938	0.740	0.994	0.873	0.281	0.967	0.642
48	0.515	0.532	0.523	0.193	0.829	0.528	0.395	0.611	0.509
Average	0.863	0.805	0.844	0.220	0.953	0.525	0.552	0.522	0.554
Std.Dev	0.083	0.105	0.069	0.184	0.060	0.124	0.335	0.387	0.118

Table A.2: Pixel-based performances of the images in the test set. For each individual image; sensitivity, specificity, and accuracy values are reported.

	Our Method			Region Growing			Directional Filtering		
Image	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**
1	0.951	0.784	0.891	0.230	0.797	0.441	0.000	0.640	0.295
2	0.870	0.891	0.879	0.216	0.967	0.527	0.038	0.676	0.378
3	0.846	0.815	0.830	0.000	1.000	0.526	0.097	0.001	0.047
4	0.453	0.957	0.754	0.000	1.000	0.594	0.539	0.845	0.721
5	0.935	0.838	0.903	0.100	0.992	0.395	0.000	0.536	0.217
6	0.899	0.949	0.929	0.693	0.838	0.786	0.812	0.720	0.752
7	0.849	0.916	0.874	0.200	0.891	0.473	0.000	0.684	0.330
8	0.893	0.867	0.885	0.223	0.853	0.443	0.012	0.763	0.314
9	0.619	0.927	0.815	0.601	0.875	0.779	0.364	0.809	0.652
10	0.890	0.884	0.887	0.193	0.712	0.469	0.033	0.733	0.417
11	0.829	0.908	0.852	0.160	0.570	0.354	0.006	0.613	0.249
12	0.871	0.923	0.897	0.000	1.000	0.488	0.058	0.000	0.030
13	0.878	0.784	0.846	0.167	0.975	0.436	0.026	0.507	0.189
14	0.681	0.773	0.720	0.000	1.000	0.401	0.128	0.774	0.429
15	0.757	0.994	0.905	0.387	0.827	0.690	0.795	0.875	0.848
16	0.951	0.743	0.900	0.000	1.000	0.242	0.055	0.038	0.050
17	0.934	0.890	0.917	0.198	0.889	0.485	0.011	0.613	0.331
18	0.629	0.953	0.832	0.000	1.000	0.627	0.472	0.808	0.712
19	0.661	0.808	0.741	0.000	1.000	0.462	0.197	0.643	0.443
20	0.843	0.794	0.824	0.218	0.819	0.457	0.025	0.365	0.158
21	0.891	0.836	0.867	0.125	0.997	0.506	0.029	0.000	0.017
22	0.950	0.839	0.916	0.024	0.884	0.309	0.131	0.526	0.305
23	0.971	0.757	0.906	0.210	0.732	0.383	0.019	0.049	0.030
24	0.932	0.907	0.919	0.188	0.834	0.557	0.184	0.662	0.498
Average	0.833	0.864	0.862	0.172	0.894	0.493	0.168	0.537	0.351
Std.Dev.	0.132	0.071	0.058	0.181	0.115	0.132	0.246	0.295	0.247

Table A.3: Modified pixel-based performances of the images in the training set. For each individual image; sensitivity, specificity, and accuracy values are reported.

	Our Method			Region Growing			Directional Filtering		
Image	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**
1	0.527	0.507	0.519	0.000	0.998	0.387	0.065	0.000	0.040
2	0.758	0.773	0.763	0.173	0.639	0.372	0.017	0.102	0.049
3	0.706	0.805	0.737	0.310	0.736	0.470	0.010	0.818	0.299
4	0.913	0.881	0.902	0.226	0.968	0.482	0.019	0.663	0.283
5	0.855	0.693	0.799	0.118	0.947	0.409	0.012	0.691	0.293
6	0.812	0.867	0.838	0.010	0.930	0.457	0.017	0.730	0.426
7	0.947	0.798	0.896	0.299	0.745	0.476	0.168	0.268	0.216
8	0.831	0.932	0.891	0.155	0.999	0.648	0.315	0.614	0.508
9	0.859	0.760	0.796	0.167	0.996	0.680	0.183	0.670	0.525
10	0.902	0.858	0.878	0.367	0.972	0.700	0.268	0.503	0.402
11	0.886	0.939	0.917	0.307	0.976	0.705	0.447	0.592	0.542
12	0.824	0.648	0.773	0.166	0.718	0.341	0.008	0.002	0.006
13	0.966	0.739	0.895	0.238	0.774	0.425	0.011	0.000	0.007
14	0.885	0.719	0.814	0.000	1.000	0.422	0.005	0.000	0.003
15	0.654	0.686	0.674	0.000	1.000	0.464	0.265	0.000	0.142
16	0.617	0.683	0.640	0.162	0.951	0.439	0.026	0.881	0.345
17	0.673	0.595	0.648	0.344	0.726	0.484	0.250	0.003	0.173
18	0.877	0.859	0.871	0.395	0.891	0.588	0.019	0.369	0.209
19	0.935	0.722	0.865	0.257	0.951	0.489	0.139	0.586	0.325
20	0.907	0.844	0.877	0.145	0.973	0.541	0.150	0.677	0.467
21	0.908	0.776	0.850	0.100	0.980	0.488	0.078	0.623	0.392
22	0.810	0.739	0.773	0.084	0.919	0.512	0.051	0.756	0.469
23	0.862	0.806	0.832	0.000	1.000	0.536	0.251	0.455	0.377
24	0.798	0.885	0.841	0.133	0.916	0.539	0.061	0.496	0.375
25	0.951	0.921	0.939	0.000	0.992	0.385	0.220	0.000	0.135
26	0.896	0.681	0.801	0.000	1.000	0.369	0.230	0.001	0.145
27	0.848	0.961	0.901	0.004	0.941	0.464	0.081	0.670	0.395
28	0.778	0.847	0.813	0.043	0.999	0.525	0.081	0.778	0.501
29	0.785	0.708	0.752	0.003	0.979	0.416	0.005	0.988	0.419
30	0.829	0.937	0.886	0.564	0.977	0.788	0.116	0.763	0.521
31	0.937	0.745	0.875	0.165	0.895	0.420	0.074	0.000	0.050
32	0.883	0.904	0.892	0.260	0.892	0.528	0.009	0.579	0.333
33	0.859	0.802	0.833	0.195	0.939	0.536	0.046	0.773	0.429
34	0.941	0.803	0.901	0.269	0.825	0.450	0.009	0.000	0.006
35	0.858	0.655	0.792	0.000	1.000	0.321	0.020	0.001	0.014
36	0.980	0.757	0.908	0.296	0.914	0.495	0.008	0.532	0.250
37	0.880	0.667	0.820	0.176	0.855	0.369	0.007	0.304	0.130
38	0.783	0.826	0.800	0.000	1.000	0.391	0.164	0.001	0.100
39	0.887	0.836	0.867	0.101	0.987	0.450	0.139	0.797	0.417
40	0.703	0.916	0.758	0.003	0.913	0.254	0.039	0.000	0.029
41	0.728	0.854	0.768	0.000	1.000	0.290	0.023	0.577	0.251
42	0.904	0.771	0.851	0.359	0.836	0.564	0.029	0.538	0.305
43	0.877	0.867	0.873	0.008	0.930	0.396	0.037	0.013	0.027
44	0.792	0.908	0.850	0.367	0.916	0.648	0.119	0.051	0.086
45	0.869	0.822	0.846	0.480	0.946	0.715	0.029	0.709	0.429
46	0.943	0.865	0.906	0.249	0.831	0.566	0.012	0.812	0.453
47	0.954	0.911	0.931	0.730	0.964	0.857	0.137	0.841	0.563
48	0.954	0.911	0.931	0.163	0.807	0.481	0.052	0.564	0.303
Average	0.844	0.800	0.829	0.179	0.918	0.494	0.094	0.433	0.274
Std.Dev.	0.097	0.101	0.084	0.166	0.091	0.126	0.102	0.330	0.177

Table A.4: Modified pixel-based performances of the images in the test set. For each individual image; sensitivity, specificity, and accuracy values are reported.

	Our Method				Region Growing				Directional Filtering			
Image	True Gland Perc.	Illeg. Border Detect. Perc.	Misdet. Gland Perc.	False Alarm Perc.	True Gland Perc.	Illeg. Border Detect. Perc.	Misdet. Gland Perc.	False Alarm Perc.	True Gland Perc.	Illeg. Border Perc. Perc.	Misdet. Gland Perc.	False Alarm Perc.
1	0.938	0.000	0.063	0.000	0.000	0.563	0.438	0.727	0.000	0.875	0.125	1.000
2	0.667	0.000	0.333	0.000	0.056	0.167	0.778	0.500	0.056	0.611	0.333	0.824
3	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
4	0.459	0.000	0.541	0.056	0.000	0.000	1.000	0.000	0.405	0.541	0.054	0.087
5	0.917	0.000	0.083	0.000	0.083	0.000	0.917	0.500	0.000	0.750	0.250	1.000
6	0.750	0.000	0.250	0.000	1.000	0.000	0.000	0.912	0.583	0.333	0.083	0.273
7	0.857	0.000	0.143	0.000	0.071	0.071	0.857	0.667	0.000	0.786	0.214	1.000
8	0.842	0.000	0.158	0.000	0.158	0.526	0.316	0.750	0.000	0.684	0.316	0.952
9	0.889	0.000	0.111	0.111	0.778	0.222	0.000	0.951	0.333	0.667	0.000	0.600
10	0.800	0.000	0.200	0.000	0.200	0.400	0.400	0.983	0.500	0.300	0.200	0.929
11	0.714	0.000	0.286	0.000	0.143	0.286	0.571	0.500	0.143	0.714	0.143	0.909
12	1.000	0.000	0.000	0.083	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
13	0.875	0.000	0.125	0.000	0.125	0.000	0.875	0.000	0.000	1.000	0.000	0.800
14	0.364	0.273	0.364	0.222	0.000	0.000	1.000	0.000	0.000	0.636	0.364	0.778
15	1.000	0.000	0.000	0.000	0.667	0.333	0.000	0.974	0.667	0.333	0.000	0.667
16	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
17	0.750	0.000	0.250	0.000	0.063	0.250	0.688	0.875	0.000	0.813	0.188	0.905
18	0.643	0.000	0.357	0.000	0.000	0.000	1.000	0.000	0.464	0.500	0.036	0.621
19	0.833	0.000	0.167	0.444	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.600
20	0.600	0.000	0.400	0.077	0.050	0.300	0.650	0.864	0.150	0.850	0.000	0.700
21	0.762	0.000	0.238	0.000	0.095	0.000	0.905	0.000	0.000	1.000	0.000	0.000
22	1.000	0.000	0.000	0.000	0.125	0.250	0.625	0.929	0.125	0.750	0.125	0.857
23	0.900	0.000	0.100	0.100	0.000	0.300	0.700	0.800	0.000	1.000	0.000	0.000
24	0.882	0.000	0.118	0.000	0.176	0.588	0.235	0.951	0.000	0.941	0.059	0.733
Average	0.810	0.011	0.179	0.046	0.158	0.177	0.665	0.495	0.143	0.754	0.104	0.593
Std.Dev.	0.171	0.056	0.147	0.101	0.266	0.200	0.347	0.416	0.219	0.229	0.121	0.378

Table A.5: Gland-based performances of the images in the training set. For each individual image; the average true gland detection percentage, the percentage of glands detected with illegitimate borders, the percentage of misdetected glands, and the percentage of false alarms are reported.

	Our Method				Region Growing				Directional Filtering			
Image	True Gland Perc.	Illeg. Border Detect. Perc.	Misdet. Gland Perc.	False Alarm Perc.	True Gland Perc.	Illeg. Border Detect. Perc.	Misdet. Gland Perc.	False Alarm Perc.	True Gland Perc.	Illeg. Border Perc. Perc.	Misdet. Gland Perc.	False Alarm Perc.
1	0.200	0.300	0.500	0.167	0.000	0.000	1.000	1.000	0.000	1.000	0.000	0.000
2	0.583	0.000	0.417	0.125	0.000	0.583	0.417	0.800	0.000	0.750	0.250	0.000
3	0.429	0.214	0.357	0.182	0.143	0.429	0.429	0.750	0.000	0.786	0.214	0.923
4	0.750	0.000	0.250	0.000	0.000	0.250	0.750	0.000	0.125	0.625	0.250	0.933
5	0.652	0.000	0.348	0.000	0.130	0.000	0.870	0.500	0.043	0.739	0.217	0.926
6	0.714	0.000	0.286	0.167	0.000	0.571	0.429	0.987	0.000	0.857	0.143	0.923
7	0.900	0.000	0.100	0.000	0.000	0.200	0.800	0.000	0.000	1.000	0.000	0.750
8	0.769	0.000	0.231	0.167	0.077	0.000	0.923	0.000	0.385	0.538	0.077	0.583
9	0.846	0.000	0.154	0.389	0.077	0.000	0.923	0.000	0.385	0.692	-0.077	0.615
10	0.824	0.000	0.176	0.000	0.294	0.000	0.706	0.333	0.235	0.824	-0.059	0.556
11	0.643	0.000	0.357	0.000	0.286	0.000	0.714	0.250	0.214	0.714	0.071	0.333
12	0.591	0.000	0.409	0.000	0.045	0.409	0.545	0.857	0.000	0.773	0.227	0.000
13	0.909	0.000	0.091	0.000	0.000	0.636	0.364	0.833	0.000	0.909	0.091	0.000
14	0.818	0.000	0.182	0.000	0.000	0.000	1.000	0.000	0.000	0.818	0.182	0.000
15	0.400	0.400	0.200	0.333	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
16	0.286	0.143	0.571	0.143	0.071	0.000	0.929	0.000	0.000	0.500	0.500	0.571
17	0.000	0.500	0.500	0.286	0.100	0.700	0.200	0.923	0.000	0.800	0.200	0.000
18	0.833	0.000	0.167	0.000	0.167	0.333	0.500	0.500	0.083	0.833	0.083	0.833
19	0.813	0.000	0.188	0.000	0.063	0.250	0.688	0.000	0.000	0.875	0.125	0.875
20	0.781	0.000	0.219	0.000	0.031	0.063	0.906	0.500	0.125	0.750	0.125	0.733
21	0.857	0.000	0.143	0.000	0.107	0.000	0.893	0.000	0.107	0.786	0.107	0.885
22	0.786	0.000	0.214	0.214	0.000	0.000	1.000	0.000	0.071	0.857	0.071	0.889
23	0.813	0.000	0.188	0.235	0.000	0.000	1.000	0.000	0.125	0.750	0.125	0.500
24	0.714	0.000	0.286	0.167	0.286	0.000	0.714	0.667	0.000	1.000	0.000	0.750
25	1.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	0.000	1.000	0.000	0.000
26	0.875	0.000	0.125	0.222	0.125	0.000	0.875	0.000	0.000	1.000	0.000	0.000
27	0.889	0.000	0.111	0.000	0.000	1.000	0.000	0.983	0.111	0.889	0.000	0.667
28	0.800	0.000	0.200	0.167	0.000	0.000	1.000	0.000	0.000	0.800	0.200	0.905
29	0.667	0.000	0.333	0.176	0.000	0.333	0.667	0.955	0.000	0.143	0.857	0.667
30	0.625	0.000	0.375	0.000	0.375	0.000	0.625	0.167	0.125	0.625	0.250	0.722
31	0.833	0.000	0.167	0.000	0.167	0.111	0.722	0.600	0.000	0.778	0.222	0.000
32	0.750	0.000	0.250	0.100	0.000	0.333	0.667	0.750	0.083	0.917	0.000	0.947
33	0.727	0.000	0.273	0.333	0.000	0.182	0.818	0.900	0.182	0.818	0.000	0.813
34	0.818	0.000	0.182	0.000	0.091	0.545	0.364	0.962	0.000	0.818	0.182	0.000
35	0.615	0.154	0.231	0.091	0.000	0.000	1.000	0.000	0.000	0.923	0.077	0.000
36	0.857	0.000	0.143	0.000	0.000	0.286	0.714	0.600	0.000	0.929	0.071	0.857
37	0.722	0.000	0.278	0.000	0.056	0.444	0.500	0.333	0.000	1.056	-0.056	0.875
38	1.000	0.000	0.000	0.143	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
39	0.600	0.000	0.400	0.000	0.050	0.000	0.950	0.000	0.100	0.700	0.200	0.800
40	0.818	0.000	0.182	0.000	0.000	0.636	0.364	0.988	0.000	0.909	0.091	0.000
41	0.846	0.000	0.154	0.083	0.000	0.000	1.000	0.000	0.000	0.923	0.077	0.905
42	0.700	0.000	0.300	0.125	0.300	0.000	0.700	0.889	0.000	1.100	-0.100	0.818
43	0.800	0.000	0.200	0.000	0.000	0.133	0.867	0.909	0.200	0.000	0.800	0.000
44	0.846	0.000	0.154	0.000	0.308	0.000	0.692	0.500	0.000	0.923	0.077	0.000
45	0.722	0.000	0.278	0.000	0.333	0.111	0.556	0.333	0.056	0.611	0.333	0.852
46	0.813	0.000	0.188	0.000	0.375	0.188	0.438	0.906	0.063	0.625	0.313	0.909
47	0.933	0.000	0.067	0.067	0.733	0.000	0.267	0.182	0.067	0.533	0.400	0.950
48	0.933	0.000	0.067	0.067	0.000	0.267	0.733	0.985	0.067	0.333	0.600	0.778
Average	0.731	0.036	0.233	0.086	0.100	0.187	0.713	0.455	0.061	0.782	0.157	0.522
Std.Dev.	0.196	0.106	0.125	0.109	0.150	0.248	0.253	0.405	0.095	0.218	0.201	0.393

Table A.6: Gland-based performances of the images in the test set. For each individual image; the average true gland detection percentage, the percentage of glands detected with illegitimate borders, the percentage of misdetected glands, and the percentage of false alarms are reported.

	Our Method			Region Growing			Directional Filtering		
Image	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**
1	0.959	0.795	0.898	0.234	0.803	0.453	0.000	0.678	0.315
2	0.943	0.902	0.924	0.234	0.970	0.573	0.000	0.723	0.412
3	0.846	0.815	0.830	0.000	1.000	0.526	0.097	0.001	0.047
4	0.462	0.963	0.779	0.000	1.000	0.631	0.499	0.897	0.750
5	0.960	0.846	0.920	0.102	0.992	0.412	0.000	0.579	0.241
6	0.957	0.951	0.953	0.714	0.874	0.819	0.807	0.764	0.778
7	0.867	0.919	0.887	0.205	0.895	0.486	0.000	0.724	0.352
8	0.919	0.875	0.904	0.221	0.871	0.460	0.000	0.796	0.328
9	0.619	0.927	0.815	0.601	0.887	0.786	0.364	0.809	0.652
10	0.911	0.887	0.901	0.196	0.733	0.486	0.034	0.748	0.431
11	0.836	0.910	0.858	0.162	0.576	0.361	0.000	0.622	0.253
12	0.869	0.929	0.900	0.000	1.000	0.504	0.000	1.000	0.504
13	0.885	0.817	0.861	0.174	0.977	0.464	0.027	0.540	0.214
14	0.690	0.777	0.727	0.000	1.000	0.408	0.130	0.780	0.437
15	0.757	0.994	0.905	0.387	0.832	0.693	0.778	0.886	0.849
16	0.951	0.743	0.900	0.000	1.000	0.242	0.055	0.038	0.050
17	0.890	0.921	0.902	0.206	0.897	0.508	0.000	0.642	0.349
18	0.644	0.956	0.847	0.000	1.000	0.650	0.475	0.829	0.731
19	0.661	0.808	0.741	0.000	1.000	0.462	0.197	0.645	0.444
20	0.907	0.815	0.868	0.234	0.837	0.501	0.000	0.913	0.390
21	0.911	0.840	0.879	0.128	0.997	0.519	0.000	1.000	0.450
22	0.950	0.839	0.916	0.024	0.886	0.309	0.131	0.527	0.305
23	0.983	0.762	0.914	0.212	0.735	0.390	0.000	1.000	0.295
24	0.958	0.909	0.932	0.192	0.843	0.569	0.156	0.708	0.511
Average	0.847	0.871	0.873	0.176	0.900	0.509	0.156	0.702	0.420
Std.Dev	0.137	0.070	0.058	0.184	0.111	0.136	0.246	0.251	0.211

Table A.7: Modified pixel-based performances of the images in the training set; when the glands with centroids closer to the image boundaries than 40 pixels are excluded. For each individual image; sensitivity, specificity, and accuracy values are reported.

	Our Method			Region Growing			Directional Filtering		
Image	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**
1	0.527	0.507	0.519	0.000	1.000	0.388	0.065	0.000	0.040
2	0.797	0.790	0.794	0.182	0.661	0.400	0.000	1.000	0.402
3	0.711	0.808	0.741	0.313	0.741	0.476	0.000	0.851	0.305
4	0.913	0.881	0.902	0.226	0.968	0.482	0.019	0.694	0.287
5	0.893	0.716	0.827	0.123	0.951	0.438	0.000	0.731	0.318
6	0.812	0.867	0.838	0.010	0.941	0.461	0.017	0.743	0.431
7	0.961	0.803	0.906	0.271	0.917	0.500	0.000	1.000	0.349
8	0.893	0.935	0.919	0.166	0.999	0.677	0.275	0.802	0.628
9	0.920	0.773	0.823	0.183	0.996	0.714	0.124	0.800	0.621
10	0.957	0.870	0.907	0.404	0.974	0.742	0.276	0.602	0.476
11	0.946	0.941	0.943	0.329	0.978	0.732	0.375	0.822	0.667
12	0.854	0.676	0.798	0.172	0.737	0.365	0.000	1.000	0.311
13	0.976	0.745	0.902	0.241	0.779	0.433	0.000	1.000	0.322
14	0.901	0.725	0.824	0.000	1.000	0.435	0.000	1.000	0.435
15	0.654	0.686	0.674	0.000	1.000	0.464	0.265	0.000	0.142
16	0.644	0.841	0.719	0.171	0.955	0.472	0.000	0.952	0.371
17	0.684	0.607	0.658	0.349	0.735	0.494	0.000	1.000	0.321
18	0.892	0.863	0.881	0.401	0.894	0.598	0.000	0.391	0.215
19	0.957	0.735	0.880	0.263	0.953	0.504	0.142	0.603	0.340
20	0.952	0.858	0.904	0.155	0.975	0.575	0.124	0.728	0.503
21	0.934	0.784	0.865	0.103	0.981	0.503	0.070	0.641	0.405
22	0.824	0.743	0.781	0.086	0.920	0.520	0.031	0.826	0.503
23	0.888	0.826	0.853	0.000	1.000	0.557	0.216	0.493	0.393
24	0.832	0.889	0.862	0.140	0.919	0.561	0.000	0.526	0.391
25	0.951	0.921	0.939	0.000	0.997	0.386	0.220	0.000	0.135
26	0.911	0.687	0.811	0.000	1.000	0.379	0.000	1.000	0.379
27	0.858	0.962	0.907	0.000	0.952	0.471	0.069	0.681	0.399
28	0.779	0.861	0.823	0.045	0.999	0.551	0.086	0.792	0.527
29	0.816	0.769	0.795	0.000	0.987	0.440	0.000	1.000	0.443
30	0.855	0.939	0.901	0.583	0.978	0.803	0.080	0.837	0.555
31	0.962	0.778	0.896	0.174	0.905	0.455	0.000	1.000	0.360
32	0.892	0.906	0.898	0.262	0.894	0.533	0.000	0.589	0.337
33	0.898	0.833	0.867	0.204	0.942	0.561	0.000	0.816	0.448
34	0.956	0.810	0.912	0.271	0.844	0.462	0.000	1.000	0.299
35	0.873	0.668	0.805	0.000	1.000	0.333	0.000	1.000	0.333
36	0.988	0.779	0.916	0.307	0.919	0.519	0.000	0.595	0.274
37	0.898	0.719	0.843	0.181	0.859	0.389	0.000	0.438	0.189
38	0.783	0.826	0.800	0.000	1.000	0.391	0.164	0.001	0.100
39	0.930	0.847	0.895	0.106	0.988	0.477	0.000	0.940	0.416
40	0.702	0.926	0.763	0.003	0.936	0.272	0.000	1.000	0.272
41	0.736	0.857	0.775	0.000	1.000	0.297	0.016	0.645	0.266
42	0.928	0.779	0.866	0.368	0.842	0.579	0.000	0.828	0.393
43	0.898	0.872	0.887	0.009	0.939	0.410	0.000	1.000	0.408
44	0.801	0.910	0.856	0.372	0.917	0.654	0.000	1.000	0.503
45	0.895	0.827	0.861	0.494	0.948	0.730	0.000	0.757	0.452
46	0.960	0.887	0.925	0.255	0.836	0.578	0.000	0.834	0.463
47	0.961	0.917	0.936	0.755	0.965	0.872	0.140	0.853	0.578
48	0.961	0.917	0.936	0.162	0.821	0.488	0.052	0.573	0.307
Average	0.865	0.814	0.844	0.184	0.926	0.511	0.059	0.737	0.375
Std.Dev.	0.102	0.095	0.083	0.171	0.084	0.129	0.095	0.283	0.133

Table A.8: Modified pixel-based performances of the images in the test set; when the glands with centroids closer to the image boundaries than 40 pixels are excluded. For each individual image; sensitivity, specificity, and accuracy values are reported.

Bibliography

[1] G.D. Thomas, M.F. Dixon, N.C. Smeeton, and N.S. Williams. Observer variation in the histological grading of rectal carcinoma. *Journal of Clinical Pathology*, 36(4):385–391, April 1983.

[2] A. Andrion, C. Magnani, P.G. Betta, A. Donna, F. Mollo, M. Scelsi, P. Bernardi, M. Botta, and B. Terracini. Malignant mesothelioma of the pleura: interobserver variability. *Journal of Clinical Pathology*, 48(9):856–860, September 1995.

[3] W.H. Wolberg, W.N. Street, D.M. Heisey, and O.L. Mangasarian. Computer-derived nuclear features distinguish malignant from benign breast cytology. *Journal of Human Pathology*, 26(7):792–796, July 1995.

[4] J.P. Thiran and B. Macq. Morphological feature extraction for the classification of digital images of cancerous tissues. *IEEE Transactions on Biomedical Engineering*, 43(10):1011–1020, October 1996.

[5] P.W. Hamilton, P.H. Bartels, D. Thompson, N.H. Anderson, and R. Montironi. Automated location of dysplastic fields in colorectal histology using image texture analysis. *Journal of Pathology*, 182(1):68–75, May 1997.

[6] H.K. Choi, T. Jarkrans, E. Bengtsson, J. Vasko, K. Wester, P.U. Malmstrom, and C. Busch. Image analysis based grading of bladder carcinoma. comparison of object, texture and graph based methods and their reproducibility. *Journal of Analytical Cellular Pathology*, 15(1):1–18, January 1997.

[7] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, and A. Murray. Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa. *IEEE Transactions on Information Technology in Biomedicine*, 2(3):197–203, September 1998.

[8] P. Spyridonos, P. Ravazoula, D. Cavouras, K. Berberidis, and G. Nikiforidis. Computer-based grading of haematoxylin-eosin stained tissue sections of urinary bladder carcinomas. *Journal of Medical Informatics*, 26(3):179–190, July 2001.

[9] M. Wiltgen, A. Gerger, and J. Smolle. Tissue counter analysis of benign common nevi and malignant melanoma. *International Journal of Medical Informatics*, 69(1):17–28, January 2003.

[10] B. Nielsen, F. Albregtsen, and H.E. Danielsen. The use of fractal features from the periphery of cell nuclei as a classification tool. *Journal of Analytical Cellular Pathology*, 19(1):21–37, January 1999.

[11] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, and A. Murray. Fractal analysis in the detection of colonic cancer images. *IEEE Transactions on Information Technology in Biomedicine*, 6(1):54–58, March 2002.

[12] B. Weyn, G. Van de Wouwer, S. Kumar-Singh, A. Van Daele, P. Scheunders, E. Van Marck, and W. Jacob. Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis. *Journal of Cytometry*, 35(1):23–29, January 1999.

[13] S.J. Keenan, J. Diamond, W.G. McCluggage, H. Bharucha, D. Thompson, B.H. Bartels, and P.W. Hamilton. An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (cin). *Journal of Pathology*, 192(3):351–362, November 2000.

[14] C. Demir, S.H. Gultekin, and B. Yener. Learning the topological properties of brain tumors. *IEEE-ACM Transactions on Computational Biology and Bioinformatics*, 2(3):262–270, July 2005.

[15] C. Gunduz-Demir. Mathematical modeling of the malignancy of cancer using graph evolution. *Journal of Mathematical Biosciences*, 209(2):514–527, October 2007.

[16] H.S. Wu, R. Xu, N. Harpaz, D. Burstein, and J. Gil. Segmentation of intestinal gland images with iterative region growing. *Journal of Microscopy*, 220(3):190–204, December 2005.

[17] H.S. Wu, R. Xu, N. Harpaz, D. Burstein, and J. Gil. Segmentation of microscopic images of small intestinal glands with directional 2-d filters. *Analytical and Quantitative Cytology and Histology*, 27(5):291–300, October 2005.

[18] H.S. Wu and J. Gil. A biased median filtering algorithm for segmentation of intestinal cell gland images. *The Scientific World Journal*, 6(1):200–220, February 2006.

[19] S. Naik, S. Doyle, M. Feldman, Tomaszewski J., and A. Madabhushi. Gland segmentation and gleason grading of prostate histology by integrating low-, high-level and domain specific information. In *Proceedings of 2nd Workshop on Microsopic Image Analysis with Applications in Biology*, 2007.

[20] Lillie R.D. *Histopathologic Technic and Practical Histochemistry*. McGraw-Hill, 1965.

[21] Online medical dictionary, http://www.medterms.com.

[22] L. Pham, X. Chenyiang, and L.P. Jerry. A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000.

[23] L. Lucchese and S.K. Mitra. Color image segmentation: A state-of-the-art survey. *Proc. of the Indian National Science Academy(INSA-A)*, 67(2):207–221, 2001.

[24] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, January 1979.

[25] Mao K.Z., Zhao P., and Tan P.H. Supervised learning-based cell image segmentation for p53 immunohistochemistry. *IEEE Transactions on Biomedical Engineering*, 53(6):1153–1163, June 2006.

[26] J.B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

[27] S.H. Park, I.D. Yun, and S.U. Lee. Color image segmentation based on 3-d clustering: Morphological approach. *Journal of Pattern Recognition*, 31(8):1061–1076, August 1998.

[28] J. Wu, H. Yan, and A.N. Chalmers. Color image segmentation using fuzzy clustering and supervised learning. *Journal of Electronic Imaging*, 3(4):397–403, October 1994.

[29] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using em and its application to content-based image retrieval. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, page 675. IEEE Computer Society, 1998.

[30] M. Celenk. Hierarchical color clustering for segmentation of textured images. In *SSST '97: Proceedings of the 29th Southeastern Symposium on System Theory (SSST '97)*, page 483, Washington, DC, USA, 1997. IEEE Computer Society.

[31] Y. Deng and B.S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):800–810, 2001.

[32] B. Cramariuc, M. Gabbouj, and J. Astola. Clustering based region growing algorithm for color image segmentation. In *Proc. of 13th International Conference on Digital Signal Processing*, volume 1, pages 281–297. University of California Press, 1997.

[33] Ali Moghaddamzadeh and Nikolaos G. Bourbakis. A fuzzy region growing approach for segmentation of color images. *Pattern Recognition*, 30(6):867–881, 1997.

[34] T. Carron and P. Lambert. Integration of linguistic knowledge for colour image segmentation. In *Proc. of 8th European Signal Processing Conf. (EUSIPCO'96)*, pages 1729–1732. University of California Press, 1996.

[35] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, January 1988.

[36] G. Healey. Using color for geometry-insensitive segmentation. pages 52–69, 1992.

[37] G. Healey. Segmenting images using normalized color. pages 166–198, 1992.

[38] B.A. Maxwell and S.A. Shafer. Physics-based segmentation of complex objects using multiple hypotheses of image formation. *Comput. Vis. Image Underst.*, 65(2):269–295, 1997.

[39] R. Farjam, H. Soltainan-Zadeh, K. Jafari-Khouzani, and R.A. Zoroofi. An image analysis approach for automatic malignancy determination of prostate pathological images. *Journal of Clinical Cytometry*, 72B(4):227–240, July 2007.

[40] R. Fernandez-Gonzalez, T. Deschamps, A. Idica, R. Malladi, and C. Ortiz-de Solorzano. Automatic segmentation of histological structures in mammary gland tissue sections. *Journal of Biomedical Optics*, 9(3):444–453, May 2004.

[41] R. Fernandez-Gonzalez, MH. Barcellos-Hoff, and C. Ortiz-de Solorzano. Quantitative image analysis in mammary gland biology. *Journal of Mammary Gland Biology and Neoplasia*, 9(4):343–359, October 2004.

[42] B. Susnik, A. Worth, J. LeRiche, and B. Palcic. Malignancy-associated changes in the breast. changes in chromatin distribution in epithelial cells in normal-appearing tissue adjacent to carcinoma. *Analytical and Quantitative Cytology and Histology*, 17(2):62–68, February 1995.

[43] S.J. Lockett, D. Sudar, C.T. Thompson, D. Pinkel, and J.W. Gray. Efficient, interactive, and three-dimensional segmentation of cell nuclei in thick tissue sections. *Journal of Cytometry*, 31(4):275–286, April 1998.

[44] B. Weyn, G. van-de Wouwer, A. van Daele, P. Scheunders, D. van Dyck, E. van Marck, and W. Jacob. Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. *Journal of Cytometry*, 33(1):32–40, September 1998.

[45] A. Krtolica, C. Ortiz-de Solorzano, S. Lockett, and Campisi J. Quantification of epithelial cells in coculture with fibroblasts by fluorescence image analysis. *Journal of Cytometry*, 49(2):73–82, October 2002.

[46] R. Albert, J.G. Mller, P. Kristen, T. Schindewolf, S. Kneitz, and H. Harms. New method of nuclear grading of tissue sections by means of digital image analysis with prognostic significance for node-negative breast cancer patients. *Journal of Cytometry*, 24(1):140–150, June 1996.

[47] R. Albert, J.G. Mller, P. Kristen, and H. Harms. Objective nuclear grading for node-negative breast cancer patients: comparison of quasi-3d and 2d image-analysis based on light microscopic images. *Lab Invest*, 78(3):247–259, March 1998.

[48] S.J. Lockett and B. Herman. Automatic detection of clustered, fluorescent-stained nuclei by digital image-based cytometry. *Journal of Cytometry*, 17(1):1–12, September 1994.

[49] H.S. Wu, J. Barba, and J. Gil. An iterative algorithm for cell segmentation using short-time fourier transform. *Journal of Microscopy*, 184(2):127–132, November 1996.

[50] F. Schnorrenberg, N. Tsapatsoulis, C.S. Pattichis, C.N. Schizus, S. Kollias, M. Vassiliou, A. Adamou, and K. Kyriacou. Improved detection of breast cancer nuclei using modular neural networks. *IEEE Engineering in Medicine and Biology Magazine*, 19(1):48–63, January 2000.

[51] H.S. Wu, J. Barba, and J. Gil. An efficient semiautomatic algorithm for cell contour extraction. *IEEE Transactions on Biomedical Engineering*, 179(1):270–276, January 1995.

[52] A.J. Einstein, J. Gil, S. Wallenstein, C.A. Bodian, M. Sanchez, D.E. Burstein, H.S. Wu, and Z. Liu. Reproducibility and accuracy of interactive segmentation procedures for image analysis in cytology. *Journal of Microscopy*, 188(2):136–148, November 1997.

[53] H.S. Wu, J. Barba, and J. Gil. A parametric fitting algorithm for segmentation of cell images. *IEEE Transactions on Biomedical Engineering*, 45(3):400–407, March 1998.

[54] G. Cong and B. Parvin. Model-based segmentation of nuclei. *Pattern Recognition*, 33(8):1383–1393, August 2000.

[55] Q. Yang and B. Parvin. Harmonic cut and regularized centroid transform for localization of subcellular structures. *IEEE Transactions on Biomedical Engineering*, 50(4):469–475, April 2003.

[56] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991.

[57] P. Belhomme, A. Elmoataz, P. Herlin, and D. Bloyet. Generalized region growing operator with optimal scanning: application to segmentation of breast cancer images. *Journal of Microscopy*, 186(1):41–50, April 1997.

[58] N. Malpica, C. Ortiz-de Solorzano, J.J. Vaquero, A. Santos, I. Vallcorba, J.M. Garcia-Sagredo, and F. del Pozo. Applying watershed algorithms to the segmentation of clustered nuclei. *Journal of Cytometry*, 28(4):289–297, August 1997.

[59] C. Ortiz-de Solorzano, E. Garcia-Rodrigues, A. Jones, D. Pinkel, J.W. Gray, D. Sudar, and S.J. Lockett. Segmentation of confocal microscope images of cell nuclei in thick tissue sections. *Journal of Microscopy*, 193(3):212–226, March 1999.

[60] K. Chin, C. Ortiz-de Solorzano, Jones A. Knowles, D. and, W. Chou, E.G. Rodriguez, W.L. Kuo, B.M. Ljung, K. Chew, K. Myambo, M. Miranda, S. Krig, J. Garbe, M. Stampfer, P. Yaswen, J.W. Gray, and S.J. Lockett. In situ analyses of genome instability in breast cancer. *Journal of Nature Genetics*, 36(9):984–988, September 2004.

[61] C. Wahlby, I.M. Sintorn, F. Erlandsson, G. Borgefors, and E. Bengtsson. Combining intensity, edge and shape information for 2d and 3d segmentation of cell nuclei in tissue sections. *Journal of Microscopy*, 215(1):67–76, July 2004.

[62] K. Rajpoot, N. Rajpoot, and M. Turner. Hyperspectral colon tissue cell classification. In *Proc. SPIE Medical Imaging*, volume 1, pages 1–2. SPIE, January 2004.

[63] K. Masood and N. Rajpoot. Hyperspectral colon biopsy classification into normal and malignant categories, 2006.

[64] K. Rajpoot and N. Rajpoot. Svm optimization for hyperspectral colon tissue cell classification. In *Proc. MICCAI 2004*, pages 829–836, 2004.

[65] L. Wenyin. On-line graphics recognition: A brief survey, grec 2003.

[66] T. Shimbashi, Y. Kokubo, and N. Shirota. Region segmentation using edge based circle growing. *ICIP*, 3:3065, 1995.