Gland segmentation of microscopic image for prostate cancer

Han, Dan

2007

Han, D. (2007). Gland segmentation of microscopic image for prostate cancer. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/39166>

<https://doi.org/10.32657/10356/39166>

AS29604 159
2X
D/Blac

Gland Segmentation of Microscopic Image for Prostate Cancer

Han Dan

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Master of Engineering

2007

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

17 sept. 07

Date

Han dan

Han Dan

To my family.

Acknowledgement

First and foremost, I would like to express my gratitude to my advisor, Prof. Mao Kezhi. He has been a constant source of inspiration, advice, and encouragement during my master study, both in my academic life and beyond. I benefited greatly from the unique perspective that he brought into the research leading to this work. I would like to thank him for sharing his profound insight in pattern recognition and image processing, and for his endless words of encouragement. His enthusiasm has always been the driving force behind my research work.

I would like to acknowledge the support from the technicians and other fellow students in the Biomedical Instrumentation Lab for their kind help and discussions during my master study. Particularly, I would like to thank Tang Wenyin and Zhao Peng. They have been a constant and reliable source of wisdom and advice for both my academic research and life.

I would like to thank the School of Electrical and Electronic Engineering, Nanyang Technological University for providing the excellent facilities to support my research work.

I owe everything I have ever achieved to my husband and parents. I will always be grateful to them for their unrelenting love and unconditional support. To them this

thesis is dedicated.

Table of Contents

Acknowledgements	i
Summary	vi
List of Figures	viii
List of Tables	xi
1 Introduction	1
1.1 Background	1
1.1.1 Prostate Cancer	1
1.1.2 Prostate Cancer Diagnosis	3
1.1.3 Prostate Cancer Grading	4
1.2 Motivation	6
1.3 Major Contributions of the Thesis	7
1.4 Organization of the Thesis	8
2 Literature Review	9

TABLE OF CONTENTS	iv
2.1 Overview of Image Segmentation	9
2.1.1 Thresholding	9
2.1.2 Edge-Based Segmentation	11
2.1.3 Region-Based Segmentation	12
2.1.4 Pixel Classification-Based Segmentation	15
2.2 Image Feature Analysis	16
2.2.1 Color Feature Analysis	16
2.2.2 Texture Feature Analysis	18
2.2.3 Shape Feature Analysis	22
2.3 Feature Selection	23
2.4 Classification	26
2.5 Segmentation Performance Evaluation	30
3 Experimental Studies of Commonly Used Segmentation Methods	35
3.1 Introduction	35
3.2 Experimental Studies of Thresholding Method	36
3.3 Experimental Studies of Edge Detection Method	46
3.4 Experimental Studies of Region-Based Segmentation Method	52
4 Pixel Classification Segmentation for Prostate Microscopic Image	61
4.1 Pixel Classification Segmentation Scheme	61
4.2 Window-Based Histogram Feature Formulation	64

TABLE OF CONTENTS		v
<hr/>		
4.3	Feature Selection	68
4.4	Pixel Classification	71
4.5	Experimental Studies	73
4.6	Experimental Performance Analysis	83
5	Conclusion and Recommendations	85
5.1	Conclusion	85
5.2	Recommendations for Future Research	86
	Bibliography	88

Summary

Computer-aided image processing has attracted much research interest in recent years. In particular, there has been growing interest in processing and analyzing biomedical images by computer. Traditionally, biomedical images used for diagnosis are examined visually by experienced pathologists. However, the subjective nature of visual examination frequently results in variant conclusions by different pathologists. Hence, there's an increasing demand for automated method to help improve the efficiency and accuracy of biomedical images diagnosis by incorporating the power of digital image processing techniques into pathological image analysis.

The research work of this thesis focuses on the image segmentation of prostate microscopic images. Prostate cancer is reported as the second leading cause of cancer death in men and the leading cause of cancer death in males over 85. It can only be confirmed by a biopsy, which makes computer-aided image processing an ideal auxiliary tool to help the pathologists make objective diagnosis. This thesis aims to develop a systematic computer-aided method to obtain precise segmentation of abnormal glands from given prostate biopsy images. This research is important and necessary since it constitutes the basis for further computer-aided processing and analyzing of prostate microscopic images. It is easy for human eyes to distinguish the gland from the image due to the color difference between the gland and stroma. However, segmentation of prostate cancer images is far from easy for computer due

to the complex nature of histology images.

In the thesis, the commonly used image segmentation methods like thresholding, edge-based segmentation and region-based segmentation are experimentally evaluated and are shown not suitable for completing this task. A pixel classification segmentation method is then developed to detect the glands in the prostate biopsy image. This method combines the knowledge from pattern classification, feature extraction and selection with the prior knowledge used by pathologist to diagnose prostate cancer. A window-based histogram feature is extracted first based on the unique nature of prostate biopsy image and feature selection algorithm is used to select the proper feature subset from the extracted histogram feature. The support vector machine (SVM) classifier is employed to differentiate the gland points from non-gland points. Experimental studies are conducted to confirm that the proposed method is efficient for the prostate gland segmentation task.

List of Figures

1.1	Anatomy of the Prostate	2
1.2	Prostate Under a Microscope	4
1.3	Gleason Histologic Grading Diagram	5
2.1	The RGB color Cube	17
2.2	Brodatz Texture Collection Diagram	20
2.3	Vistex Texture Collection Diagram	20
2.4	The k -nearest-neighbor query starts at the test points x and grows a spherical region until it encloses k training samples, and it labels the test point by a majority vote of these samples. In this $k=5$ case, the test point x would be labeled the category of the red points.	28
2.5	Segmentation Evaluation Diagram	30
2.6	Segmentation Performance Evaluation	34
3.1	Prostate Biopsy Image	36
3.2	Illustration of thresholding using global histogram	38

LIST OF FIGURES

ix

3.3	Global Histogram of Gray Image Corresponding to Color Prostate Image	38
3.4	Histogram Obtained from R, G, B Channel of the Color Prostate Image	39
3.5	Failed Thresholding Segmentation using Gray Histogram	40
3.6	Thresholding Segmentation Result From Color Histogram	40
3.7	Graphic Illustration	41
3.8	Prostate Biopsy Image	42
3.9	Thresholding Segmentation Result From Color Histogram	43
3.10	Prostate Biopsy Image	43
3.11	Thresholding Segmentation Result From Color Histogram	44
3.12	Prostate Biopsy Image	44
3.13	Thresholding Segmentation Result From Color Histogram	45
3.14	Sobel convolution masks	47
3.15	Original Prostate Biopsy Image	49
3.16	Detected Edges Using Sobel Edge Detector	49
3.17	Apply Sobel Detected Edge to the Original Image	50
3.18	Detected Edges Using Canny Edge Detection	50
3.19	Apply Canny Detected Edge to the Original Image	51
3.20	Region splitting and merging quatree	53
3.21	Example of region splitting and merging	54
3.22	Original Image	55

LIST OF FIGURES	x
3.23 Region Segmentation Result ($t_s = 600, t_m = 300, t_g = 30$)	56
3.24 Original Image	56
3.25 Region Segmentation Result ($t_s = 600, t_m = 300, t_g = 30$)	57
3.26 Original Image	57
3.27 Region Segmentation Result ($t_s = 600, t_m = 300, t_g = 30$)	58
3.28 Original Image	58
3.29 Region Segmentation Result ($t_s = 600, t_m = 300, t_g = 30$)	59
4.1 Pixel Classification Segmentation Scheme	63
4.2 Image of prostate needle biopsy	64
4.3 Histogram of point 1 in Fig 4.2	66
4.4 Histogram of point 2 in Fig 4.2	66
4.5 Histogram of point 3 in Fig 4.2	67
4.6 Histogram of point 4 in Fig 4.2	67
4.7 Separability of the feature obtained by window 13×13	69
4.8 Separability of the feature obtained by window 19×19	70
4.9 Separability of the feature obtained by window 25×25	70
4.10 Training a support vector machine consists of finding the optimal hyperplane, that is, the one with the maximum distance from the nearest training patterns. The support vectors are those (nearest) patterns, a distance b from the hyperplane. The three support vectors are shown as black dots.	73

<i>LIST OF FIGURES</i>	xi
4.11 Original Image 1	75
4.12 Detected Gland in Image 1	75
4.13 Original Image 2	76
4.14 Detected Gland in Image 2	76
4.15 Original Image 3	77
4.16 Detected Gland in Image 3	77
4.17 Original Image 4	78
4.18 Detected Gland in Image 4	78
4.19 Original Image 5	79
4.20 Detected Gland in Image 5	79
4.21 Original Image 6	80
4.22 Detected Gland in Image 6	80
4.23 Original Image 7	81
4.24 Detected Gland in Image 7	81

List of Tables

3.1 Segmentation Evaluation Table 45

3.2 Segmentation Evaluation Table 59

4.1 Segmentation Evaluation Table (Image 1-4 are the same images as
used in Chapter 3) 82

Chapter 1

Introduction

1.1 Background

1.1.1 Prostate Cancer

The prostate is a variable-sized gland located in the male pelvis, usually the size of a walnut measuring 3-4 centimeters long by 3-5 centimeters in width. On average the gland weighs about 20 grams and it consists of a base, an apex, an anterior, a posterior and two lateral surfaces. It is found behind the pubic bone, in front of the rectum, and below the bladder, surrounded by the pelvic muscles. Its main function is to store and secrete a clear, slightly basic fluid that constitutes up to one-third of the volume of semen [1] [85] . Figure 1.1 illustrates the anatomy of the prostate [88].

Figure 1.1: Anatomy of the Prostate

Prostate cancer is a malignant (cancerous) tumor (growth) that consists of cells from the prostate gland. The tumor usually grows slowly and remains confined to the gland for many years. During this time, the tumor produces little or no symptoms or outward signs (abnormalities on physical examination). As the cancer advances, however, it can spread beyond the prostate into the surrounding tissues (local spread). Moreover, the cancer also can metastasize (spread even farther) throughout other areas of the body, such as the bones, lungs, and liver.

It has been reported that prostate cancer is the second leading cause of cancer death in men, following only lung cancer, and the leading cause of cancer death in males over 85 [12] [87]. It's estimated that a man has about a 1 in 10 chance of being diagnosed with prostate cancer and one of every six men will get prostate cancer during his lifetime. Over 200,000 new cases of prostate cancer are diagnosed annually

in the U.S and it is the second most frequently diagnosed cancer in American men, following only skin cancer [109]. In Europe, an estimated number of 2.6 million new cases are diagnosed each year. Although prostate cancer is most common in men over 65, it has been diagnosed in men as young as 40. Among all the cases, prostate cancer constitutes about 11% of all male cancers and accounts for 9% of all cancer deaths among men within the European Union [3].

1.1.2 Prostate Cancer Diagnosis

Although the exact causes of prostate cancer is still unknown, prostate cancer can be diagnosed through a series of tests and exams. These tests mainly include Digital Rectal Exam (DRE), Prostate-Specific Antigen (PSA) Test and Transrectal Ultrasound (TRUS) [42]. In a DRE, the doctor can feel the prostate by inserting a gloved, lubricated finger into the rectum. It helps the doctor to determine if the prostate is enlarged, or has lumps or other types of abnormal texture. Healthy prostate tissue is soft, like the fleshy tissue of the hand where the thumb joins the palm. Malignant tissue is firm, hard, and often asymmetrical or stony, like the bridge of the nose. However, as many as one-third of patients diagnosed with prostate cancer have a normal DRE [29]. PSA is a glycoprotein manufactured almost exclusively by the prostate gland. PSA is normally present in the blood at very low levels; normal PSA levels are defined as between 0-4.0 nanogram per milliliter [83]. Increased levels of PSA may suggest the presence of prostate cancer. Transrectal Ultrasound (TRUS) is used to detect abnormal prostate growth and to guide a biopsy of the abnormal prostate area. This procedure uses sound waves to create an image of the prostate. The image is projected on a monitor, so the physician can examine the gland and surrounding tissue for tumors. When PSA exceeds a normal value, only a biopsy can definitely confirm prostate cancer. With a biopsy a sample of tissue is taken out and examined under a microscope to check

for cancerous changes. Figure 1.2 shows the image of the microscopic glands of the prostate [111].

Figure 1.2: Prostate Under a Microscope

1.1.3 Prostate Cancer Grading

When a biopsy confirms the presence of prostate cancer, the tumor is examined under a microscope and categorized into grades. The most commonly used prognostic

indicator of patient outcome is the Gleason scoring system [47]. This system is based on the match of glandular architectural patterns with standard patterns described by D. F. Gleason during the 1970's. The Gleason grading system constitutes the most powerful tool for Urologists, Oncologists and Pathologists to determine the long-term prognosis of prostate cancer cases. This system defines a five level scoring system based on patterns of glandular organization and currently has a profound impact on patient treatment. Figure 1.3 [34] is a reproduction of Gleason's original diagram showing the five patterns he recognized, ranging from pattern 1 (well differentiated) to pattern 5 (poorly differentiated). The grading is based solely on the degree of glandular organization and specifically excludes other features such as nuclear density and shape. The Gleason method has been recommended by the prostate cancer working group operating under the National Cancer Institute as a common reference system for histological grading which should accompany any other grading system being presented in the literature [34].

Figure 1.3: Gleason Histologic Grading Diagram

1.2 Motivation

Due to the high threat of prostate cancer to health of men, especially of elderly people, more and more research interest has been drawn to find efficient and accurate methods to assist doctors in diagnosing and treating prostate cancer.

The Gleason sorting system, as a commonly used prognostic indicator, defines a five level grading system and currently has a profound impact on patient treatment. However, the Gleason sorting system traditionally has been relying on pathologist's own experience to conduct a visual evaluation of the biopsy sample images, which unfortunately leads to the fact that the subjective nature of visual grading frequently results in variation in scoring by different pathologists. Moreover, due to the lack of effective automated methods prostate biopsy images are typically annotated in a manual way which is cumbersome, error prone, and prohibitively expensive especially when a large number of prostate images are produced everyday. Hence, there's an increasing demand for automated method to help improve the efficiency and accuracy of prostate cancer diagnosis by incorporating the power of digital image processing techniques into pathological image analysis.

Several image analysis approaches [61] have been proposed for this purpose. However these approaches relied on the analysis of the cellular information instead of gland information on which the Gleason scoring system is based. This greatly limits their applications as in practice most pathologists adopt Gleason scoring system in prostate cancer diagnosis. As mentioned in last subsection, the Gleason grading system uses prostatic gland architecture in determining the tumor grades. Therefore, if accurate prostate gland segmentation can be achieved it will provide the pathologists objective information about the prostate biopsy image and provide the pathologists more assistance compared with the cellular information as by the afore-

mentioned approaches [61]. Accurate prostate gland segmentation can also be used as the basis of further analysis of the gland itself. Hence development of method for prostate gland detection and segmentation is selected as the topic of this thesis.

1.3 Major Contributions of the Thesis

The major contributions of this thesis are summarized as follows:

- As far as our knowledge, there has been no study in literature devoted to the specific topic of gland segmentation for prostate cancer using microscopic images. Published studies about prostate image segmentation mainly utilize MR (Magnetic Resonance) images or Ultrasound images, contrastingly the method proposed in this thesis is devoted to segmentation of microscopic image for prostate cancer which has not been covered by existing literature.
- Based on the characteristics of prostate biopsy images, a pixel classification based segmentation method is developed to detect the prostate glands. The window-based histogram feature is proposed to detect the prostate gland. This is motivated by the difference between the gland point and non-gland point. By using the histogram as a feature, techniques of pattern classification, feature extraction and selection area were combined to perform gland segmentation.
- The commonly used image segmentation methods (thresholding, edge-based segmentation and region-based segmentation) are experimentally evaluated. The experiment studies show that they are not suitable for prostate glands segmentation task.

- Experimental studies are carried out to verify the effectiveness of the proposed method. The method is easy to implement in practice. Results with good accuracy are obtained. Segmentation performance evaluation is also provided.

1.4 Organization of the Thesis

The remainder of this thesis is organized as follows:

- Chapter 2 provides a literature review on image segmentation methods. Commonly used image features like color, texture and shape are presented. Feature selection and classification methods are briefly introduced and methods to evaluate segmentation performance are also described.
- Chapter 3 presents the experimental studies to segment glands from prostate biopsy images using some commonly used segmentation methods including thresholding, edge-based segmentation and region-based segmentation.
- Chapter 4 proposes a pixel classification segmentation scheme first, then the detailed method using window-based histogram feature is elaborated. Experimental results are provided to confirm the performance of the proposed method.
- Chapter 5 concludes this thesis and suggests some future research topics.

Chapter 2

Literature Review

2.1 Overview of Image Segmentation

Image segmentation is a major step in image processing. Segmentation divides an image into constituent regions or objects. The level to which the subdivision is carried depends on the problem being solved, i.e., segmentation should stop when the objects of interest in an application have been isolated. In this section, a brief review on major image segmentation techniques is introduced.

2.1.1 Thresholding

Thresholding is one of the old, simple and popular techniques for image segmentation [37]. If only one threshold is used for the entire image then it is called global thresholding. On the other hand, when the image is partitioned into several subregions and a threshold is determined for each of the subregions, it is referred to as local thresholding [108]. Some authors [84, 115] call these local thresholding methods adaptive thresholding schemes. Thresholding can also be classified as bilevel thresholding and multithresholding. In bilevel thresholding the image is partitioned

into two regions - object and background. The threshold T is chosen and applied to the image $I[m, n]$ as follow:

$$\begin{aligned} \text{If } I[m, n] \geq T, \quad I[m, n] &= \text{object} \\ \text{Else} \quad I[m, n] &= \text{background} \end{aligned} \quad (2.1)$$

When the image is composed of several objects with different surface characteristics one needs several thresholds for segmentation. This is known as multithresholding. In such a situation we try to get a set of thresholds $(t_1, t_2, \dots, t_k)$ such that all pixels with

$$f(x, y) \in [t_i, t_i + 1), i = 0, 1, \dots, k \quad (2.2)$$

constitute the i th region type.

There are various methods available for threshold selection. For example, Otsu [89] maximize a measure of class separability. He maximized the ratio of the between class variance to the local variance to obtain threshold. Nakagawa and Rosenfeld [84] assumed that the object and background populations are distributed normally with distinct means and standard deviations. Under this assumption they selected the threshold by minimizing the total misclassification error. Kittler and Illingworth [66] under the same assumption of normal mixture, proposed a method which optimizes a criterion function related to average pixel classification error rate that finds out an approximate minimum error threshold. Some thresholding schemes use spatial information. For example, the busyness measure of Weszka and Rosenfeld [114] is dependent on the co-occurrence of adjacent pixels in an image. They minimized the busyness measure in order to arrive at the threshold for segmentation. Deravi and Pal [26] minimized the conditional probability of transition across the boundary between two regions. This method also uses the local information contained in the co-occurrence matrix in the image.

2.1.2 Edge-Based Segmentation

Edge-based segmentations rely on edges found in an image by edge detecting operators - these edges mark image locations of discontinuities in grey level, color, context, etc. If edges are considered as boundaries between segments, it is reasonable to expect that by applying an edge detector to detect the boundaries between segments, and thus the segments in an image, may be achieved. Edge detection refers to the process of identifying and locating sharp discontinuities in an image. Edges are defined as discontinuities in the image intensity due to changes in scene structure. These discontinuities originate from different scene features and can describe the information that an image of the external world contains [46].

There are an extremely large number of edge detection operators available, each designed to be sensitive to certain types of edges. For example, Roberts gradient, Sobel gradient, Prewitt gradient and the Laplacian operator [45, 48, 97]. The gradient operators not only respond to edges but also to isolated points. For Prewitt's operator the response to the diagonal edge is weak, while for Sobel's operator it is not that weak as it gives greater weights to points lying close to the point under consideration. The preceding operators are called the first difference operator. Laplacian is a second difference operator. The Laplacian operator is given by

$$\nabla^2 = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \quad (2.3)$$

It responds strongly to corners, lines, and isolated points. Thus for a noisy picture, unless it has a low contrast, the noise will produce higher Laplacian values than the edges. To improve, Laplacian of Gaussian (LG) operator was proposed. It is normally denoted by $\nabla^2 G$; where Laplacian is given by equation 2.3 and

$$G = e^{-(x^2+y^2)/(2\pi\sigma^2)} \quad (2.4)$$

is a two-dimensional Gaussian distribution, with standard deviation σ . The gaussian part of the LG operator blurs the image, wiping out all structures at scales much smaller than the σ of the Gaussian [79]. Canny edge detector [13] maximized the product of SNR (signal to noise ratio) and the reciprocal of standard deviation (approximate) of the displacement of edge points to maximize simultaneously both good detection and localization criteria.

Within the context of the edge based techniques we can fit also the framework for object segmentation based on color snakes. Snakes or active contours were originally proposed by Kass and Witkin [63] and have received considerable attention since then. The classical snakes approach consists of deforming an initial contour towards the boundary of an object to be detected; the deformation is obtained by minimizing a global energy designed in such a way that its local minimum is attained in correspondence of the boundary of the object.

Ideally, the edge detection should yield pixels lying only on edges. In practice, this set of pixels seldom characterizes an edge completely because of noise, breaks in the edge from nonuniform illumination, and other effects that introduce spurious intensity discontinuities. Edge-based methods center around contour detection: their weakness in connecting together broken contour lines make them, too, prone to failure in the presence of blurring. Thus the image resulting from edge detection usually can not be used as a segmentation result. Other processing steps must follow to combine edges into contours that correspond better with borders in the image.

2.1.3 Region-Based Segmentation

The object of segmentation is to partition an image into regions. Segmentation accomplished via thresholds is based on the distribution of pixel properties, such as gray-level values or color. Edge-based segmentation approaches segmentation

problem by finding boundaries between regions based on discontinuities in gray levels. The region-based segmentation techniques are based on finding the regions directly. In this section we review the split-and-merge techniques and region growing techniques.

A common characteristic of split-and-merge methods is that they start with an initial inhomogeneous partition of the image (usually the initial segment is the image itself) and they keep performing splitting until homogeneous partitions are obtained. A common data structure used to implement this procedure is the quadtree representation [8, 94] which is a multiresolution scheme. After the splitting phase, there usually exist many small and fragmented regions which have to be somehow connected. The merging phase accomplishes this task by associating neighboring regions and guaranteeing that homogeneity requirements are met until maximally connected segments can be produced. The region adjacency graph (RAG) is the data structure commonly adopted in the merging phase [8, 94]. In many algorithms, smoothness and continuity of color regions are enforced with the adoption of a Markov Random Field (MRF) [25, 41] which basically is a stochastic process characterized by the following property: the conditional probability of a particular pixel taking in a certain value is only a function of the neighboring pixels, not of the entire image. Besides, the Hammersley- Clifford theorem establishes the equivalence between MRF's and Gibbs distributions [41].

Numerous variations in the split-and-merge strategies have been investigated. In [14] a k-means algorithm is used for both classifying the pixels in the splitting phase and grouping pattern classes in the merging phase. In [22] the splitting is initially performed by segmenting the luminance and then refined by checking the chrominance homogeneity of the obtained regions; the merging is based on an ad hoc cost function. In [58] the splitting is operated with the watershed transform [30] of the gradient image of the luminance component simplified by a morphological gray-

scale opening [44,94]; the merging step is realized with a Kohonen's self-organizing map (SOM) [8]. Shafarenko et al. [104] apply instead the watershed transform to the $L^*u^*v^*$ gradient of images and merge the patches of the watershed mosaic according to their color contrast until a termination criterion is met. A similar splitting approach is adopted in [100] whereas the merging phase is performed by iteratively processing the RAG constructed upon the resulting oversegmented regions. Also Round et al. [98] employ a split-and-merge strategy for segmentation of skin cancers; the splitting phase is based on a quad-tree representation of the image and the following conservative merging is performed with a RAG.

An homogeneous region of an image may be obtained through a growth process which, starting from a preselected seed, progressively agglomerates points around it satisfying a certain homogeneity criterion; the growth process stops when no more points can be added to the region. As far as gray-level image segmentation is concerned, several region growing strategies can be found in the literature [36, 49,97]. For color images some new interesting strategies for region growing based segmentation have been recently advanced. Tremeau and Borel [110] suggest several different homogeneity criteria operating in RGB coordinates. In a first phase, they generate a certain number of connected regions with a growing process and, in a second phase, they merge all the regions having similar color distributions; after the second phase, the regions have therefore homogeneous colors but they may be disconnected. Kanai [62] develops a segmentation algorithm which resorts to both color and intensity information. The markers (seeds) are extracted from intensity via morphological open-close operations and from color through quantization of the HSV space; joint markers are defined as the sets comprising both kinds of markers. A region growing process based on a watershed algorithm starts from these joint markers. A region merging process eventually reduces the number of segmented regions.

2.1.4 Pixel Classification-Based Segmentation

Pixels classification-based segmentation is another approach to segment images. Most classification techniques [43, 55, 69] ignore the spatial relationships between pixels. Instead, they use statistical measures to classify image pixels into classes and group them into a few homogenous regions. There are two basic approaches to classify pixels. They are supervised classification and unsupervised classification.

In the supervised classification, we act as teacher to supervise the classification process by initially selecting some pixels from each possible class. These pixels constitute the training data. Then the classification algorithm learns from these training data what each class "looks like", and classify every pixel in the image to one of the classes. Hence we can segment out the desired object in the image as one of the classes. In the unsupervised classification, there is no user intervention during image segmentation. Instead, image pixels are clustered into a number of classes based on specific image properties such as, color histogram, pixel values, and image vectors. The problem of clustering can be analytically stated as follows: Let us suppose that we have M patterns $x_1, \dots, x_M$ within a certain pattern space S ; in our case, the space S is a preselected color space while the patterns x_m are the representations of the image pixels within S . The process of clustering consists in determining the regions $S_1, \dots, S_K$ such that every x_m , $m = 1, \dots, M$, belongs to one of these regions and no x_m belongs to two regions at the same time, i.e., $\bigcup_{k=1}^K S_k = S$ and $S_i \cap S_j = \emptyset \forall i \neq j$. The classification of patterns into classes follows the general common sense principle that objects within each class should show a high degree of similarity while across different classes they should exhibit very low affinity. A great many techniques have been proposed in the literature of cluster analysis [2, 8]. A classical technique for color image segmentation is the *k-means* (or *c-means*) algorithm [81], widely adopted also for vector quantization

and data compression.

2.2 Image Feature Analysis

In this thesis, the pixel classification-based segmentation is proposed to accomplish the gland detection task. In this section (2.2) and the following sections (2.3 and 2.4), the literature review about image feature, feature selection and classification are given since these three parts are the central components that constitute the pixel classification segmentation system.

To segment an object out of an image, we must first determine some features that can be used to distinguish the object we are interested in so that we can develop the segmentation algorithm. In this section, we review feature information (color, texture, shape) that can be extracted and characterized in images.

2.2.1 Color Feature Analysis

The use of color in image processing is motivated by two principal factors. First, color is a powerful descriptor that often simplifies object identification and extraction from a scene. Second, humans can discern thousands of color shades and intensities, compared to about only two dozen shades of gray. This second factor is particularly important in manual (i.e., when performed by human) image analysis [46].

Color model (also called color space or color system) is used to facilitate the specification of colors in some standard, generally accept way. In essence, a color model is a specification of a coordinate system and a subspace within that system where each color is represented by a single point. In digital image processing, colour mod-

els such as RGB (red, green, blue), CMY (cyan, magenta, yellow), CMYK (cyan, magenta, yellow, black) and HSI (hue, saturation, intensity) are commonly used. Media that transmit light (such as television) use additive color mixing with primary colors of red, green, and blue, each of which stimulates one of the three types of the eye's color receptors with as little stimulation as possible of the other two. This is called "RGB" color model. Cyan, magenta, yellow are the secondary colors of light or, alternatively, the primary colors of pigments. The RGB values can be obtained easily from a set of CMY values by subtracting the individual CMY values from 1. Following figure illustrates the relationship between RGB model and CMY model.

Figure 2.1: The RGB color Cube

Often a fourth black is added to improve reproduction of some dark colors. That is CMYK color model. HSI model corresponds closely with the way humans describe and interpret color. The HSI model also has the advantage that it decouples the

color and gray-scale information in an image, making it suitable for many of the developed gray-scale techniques.

Color has been used as a low level feature in image retrieval [20, 78, 99]. Many of the existing techniques are based on the color histogram. In [107], except color histogram, they also propose other color features: color depth, average color, eight-color vector, dynamic color vector, HSB average color, HSB eight-color vector and HSB dynamic color vector. According to the characteristics of human color perception that the human eye cannot simultaneously perceive a large number of colors, Ma et al. [6] first proposed a compact color representation in terms of dominant colors for image segmentation and retrieval. The representation they proposed consists of the dominant colors and the corresponding percentage of occurrence of each color:

$$f_c = (c_i, p_i), i = 1, \dots, N, p_i \in [0, 1]. \quad (2.5)$$

where each of the dominant colors, c_i , is a three dimensional vector in RGB space, and p_i are the corresponding percentages. Chen [19] and Wang [113] implement image segmentation using spatially adaptive dominant color features. The extraction of high-level color descriptors is given by A. Mojsilovic [82].

Color is a low level feature and ignores the spatial relationships between pixels. However, color information is a useful feature in image processing as it is conceptually meaningful and easily computed.

2.2.2 Texture Feature Analysis

Texture is a key component of human visual perception. Like color, this makes it an essential feature to consider when processing images. Textures can be considered to be repeating patterns of local variation of pixel intensities. Everyone can recognize

texture, but it is more difficult to define. Unlike color, texture occurs over a region rather than a point. It is normally perceived by intensity levels and as such is orthogonal to color. Texture has quantities such as periodicity and scale; it can be described in terms of direction, coarseness, contrast and so on. It is this that makes texture a particularly interesting facet of images and results in a plethora of ways of extracting texture features.

Textures are homogeneous patterns or spatial arrangements of pixels and can not be sufficiently described by regional intensity or color alone. Textures have statistical and structural properties, or both. They may consist of the structures and/or random placement of elements; but also may be without fundamental subunits. Texture has been found to provide cues to scene depth and surface orientation and have been used to identify contents for many real-world images that have textural characteristics, such as clouds, trees, bricks, hair, and fabric [50,68,76]. For example, figure 2.2 and 2.3 illustrate the wide variety of textures from the Brodatz collection and Vistex collection. However, the appearance of texture in unconstrained imagery is substantially different from that represented in these textures set [43]. Due to the diversity of textures appearing in natural images it is difficult to narrowly define texture.

Figure 2.2: Brodatz Texture Collection Diagram

Figure 2.3: Vistex Texture Collection Diagram

In past decades, a variety of texture models have been proposed [50, 68, 76, 105]. They can be generally divided into two classes as following.

In one class, models are based on heuristic features that strive to quantify human perception of textual patterns, such as coarseness, contrast, and directionality [35]. Coarseness measures the scale of a texture (pebbles versus boulders), contrast describes its vividness, and directionality describes whether it has a favored direction (like grass) or not (like a smooth object). For example, the second moment (see the following equation, where z is a random variable denoting gray levels, L is the number of distinct gray levels, m is the mean value of z , $p(z_i)$ is the corresponding histogram) of the gray-level histogram of an image or region is a measure of gray-level contrast that can be used to establish descriptors of relative smoothness.

$$\mu_2(z) = \sum_{i=0}^{L-1} (z_i - m)^2 p(z_i). \quad (2.6)$$

Another class of models are derived from spatial-frequency distribution and statistical analysis, such as Gabor wavelet and QMF (Quadrature Mirror Filter) wavelet transform [17, 70, 105]. The wavelet transform is a new spatial/frequency analytical tool that allows both structure and spectral analysis of texture features. The texture research community is currently devoting considerable effort to wavelet application in texture analysis. [17, 50, 68] applied wavelet transforms to images and use wavelet transform algorithm in texture classification. The traditional pyramid-type wavelet transform decomposes subsignals in low frequency channels. However, in some textured image the most significant information appears in the middle frequency channels. In such situation, the conventional wavelet transform does not work properly. To overcome this difficulty, Chang and Kuo proposed a multires-

olution texture model based on the tree-structured wavelet transform [17]. Liu and Wang using spectral histogram to address texture classification and image segmentation and have obtained very good results on natural images. [74, 75]

In general image domains, where a single image object is normally composed of multiple non-homogeneous texture regions, the use of textures requires a texture segmentation process first, which remains a challenging and computational intensive task. In addition, texture-based techniques lack robust texture models and correlation with human perception. Nevertheless, texture feature can still help to describe characteristics of spectra pattern in images details.

2.2.3 Shape Feature Analysis

Shape is an important criterion in object level for matching objects based on their profile and physical structure. There are many techniques of shape description and recognition. These methods can be categorized as information preserving or unambiguous, and non-information preserving or ambiguous, depending on whether or not it is possible to reconstruct a reasonable approximation of the object from the shape descriptor [90, 91]. The information preserving methods such as [64, 65] are mostly used in data compression application. The non-information preserving methods are very useful as approximate indicators of shape.

The objective of shape measures (or shape descriptors) is to measure geometric attributes of an object, that can be used for quantifying shapes [46, 59, 97], matching shapes and recognizing objects. Some of the simpler attributes are length, area, and radius of shape boundary, and can be computed by just counting the number of pixels on the boundary. The disadvantage of such simple measures is that they are scale and size dependent, whereas the concept of shape is invariant to scale, size and

orientation. Measures such as corners, Euler number, roundness (or compactness), symmetry, and shape number are invariant to the factors mentioned. The Euler number [46] is defined as difference between the numbers of connected regions and of holes in an object. Roundness [59] is defined as the ratio of the square of perimeter to the area. In [15], a new kernel-based deformable model is proposed for detecting deformable shapes. To incorporate valuable information for shape detection, such as edge orientations into the shape representation.

Generally speaking, shape recognition has two major parts: shape representation and shape matching. Shape representation has been widely discussed in recent decades [6]. Chain coding [10], Fourier descriptor [72], invariant moments [60], autoregressive models [32], polar signatures [71] and syntactic approaches [18] are well known methods of shape representation. Among them, Fourier descriptors, invariant moments, autoregressive models and polar signature methods are independent of translation and rotation. As for shape matching, numerous approaches have also been proposed based upon the shape representation methods.

The theoretically best way to enhance object-specific information contained in images is by segmenting the object in the image using shape features, because shape feature is meaningful and match objects based on their profile and physical structure. But due to lack of mathematically exact definition of shape, shape features are difficult to extract and to be used in segmentation.

2.3 Feature Selection

In pixel classification-based segmentation, image features are used as the inputs to a classifier to realize the segmentation task. These features often have a large

number of components which may exceed hundreds or even hundreds of thousands. So they will inevitably contain information which is either redundant or irrelevant to the classification task [28]. Therefore the fundamental function of the feature selector is to choose a subset of input feature by eliminating feature components with little or no predictive information. Feature selection can significantly improve the comprehensibility of the resulting classifier and often build a model that generalizes better to unseen points. The problem of feature selection is defined as follows: given a set of d features, select a subset of size m that leads to the smallest classification error.

Let Y be the given set of features, with cardinality d and let m represent the desired number of features in the selected subset X , $X \subseteq Y$. Let the feature selection criterion function for the set X be represented by $J(X)$. Let us assume that a higher value of J indicates a better feature subset; a natural choice for the criterion function is $J = (1 - Pc)$, where Pc denotes the classification error. The use of Pc in the criterion function makes feature selection procedures dependent on the specific classifier that is used and the size of the training and test sets [56]. The most straightforward approach to the feature selection problem would require 1) examining all possible subset of size m , and 2) selecting the subset with the largest value of $J(\cdot)$. However, the number of possible subsets grows combinatorially, making this exhaustive search impractical for even moderate values of m and d . Cover and Van Campenhout [24] showed that nonexhaustive sequential feature selection procedure can be guaranteed to produce the optimal subset. They further showed that any ordering of the classification errors of each of the 2^d feature subsets is possible. Therefore, in order to guarantee the optimality of, say, a 12-dimensional feature subset out of 24 available features, approximately 2.7 million possible subsets must be evaluated. The only "optimal" (in terms of a class of monotonic criterion functions) feature selection method which avoids the exhaustive search is based on the branch and bound algo-

rithm. The computational efficiency of this method lies in an effective organization of the search process. This is achieved by using the monotonicity property shared by the majority of feature selection criterion functions; given two features subsets X_1 and X_2 , if $X_1 \subset X_2$, then $J(X_1) < J(X_2)$. By a straightforward application of this property, many combinations of features can be rejected from the set of candidate feature sets [28].

Most of the well-known feature selection methods which have been proposed in the literature [57] include: Exhaustive Search, Branch-and-Bound Search, Best Individual Features, Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), "Plus l take away r " Selection, Sequential Forward Floating Search (SFFS) and Sequential Backward Floating Search (SBFS). Only the first two methods reviewed in previous paragraph guarantee an optimal subset. All other strategies are suboptimal due to the fact that the best pair of features need not contain the best single features [23]. Best Individual Features selection evaluates all the m features individually and selects the best ones with required number. SFS (SBS) adds (deletes) one feature at a time and make the selected features maximize the criterion function. In SFS (SBS), once a feature is retained (deleted), it can't be discarded (brought back) into the optimal subset. "Plus l take away r " selection first enlarge the feature subset by l features using forward selection and then delete r features using backward selection. SFFS and SBFS are generalization of "Plus l take away r " method; the values of l and r are determined automatically and updated dynamically.

2.4 Classification

Once a feature selection procedure finds a proper representation, a classifier can be designed using a number of possible approaches. In practice, the choice of a classifier is a difficult problem and it is often based on which classifiers happen to be available, or best known, to the user.

The simplest and the most intuitive approach to classifier design is based on the concept of similarity: patterns that are similar should be assigned to the same class. So, once a good metric has been established to define similarity, template matching or the minimum distance classifier using a few prototypes per class can classify pattern. The choice of the metric and prototypes is crucial to the success of this approach. In the nearest mean classifier, selecting prototypes is very simple and robust; each pattern class is represented by a single prototype, which is the mean vector of all the training patterns in that class. More advanced techniques for computing prototypes are vector quantization [86], [5] and learning vector quantization [67], and the data reduction methods associated with the one-nearest neighbor decision rule (1-NN), such as editing and condensing [28]. The most straightforward 1-NN rule can be conveniently used as a benchmark for all the other classifiers since it appears to always provide a reasonable classification performance in most applications. Further, as the 1-NN classifier does not require any user-specified parameters (except perhaps the distance metric used to find the nearest neighbor, but Euclidean distance is commonly used), its classification results are implementation independent.

The probabilistic concept is also used for designing pattern classifier. The Bayes decision rule assigns a pattern to the class with the maximum posterior probability. The Bayes formula is shown as follows:

$$P(w_j|x) = \frac{p(x|w_j) \times P(w_j)}{p(x)}. \quad (2.7)$$

where w_j denotes the state of nature. $P(w_j)$ is the posterior probability—the probability of the state of nature being w_j given the feature value x has been measured. $p(x|w_j)$ is the class-conditional probability density function, the probability density function for x given that the state of nature is w . $P(w_j)$ is a priori probability. $p(x)$ can be viewed as merely a scale factor that guarantees that the posterior probability sum to one [33]. In practice, the empirical Bayes decision rule is used: the estimates of the densities are used in place of the true densities. These density estimates are either parametric or nonparametric. Commonly used parametric models are multivariate Gaussian distributions [38] for continuous features, binomial distributions for binary features, and multinormal distributions for integer-valued (and categorical) features.

The two well-known nonparametric decision rules, the k-nearest neighbor (k-NN) rule and the Parzen classifier (the class-conditional densities are replaced by their estimates using the Parzen window approach), while similar in nature, give different results in practice [33]. They both have essentially one free parameter, the number of neighbors k , or the smoothing parameter of the Parzen kernel, both of which can be optimized by a leave-one-out estimate of the error rate. Further, both these classifiers require the computation of the distances between a test pattern and all the patterns in the training set. We give a graphic illustration in Figure 2.4. The most convenient way to avoid these large numbers of computations is by a systematic reduction of the training set, e.g., by vector quantization techniques possibly combined with an optimized metric or kernel [39], [40]. Other possibilities like table-look-up and branch-and-bound methods [31] are less efficient for large dimensionalities.

Figure 2.4: The k -nearest-neighbor query starts at the test points x and grows a spherical region until it encloses k training samples, and it labels the test point by a majority vote of these samples. In this $k=5$ case, the test point x would be labeled the category of the red points.

Another kind of classifiers is to construct decision boundaries directly by optimizing certain error criterion. While this approach depends on the chosen metric, sometimes classifiers of this type may approximate the Bayes classifier asymptotically. The driving force of the training procedure is, however, the minimization of a criterion such as the apparent classification error or the mean squared error (MSE) between the classifier output and some preset target value. A classical example of this type of classifier is Fisher’s linear discriminant. Fisher’s linear discriminant is a classification method that projects high-dimensional data onto a line and performs classification in this one-dimensional space. Take a 2-class case: the projection maximizes the distance between the means of the two classes while minimizing the

variance within each class. This defines the Fisher criterion, which is maximized over all linear projection, w :

$$J(w) = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2}. \quad (2.8)$$

where m represents a mean, s^2 represents a variance, and the subscripts denote the two classes. Another example is the single-layer perceptron, where the separating hyperplane is iteratively updated as a function of the distances of the misclassified patterns from the hyperplane. If the sigmoid function is used in combination with the MSE criterion, as in feed-forward neural nets (also called multilayer perceptrons), the perceptron may show a behavior which is similar to other linear classifiers [95].

A special type of classifier is the decision tree [9], [21], [93], which is trained by an iterative selection of individual features that are most salient at each node of the tree. The criteria for feature selection and tree generation include the information content, the node purity, or Fisher's criterion. During classification, just those features are under consideration that are needed for the test pattern under consideration, so feature selection is implicitly built-in. The most commonly used decision tree classifiers are binary in nature and use a single feature at each node, resulting in decision boundaries that are parallel to the feature axes [103]. Consequently, such decision trees are intrinsically suboptimal for most application.

One of the most interesting recent developments in classifier design is the introduction of support vector classifier by Vapnik [112] which has also been studied by other authors [11, 101, 102]. It is primarily a two-class classifier. The optimization criterion here is the width of the margin between the classes, i.e., the empty area around the decision boundary defined by the distance to the nearest training

patterns. These patterns, called support vectors, finally define the classification function. Their number is minimized by maximizing the margin.

2.5 Segmentation Performance Evaluation

Figure 2.5: Segmentation Evaluation Diagram

Objective and quantitative segmentation evaluation plays an important role in image segmentation [117]. There are two classes of the segmentation evaluation. One is an inter-technique process for ranking the performance of different techniques in segmenting the same type of images that is segmentation comparison. Another is an intra-technique process for recognizing the behavior of the considered technique in segmenting various kinds of images that is segmentation characterization [118].

As shown in Figure 2.5, a simplified version of the general scheme for segmentation and its evaluation [116] is presented. The portion enclosed by the rounded square with dot line corresponds to the image segmentation procedure that can be considered as a black box in which the image to be segmented is the input and the segmented image is the output. Three groups of evaluation methods are distinct: analytical method, (empirical) goodness method and (empirical) discrepancy method. The access points for these three groups of methods are depicted in Figure 2.5 with thick arrows.

The analytical methods treat the algorithms for segmentation directly by considering the principles, requirements, utilities, complexity, etc., of algorithms. Although it looks simple, the analytical study can not obtain all properties of segmentation algorithms. On the other side, both analysis and practice show that the analytical methods themselves can only provide some additional information to that of other methods do and thus are seldomly used in isolation.

The goodness methods evaluate the performance of algorithms indirectly by judging the segmented images with certain quality measures established according to human intuition. Various goodness measures have been proposed to cover different aspects of an "ideal" or a "good" segmentation. Liu and Yang [73] proposed a goodness measure function which is parameter-free. Their evaluation function is empirically defined as:

$$F(I) = \frac{1}{1000(N \times M)} \sqrt{R} \sum_{i=1}^R \frac{e_i^2}{\sqrt{A_i}}, \quad (2.9)$$

where I is the segmented image $N \times M$ the image size, and R the number of regions of the segmented image, while A_i and e_i are, respectively, the area and the average color error of the i th region; e_i is defined as the sum of the Euclidean distances between the RGB color vectors of the pixels of region i and the color vector attributed to region i

in the segmented image. The smaller the value of $F(I)$, the better the segmentation result should be. Borsotti [7] designed two enhanced versions of Liu and Yang's measure. The first modified version enhances the contribution of small region for the penalization factor. The second modified version enhances the contribution of the number of small regions and the regions having a large segmentation error. Rosenberger [96] proposed a genetic technique capable of combining the results of several segmentations. In this technique, adapted to the texture segmentation case, the fusion is based on an evaluation criterion that takes into the consideration of both intra-region homogeneity and inter-region distinction.

The discrepancy methods count the difference between an actually segmented image and a correctly or ideally segmented image (reference image, also called gold standard or ground truth) to assess the performance of segmentation algorithms. In other words, these methods try to determine how far the actually segmented image is from the reference image. Comparative experiments have shown that the discrepancy methods are more effective than the goodness methods [116]. Betanzos [4] defined an accuracy measure for the case of segmenting images with multi-types of object. Suppose the image contains N types of object, then the accuracy measure is computed by:

$$Accuracy = \sum_{i=1}^N \frac{\text{correct segmented pixels in } i^{\text{th}} \text{ object}}{\text{total number of pixels in } i^{\text{th}} \text{ object}}. \quad (2.10)$$

Chang [16] used two criteria in evaluation, one is a pixel-based measure - misclassified pixel [116], and another is a region-based measure that is adapted from [51]. Huo [54] evaluated the performance of the segmentation algorithm in two steps: (1) computer the area of overlap between segmented regions with expert-identified region; (2) substitute the segmented boundary by expert-identified one, and compare

the results. The area of overlap for region A and B is computed as:

$$\text{area of overlap} = \frac{A \cap B}{A \cup B}. \quad (2.11)$$

In our research work, the discrepancy method is employed to evaluate the segmentation performance. They are area error metrics [77], which determine the discrepancy between the ideally or manually segmented image and the actually segmented image, are applied. For area error metrics, three commonly used measures are false positive (FP) volume fraction, false negative (FN) volume fraction, and true positive (TP) volume fraction where A_m denotes the area of the glands determined by manual segmentation and A_c is the area of the glands determined by computer-aided approaches. $\cup$ and $\cap$ denote union and intersection operations, respectively.

$$\begin{cases} FP = \frac{A_c \cup A_m - A_m}{A_m}, \\ FN = \frac{A_c \cup A_m - A_c}{A_m}, \\ TP = \frac{A_c \cap A_m}{A_m}, \end{cases} \quad (2.12)$$

The graphic illustration is shown in following:

Figure 2.6: Segmentation Performance Evaluation

Chapter 3

Experimental Studies of Commonly Used Segmentation Methods

3.1 Introduction

The quality of the prostate glands segmentation has an determinant influence on the success of computer-aided prostate cancer image analysis. From the prostate gland image as illustrated in *Chapter1*, the glands can be easily distinguished from the image by human eyes due to the color difference between the gland and stroma. However, to computer, segmentation of prostate cancer images is far from easy due to the complex nature of histology images. In this chapter, experimental studies will be conducted using the three commonly used segmentation methods as introduced in *Chapter2* to test their abilities to accomplish the specific segmentation task. The area error metrics [77] introduced in last chapter is employed to give the segmentation performance evaluation.

3.2 Experimental Studies of Thresholding Method

In the analysis of the objects in images it is essential that we can distinguish between the objects of interest and “the rest.” This latter group is also referred to as the background. For our problem, the object of interest is the prostate gland and “the rest” or background is the area between glands, i.e. stroma. A prostate biopsy image is shown in figure 3.1 for illustration. From this image, it can be observed that the gland has darker color than the stroma. As introduced in *Chapter2*, thresholding can be used to segment the darker object out of the brighter background due to their dichotomous nature. Hence, in this section experimental study is conducted to test its performance in segmenting the prostate biopsy image.

Figure 3.1: Prostate Biopsy Image

In thresholding implementation, the segmentation is determined by a single parameter known as threshold. Each pixel in the image is compared with this threshold.

If the pixel's intensity is higher than the threshold, the pixel is set to, say, white, in the output. If it is less than the threshold, it is set to black. Generally, an intensity histogram is used to determine whether an image can be segmented this way. An image histogram refers to the probability mass function of the image intensities. For gray scale image, the histogram of a image with gray levels in the range $[0, L - 1]$ is a discrete function $h(r_k) = n_k$, where r_k is the k th gray level and n_k is the number of pixels in the image having gray level r_k . This is extended for color images to capture the joint probabilities of the intensities of the three color channels. More formally, the color histogram is defined by,

$$h_{A,B,C} = N \cdot \text{Prob}(A = a, B = b, C = c), \quad (3.1)$$

where A, B and C represent the three color channels (R, G, B) and N is the number of pixels in the image. Computationally, the color histogram is formed by discretizing the colors within an image and counting the number of pixels of each color. The color histogram can be thought of as a set of vectors. For gray-scale images these are two-dimensional vectors. One dimension gives the value of the gray-level and the other the count of pixels at the gray-level. For color images the color histograms are composed of 4-D vectors. This makes color histograms very difficult to visualize. There are several lossy approaches for viewing color histograms, one of the easiest is to view separately the histograms of the color channels. This type of visualization does illustrate some of the salient features of the color histogram. In the following figures, we first give an illustration of the histogram used to do thresholding segmentation, and then we show the histogram obtained from the gray image corresponding to the color image in Figure 3.2, 3.3, 3.4 and from the R, G, B channel of the color image separately.

Figure 3.2: Illustration of thresholding using global histogram

Figure 3.3: Global Histogram of Gray Image Corresponding to Color Prostate Image

Figure 3.4: Histogram Obtained from R, G, B Channel of the Color Prostate Image

From Figure 3.3 and Figure 3.4, it can be concluded that except the histogram of R channel, the other three histograms obviously have two distinct peaks similar to the histogram in Figure 3.2. From distribution of the histogram, it seems possible to use thresholding to segment the image. We hence conduct experimental studies as in the following where Otsu’s histogram thresholding [89] is employed. Two cases are considered. In the first experiment, the original color image is converted to gray image first and then thresholding is applied. For color histogram, we set different thresholds for each channel and combine the thresholding segmentation results from three channels together to obtain the second segmentation result.

Figure 3.5: Failed Thresholding Segmentation using Gray Histogram

Figure 3.6: Thresholding Segmentation Result From Color Histogram

As shown in Figure 3.5, the thresholding result obtained from gray histogram presents a failure to segment the prostate biopsy image. This can be explained by the following reason. Although the histogram shown in Figure 3.3 has two distinct peaks, the meaning represented by these two parts is different from the Figure 3.2. Our object is to detect the darker gland from the brighter stroma, but it is noticed that the lumen of the gland is also bright and even brighter than the stroma. As a result, when we convert the color image into gray image the gland wall and the stroma are darker than the lumen of the gland. Hence the thresholding segmentation algorithm mistakenly takes the brighter lumen and border of the image as the background and takes the gland wall and stroma as the object to be segmented. It can be illustrated in the following figure.

Figure 3.7: Graphic Illustration

In the Figure 3.6, the segmentation result is obtained by thresholding the color histogram. The result shows an obvious improvement over the result got from gray

histogram due to the combination of the color information. But the result is still far from satisfactory because of the existence of the noisy dark dots intersperse among the stroma. It can be explained as that these dots are colorfully similar to the nucleoli in the wall of gland and hence unable to be got rid of through the thresholding algorithm.

We give more color thresholding segmentation result as follows:

Figure 3.8: Prostate Biopsy Image

Figure 3.9: Thresholding Segmentation Result From Color Histogram

Figure 3.10: Prostate Biopsy Image

Figure 3.11: Thresholding Segmentation Result From Color Histogram

Figure 3.12: Prostate Biopsy Image

Figure 3.13: Thresholding Segmentation Result From Color Histogram

Color thresholding segmentation performance evaluation for Figures 3.1, 3.8, 3.10 and 3.12 are shown in the following table.

Table 3.1: Segmentation Evaluation Table

	FP	FN	TP
Figure 3.1	0.1895	0.7593	0.2408
Figure 3.8	0.5840	0.8008	0.1997
Figure 3.10	0.2349	0.5756	0.4246
Figure 3.12	0.2148	0.7288	0.2717
Average	0.3058	0.7161	0.2842

The experimental studies show that good segmentation performance cannot be ob-

tained by directly applying the thresholding method to the prostate biopsy image.

3.3 Experimental Studies of Edge Detection Method

Edge-based segmentation represents a large group of methods based on information about edges in the image; it is one of the earliest segmentation approaches and is still very important. Edge-based segmentations rely on edges found in an image by edge detecting operators, which mark locations of discontinuities in gray level, color, texture, etc. The close contours may be obtained from the regions by edge detection and regions of interest may be obtained from the closed contours by boundary filling. In this section, we will experimentally test the prostate glands segmentation performance of two commonly used edge detectors: Sobel edge detector [106] and canny edge detector [13].

The Sobel operator performs a 2-D spatial gradient measurement on an image and emphasizes regions of high spatial gradient that corresponds to edges. Typically it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image. In theory at least, the operator consists of a pair of 3×3 convolution masks as shown in Figure 3.14. One mask is simply the other rotated by 90° . These masks are designed to respond maximally to edges running vertically and horizontally relative to the pixel grid, one mask for each of the two perpendicular orientations. The masks can be applied separately to the input image, to produce separate measurements of the gradient component in each orientation (call these G_x and G_y). These can then be combined together to find the absolute magnitude of the gradient at each point and the orientation of that gradient. The gradient magnitude is given by:

$$\|G\| = \sqrt{G_x^2 + G_y^2} \quad (3.2)$$

An approximate magnitude is often computed using:

$$\|G\| = \|Gx\| + \|Gy\| \tag{3.3}$$

which is much faster to compute.

The angle of orientation of the edge (relative to the pixel grid) giving rise to the spatial gradient is given by:

$$\theta = \arctan(Gy/Gx) - 3\pi/4 \tag{3.4}$$

In this case, orientation 0 is taken to denote that the direction of maximum contrast from black to white runs from left to right on the image, and other angles are measured anti-clockwise from this.

Figure 3.14: Sobel convolution masks

The Canny edge detector is considered as a "standard method" or as the optimal edge detector. Canny edge detection algorithm followed a list of criteria to improve current methods of edge detection. The first and most obvious is low error rate. It is important that edges occurring in images should not be missed and that there be NO responses to non-edges. The second criterion is that the edge points should be well localized. In other words, the distance between the edge pixels as found by the detector and the actual edge is to be at a minimum. The third criterion is to have only one response to a single edge. This was implemented because the first 2 were not substantial enough to completely eliminate the possibility of multiple responses to an edge. Based on these criteria, the canny edge detector first smoothes the image to eliminate noise. It then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum (nonmaximum suppression). The gradient array is now further reduced by hysteresis. Hysteresis is used to track along the remaining pixels that have not been suppressed. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is set to zero (made a nonedge). If the magnitude is above the high threshold, it is made an edge. And if the magnitude is between the 2 thresholds, then it is set to zero unless there is a path from this pixel to a pixel with a gradient above the high threshold.

The experiments are based on the same prostate biopsy image as shown in Figure 3.1. The experimental results using Sobel and Canny edge detector are shown in Figures 3.16-3.17 and 3.18-3.19 respectively.

Figure 3.15: Original Prostate Biopsy Image

Figure 3.16: Detected Edges Using Sobel Edge Detector

Figure 3.17: Apply Sobel Detected Edge to the Original Image

Figure 3.18: Detected Edges Using Canny Edge Detection

Figure 3.19: Apply Canny Detected Edge to the Original Image

Ideally, it's expected that the edge detector will detect the edges of the whole gland and then the detected edges can be further utilized to segment the whole gland. But as shown in the experimental results, the obtained images by using edge detection fail to present as satisfactory segmentation results. The reason can be explained as: the edge-detecting operators find edges based on the discontinuities in gray level, but due to the brightness contrast between the dark nucleoli and the stroma the edge detector mark the edges of the nucleoli rather than the edges of the whole gland. And because Canny edge detector is more accurate than Sobel edge detector in edge detection, this quality make the Canny detector more prone to be affected by the noisy dark points as can be seen in Figures 3.16 and 3.18. Also it can be perceived that it will be very difficult to meaningfully combine edge segments into edge chains that correspond with borders of the glands in the image, due to the following reasons: the existence of the noisy dark points which are not part of the

gland; the edge detection results are from the nucleoli lying in the wall of the gland instead of the gland itself. So the segmentation can not further proceed. Hence the edge based segmentation is not suitable for the gland detection task.

3.4 Experimental Studies of Region-Based Segmentation Method

The aim of the segmentation methods is to find interested regions or objects; the region-based segmentation methods construct regions directly. Region-based segmentation techniques extract the homogeneous zones from the image. Each region regroups a related set of pixels having at least one common property. Given a more rigorous definition: region-based segmentation is a partition of the image in topological meaning, in order that each region verifies properties (or predicates) of homogeneity. By using a mathematical formalism, we obtain the next definition: region-based segmentation of an image I , using a predicate of homogeneity P , is defined as a partition $S = R_1, R_2, \dots, R_n$ of I such that [27]:

1. $I = \cup R_i, i \in [1, n]$
2. $R_i \cap R_j = \emptyset$ for $i \neq j, i$ and $j \in [1, n]$
3. $P(R_i) = \text{true}$ for $i \in [1, n]$
4. $P(R_i \cup R_j) = \text{false}$ for $i \neq j$ and for all pair of adjacent regions

The predicate P is a constraint, a condition that we impose to differentiate pixels and that a region has therefore to satisfy.

As introduced above, the aim of region-based segmentation techniques is to extract the homogeneous zones from the image. In prostate biopsy image, the wall of

the gland is obviously not homogeneous with the lumen of the gland. But if the region-based segmentation can accurately detect these two separate regions, then it's possible to combine them together to form a complete gland and hence to accomplish the segmentation task.

In the following part of this section, we conduct the experimental studies using split and merge algorithm. The classical split and merge technique proposed by Horowitz and Pavlidis [52] is adopted. This technique uses the pyramidal structure of the quadtree as illustrated in Figure 3.20. On the top of the pyramid, we have the whole image as the same block. To pass from a level to immediately inferior one, each block is divided in four square blocks of same size. At the pyramid basis, each block corresponds therefore to a pixel of the image. Figure 3.21 give the illustration of region splitting and merging [46,80].

Figure 3.20: Region splitting and merging quadtree

Figure 3.21: Example of region splitting and merging

The split and merge algorithm works in two steps:

1. Split the image into homogeneous regions (based on quadrees; it splits if following condition is true: if $((V_r + V_g + V_b) \geq t_s)$ then split, where V is the variance of region in one color channel).
2. Merge regions (based on quadrees; it merges regions if the following condition is true: if $(\|MR_i - MR_j\| + \|MG_i - MG_j\| + \|MB_i - MB_j\| \leq t_m)$ then merge, where M is the mean color of region in one color channel).

where t_s denotes the threshold for split, higher values for t_s means less splitting of regions; t_m denotes threshold for merge, higher values for t_m means more merged regions [92].

To further improve the segmentation quality, we group regions based on the splitting and merging results. Region growing technique melts two regions if the following condition is true: if $(\|M_r - m_r\| + \|M_g - m_g\| + \|M_b - m_b\| \leq t_g)$ then group, where M and m is the mean color of the two regions in one channel. Higher values for t_g results more merged regions.

Following the split, merge and group algorithm, experimental studies were conducted with different groups of t_s , t_m and t_g values. Among the obtained experimental results, several are presented as follows:

Figure 3.22: Original Image

Figure 3.23: Region Segmentation Result ($t_s = 600, t_m = 300, t_g = 30$)

Figure 3.24: Original Image

Figure 3.25: Region Segmentation Result ($t_s = 600, t_m = 300, t_g = 30$)

Figure 3.26: Original Image

Figure 3.27: Region Segmentation Result ($t_s = 600, t_m = 300, t_g = 30$)

Figure 3.28: Original Image

Figure 3.29: Region Segmentation Result ($t_s = 600, t_m = 300, t_g = 30$)

Region-based segmentation performance evaluation for Figures 3.22, 3.24, 3.26 and 3.28 are shown in the following table.

Table 3.2: Segmentation Evaluation Table			
	FP	FN	TP
Figure 3.22	0.5429	0.1236	0.8765
Figure 3.24	1.8466	0.3791	0.6205
Figure 3.26	0.6703	0.1696	0.8304
Figure 3.28	0.4358	0.2476	0.7523
Average	0.8739	0.2299	0.7699

According to the experimental result, it can be noted that the split and merge method does not lead a good segmentation result. The resulting image suffers from “block effect”. This method would suit better for more homogeneous images or for images having a large proportion of rectangular primitive, but not suitable for textual image. However, the prostate biopsy images intrinsically have complex texture and hence the split and merge algorithm does not constitute an efficient solution for the prostate glands segmentation problem.

Chapter 4

Pixel Classification Segmentation for Prostate Microscopic Image

In the previous Chapter, classical segmentation methods - thresholding, edge-based segmentation and region-based segmentation - have been experimentally studied to solve the prostate glands segmentation problem. But the results obtained are not satisfactory. In this chapter, we propose a segmentation method based on pixel classification, with the goal of achieving improved segmentation. The segmentation algorithm is presented in details in Section 4.2 - 4.4 and experimental studies using the proposed method are provided in Section 4.5. At last, the segmentation performance is evaluated.

4.1 Pixel Classification Segmentation Scheme

In *Chapter3*, it's shown through experimental studies that the classical segmentation methods are not applicable to the prostate biopsy images due to their intrinsic limitations of segmentation concepts. Hence, we examine this segmentation task

from another perspective: if the pathologists' visual judgement criteria can be represented by certain image feature such that a pixel in the image can be identified as a gland pixel or non-gland pixel, then pixel classification can be carried out to extract the glands out from the image.

The segmentation scheme can be illustrated by the following figure 4.1. The feature formulation step aims to find the feature that can be used to represent the image pixels meaningfully and can give the good separability between two different kinds of pixels: gland pixel and non-gland pixel. Feature selection step is conducted to extract with appropriate tools the most useful information from the feature vector. The pixel classification step then determines for each pixel of the image whether it belongs to gland class or non-gland class. In evaluation step, the discrepancy method is adopted to evaluate the segmentation performance, which gives the segmentation accuracy of the segmented object in relation to a ground truth reference image.

Figure 4.1: Pixel Classification Segmentation Scheme

4.2 Window-Based Histogram Feature Formulation

Through the experimental results in the previous chapter, it can be concluded that the color information obtained from single pixel or the brightness difference between single pixels are not effective enough to segment the gland out of the image. By analyzing the image of the prostate needle biopsy, it's found that the points on the gland and their neighbors form continuous darker zone. This provides a clue to differentiate the pixels which belong to and not belong to the gland. In other words, the features used to distinguish the gland points and non-gland points are based on the neighboring area of the pixel rather than the single pixel.

Figure 4.2: Image of prostate needle biopsy

As introduced, the histogram of a digital image with gray levels in the range $[0, L-1]$

is a discrete function $h(r_k) = n_k$, where r_k is the k th gray level and n_k is the number of pixels in the image having gray level r_k . It is reasonable to expect that the histogram of a dark region will be different from the histogram of a bright region. For a dark region, the histogram will concentrate in the low gray level area. On the other hand, for the bright region, the histogram will be concentrate in the high gray level area. The observation of the visual difference between points on gland and those not on gland enable us to propose the following method to acquire the feature that can distinguish whether a point locates on gland or not. First a gray image is obtained from the original color image, then an appropriate window should be chosen and centered on each point in the gray image of prostate needle biopsy sequentially, next histograms are obtained according to each point and used as the feature vector to discriminate the points. Too big or small window size will weaken the feature separability. As an example, Fig 4.2 shows a gray image of prostate needle biopsy. Four points on the image are choosen as indicated by points 1, 2, 3 and 4. Points 1 and 2 are example points not on the gland and point 3 and 4 are example points on the gland. The histograms of points in a window centered at points 1-4 are illustrated in Fig 4.3-4.6 respectively.

Figure 4.3: Histogram of point 1 in Fig 4.2

Figure 4.4: Histogram of point 2 in Fig 4.2

Figure 4.5: Histogram of point 3 in Fig 4.2

Figure 4.6: Histogram of point 4 in Fig 4.2

4.3 Feature Selection

The histogram feature vector have all 256 components. These components will inevitably contain information which is either redundant or irrelevant to the classification task. It is therefore the fundamental function of the feature selector to extract with appropriate tools the most useful information from the representation vector and present it in a form of a pattern vector of a lower dimensionality whose elements represent only the most significant aspects of the input data. In this study, sequential forward selection (SFS) is adopted to select the desired feature subset and Mahalanobis Distance Measure is used as the criterion function.

SFS is a bottom up search procedure where one measurement at a time is added to the current feature set. At each stage, the attribute to be included in the feature set is selected from among the remaining available measurements so that the new enlarged set of features yields a maximum value of the criterion function used. Suppose k features have been selected to form feature set X_k . Rank the elements, ξ_j , of the set of available measurements, $Y - X_k$, so that

$$J(X_k + \xi_1) \geq J(X_k + \xi_2) \geq \cdots \geq J(X_k + \xi_{D-k}). \quad (4.1)$$

Then feature set X_{k+1} is given as

$$X_{k+1} = X_k + \xi_1. \quad (4.2)$$

The algorithm is initialized by setting X_0 is empty. The SFS algorithm selects successive attributes with reference to the current set of features.

The quality of a feature subset can be evaluated based on its ability to provide large class separation. The Mahalanobis distance measure is such a criteria for

class separability measure. For a two-class classification problem, the Mahalanobis distance measure is defined as (see, for example, [28])

$$J(i, j) = [m_i - m_j]^T C_{ij}^{-1} [m_i - m_j], \tag{4.3}$$

where $m_i = [m_{1i}, m_{2i}, \dots, m_{ni}]^T$ is the mean vector of samples in class i . $C_{ij} = C_i + C_j$, and C_i and C_j are the covariance matrices of classes i and j , respectively.

As mentioned in last section, to obtain the good histogram feature an appropriate window size should be chosen, too big or small window size will weaken the feature separability. To give an illustration, a comparison of the separability of histograms obtained using three different window size 13×13 , 19×19 , 25×25 are shown by Fig 4.7-4.9. By comparison, 25×25 window size is adopted in the research work.

Figure 4.7: Separability of the feature obtained by window 13×13

Figure 4.8: Separability of the feature obtained by window 19×19

Figure 4.9: Separability of the feature obtained by window 25×25

4.4 Pixel Classification

The ultimate goal of the segmentation system is to fulfill the image segmentation through pixel classification. The task of the classifier component of the full system is to use the feature vector provided by the feature selector to assign the object to a category so that the segmentation can be realized. As to our problem that is to implement a classifier to use the feature vector selected by SFS to separate the pixels in the image of prostate biopsy into two categories: one is the class in which the pixels belong to the gland, one is not. In this study the SVM classifier is adopted to carry out pixel classification.

Support vector machines rely on preprocessing the data to represent patterns in a high dimension-typically much higher than the original feature space. With an appropriate nonlinear mapping $\varphi(\cdot)$ to a sufficiently high dimension, data from two categories can always be separated by a hyperplane. Here we assume each pattern x_k has been transformed to $y_k = \varphi(x_k)$; in general RBF Kernel is a reasonable first choice of $\varphi(\cdot)$ ([53]). For each of the n patterns, $1, 2, \dots, n$, we let $z_k = \pm 1$, according to whether pattern k is in ω_1 or ω_2 . A linear discriminant in an augmented y space is

$$g(y) = a_t y, \quad (4.4)$$

where both the weigh vector and the transformed pattern vector are augmented. Thus a separating hyperplane ensures

$$z_k g(y_k) \geq 1, k = 1, \dots, n. \quad (4.5)$$

The margin was any positive distance from the decision hyperplane. The goal in training a Support Vector Machine is to find the separating hyperplane with the largest margin: we expect that the larger the margin, the better generalization of

the classifier. The distance from any hyperplane to a (transformed) pattern y is $\frac{\|g(y)\|}{\|a\|}$, and assuming that a positive margin b exists, Eq 4.6 implies

$$\frac{z_k g(y_k)}{\|a\|} \geq b, k = 1, \dots, n. \quad (4.6)$$

the goal is to find the weight vector a that maximizes the b . Of course, the solution vector can be scaled arbitrarily and still preserve the hyperplane, and thus to ensure uniqueness we impose the constraint $b = 1$; that is, we demand the solution to Eqs 4.4 and 4.5 also minimize $\|a\|$.

The support vectors are the (transformed) training patterns for which Eq 4.5 represents an equality—that is, the support vectors are (equally) close to the hyperplane (Fig 4.10). The support vectors are the training samples that define the optimal separating hyperplane and are the most difficult patterns to classify. Informally speaking, they are the patterns most informative for the classification.

Figure 4.10: Training a support vector machine consists of finding the optimal hyperplane, that is, the one with the maximum distance from the nearest training patterns. The support vectors are those (nearest) patterns, a distance b from the hyperplane. The three support vectors are shown as black dots.

4.5 Experimental Studies

After getting a whole notion of used methods, we will describe the procedure of the prostate gland detection and present the experimental results.

Firstly, we should choose an appropriate window size. As we have shown before, the separability of histogram feature obtained by small and large window are both not good. Through experiments we found that the window size 25×25 is appropriate.

Secondly, we get the training data. We choose ten prostate needle biopsy images and sample two different kinds of points in the image: one belongs to the gland and one doesn't. Then we put the window centered on each sample point to extract the histogram of the area covered by the window. Corresponding to every sample point, we obtain a histogram feature vector to represent it. All the feature vectors of sample points constitutes the training data.

Then, as explained in section 4.3, the presence in the input data of more redundant or/and irrelevant information detrimentally affects the reliability of the classifier. So in this step we implement feature selection using Sequential Forward Selection.

Next, training the classifier. By this time, we gained the training sample set to estimate the various " parameters " of a classifier. In the training process, the input training data will first be normalized and then be input to the SVM classifier to train it.

Finally, testing the classifier. Now we can use the trained classifier to detect prostate gland. The all points in the testing image are represented by its corresponding histogram. After classification we can get a binary image with the black points that are on the gland and the rest as the white background.

In the following, several examples with the original images and images after gland detection are provided to illustrate the performance of the developed segmentation method.

Figure 4.11: Original Image 1

Figure 4.12: Detected Gland in Image 1

Figure 4.13: Original Image 2

Figure 4.14: Detected Gland in Image 2

Figure 4.15: Original Image 3

Figure 4.16: Detected Gland in Image 3

Figure 4.17: Original Image 4

Figure 4.18: Detected Gland in Image 4

Figure 4.19: Original Image 5

Figure 4.20: Detected Gland in Image 5

Figure 4.21: Original Image 6

Figure 4.22: Detected Gland in Image 6

Figure 4.23: Original Image 7

Figure 4.24: Detected Gland in Image 7

The performance evaluation of the segmentation results are given in the following Table 4.1.

Table 4.1: Segmentation Evaluation Table (Image 1-4 are the same images as used in Chapter 3)

	FP	FN	TP
Image No.1	0.5429	0.1984	0.8016
Image No.2	0.3549	0.1738	0.8262
Image No.3	0.1359	0.2028	0.7972
Image No.4	0.2722	0.2543	0.7457
Image No.5	0.0891	0.2736	0.7264
Image No.6	0.5245	0.2145	0.7855
Image No.7	0.3701	0.1938	0.8062
Average of Image 1-4	0.3264	0.2073	0.7927

The performance of segmentation method for certain task is normally numerically evaluated by several criteria: FP (false positive, as illustrated in Figure 2.6) is expected as small as possible, FN (false negative, as illustrated in Figure 2.6) is expected as small as possible and TP (true positive, as illustrated in Figure 2.6) is expected as large as possible. The FP/FN/TP values should not be treated separately but considered together to give an overall and more meaningful evaluation of the performance. As shown in Table 3.1, 3.2, 4.1, the average FN and TP values for our method is the best among all methods under comparison; the threshold method possesses the best average FP value which is very close to our method but suffers very poor FN and TP average values. Hence through overall numerical evaluation our method is shown to give the best performance, and actually this can

also be easily confirmed through visual examination of the segmented results.

So far, there has been no recognized and rigorous measure on effectiveness of segmentation methods for prostate microscopic images. But from practical clinic research point of view, the segmentation results obtained through our method are satisfactory in the sense that glands are clearly segmented, comparatively the commonly used methods studied in Chapter 3 fail to do a good job.

4.6 Experimental Performance Analysis

Compared with the experimental performance of the three commonly used methods implemented in prostate microscopic image segmentation in Chapter 3, our proposed segmentation method is superior as confirmed by the experiment studies in this chapter.

Thresholding algorithm is unable to get rid of the existence of the noisy dark dots intersperse among the stroma because these dots are colorfully similar to the nucleoli in the wall of gland. The edge-detecting method find edges based on the discontinuities in gray level, but due to the brightness contrast between the dark nucleoli and the stroma the edge detector mark the edges of the nucleoli rather than the edges of the whole gland. The resulting image from region-based method suffers from “block effect”. This method would suit better for more homogeneous images or for images having a large proportion of rectangular primitive, but not suitable for textual image.

Contrasting the aforementioned three methods, our method did better job for this specific task because it combines the techniques from pattern classification, feature extraction and selection with the prior knowledge used by pathologist to diagnose prostate cancer. A window-based histogram feature is extracted first based on the

unique nature of prostate biopsy image and sequential forward selection algorithm is implemented to select the proper feature subset from the histogram feature extracted. As a result, the gland points and non-gland points can be differentiated efficiently by support vector machine (SVM) classifier.

Chapter 5

Conclusion and Recommendations

5.1 Conclusion

In this thesis, the gland segmentation problem for prostate biopsy images is studied. Firstly, some commonly used image segmentation methods like thresholding, edge-based segmentation and region-based segmentation are tested through experimental studies, and the results show that they are not suitable for completing this task. A pixel classification segmentation method is then developed to detect the glands in the prostate biopsy image.

The proposed method combines the techniques from pattern classification, feature extraction and selection with the prior knowledge used by pathologist to diagnose prostate cancer. A window-based histogram feature is extracted first based on the unique nature of prostate biopsy image and sequential forward selection (SFS) algorithm is used to select the proper feature subset from the histogram feature extracted. The support vector machine (SVM) classifier is employed to differentiate the gland points from non-gland points.

Experimental studies are carried out to show that the proposed method is efficient in the sense that only the glands of interest are displayed and the segmentation accuracy is quite high.

5.2 Recommendations for Future Research

In this section, we present some recommendations for future research:

1. An important aspect of the pixel-classification segmentation approach is the choice of appropriate feature to provide medically useful categorizations. Although we have had good initial performance with the window-based histogram feature, we recognize the need to explore more medically informative features in order to avoid the problems of over-simplified algorithm and unpredictable segmentation performance led by the use of only one or two image features. The exploration of how the linear and nonlinear combination of several features can help generate a very effective classification metric leading to efficient segmentation will be implemented.
2. High dimensionality of feature brings computation overhead in the training and classification process and over-fitting is unavoidable as real world problems often do not provide enough data sample to support a complex high dimensionality model. Therefore, it is useful to create a simple model that gives satisfactory performance. This simple model can be constructed from many candidate features through feature selection and dimensionality reduction. We used SFS feature selection method in our research due to its simplicity and computational efficiency, while it is still worthy to explore the possibility of adopting other feature selection and dimensionality reduction methods because of their respective potential advantages.
3. The segmentation results obtained in this thesis can be further utilized and

analyzed to establish a content-based image retrieval system. Content-based image retrieval (CBIR) has attracted much research interest in recent years. In particular, there has been growing interest in indexing biomedical images by content. CBIR is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large database. However, due to the lack of effective automated methods biomedical images are typically annotated in a manual way which is cumbersome, error prone, and prohibitively expensive. Hence the demand for automatically recognizing and retrieving medical images for screening, reference, and management is growing faster than ever. It is also true for prostate images. The use of an objective tool such as the CBIR system will allow pathologists to more accurately grade prostate adenocarcinomas and assist in the evaluation of a broad range of additional conditions. This will be accomplished by allowing pathologists to “check” their visual evaluations against a gold standard collection of historical cases as a result of easy access to an on-line database of high quality images with attached diagnostic annotations. As the first step we have developed a method to detected the prostate glands, then the cellular and glandular differences between prostate cancer of different grades should be analyzed and used to represent the prostate biopsy image in constructing the CBIR system.

Bibliography

- [1] Anatomy. Retrieved June 9, 2006, from [http : //www.prostatehealth.com/anatomy.htm](http://www.prostatehealth.com/anatomy.htm).
- [2] M. R. Anderberg. *Cluster Analysis for Applications*. Academic Press, New York, 1973.
- [3] G. Aus, C.C. Abbou, M.Bolla, A. Heidenreich, H. van Poppel, H-P. Schmid, J.M. Wolff, and F. Zattoni. Guidelines on prostate cancer. *European Association of Urology*, 2005.
- [4] A. A. Betanzos and et al. Analysis and evaluation of hard and fuzzy clustering segmentation techniques in burned patient images. *IVC*, 18(13):1045–1054, 2000.
- [5] S. K. Bhatia and J. S. Doogun. Conceptual clustering in information retrieval. *IEEE Trans. Systems, Man, and Cybernetics*, 28(3):427–436, 1998.
- [6] N. Blatt and J. Rubinstein. The canonical coordinates method for pattern recognition ii. isomorphism with affine transformation. *Pattern Recognition*, 27(1):99–107, 1994.
- [7] M. Borsotti and et al. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letter*, 19(8):741–747, 1998.
- [8] S. T. Bow. *Pattern Recognition and Image Preprocessing*. Marcel Dekker, Inc., New York, 1992.
- [9] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth, Calif., 1984.
- [10] H. L. Bues and S. H. Tu. An improved corner detection algorithm based on chain coded plane curve. *Pattern Recognition*, 20(3):291–296, 1987.

- [11] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- [12] Prostate Cancer. Retrieved June 9, 2006, from <http://www.prostate.com/prostatecancer/>.
- [13] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 8:679–698, Nov. 1986.
- [14] M. Celenk. Hierarchical color clustering for segmentation of textured images. In *Proc. of the 29th Southeastern Symposium on System Theory*, pages 483–487, Cookeville, TN, 9–11, Mar. 1997.
- [15] C. C. Chang. Deformable shape finding with models based on kernel methods. *IEEE Transactions on Image Processing*, 15(9):2743–2754, Sep. 2006.
- [16] K. I. Chang and et al. Evaluation of texture segmentation algorithms. In *Proc. CVPR*, pages 294–299, 1999.
- [17] T. H. Chang and C. C. J. Kuo. A wavelet transform approach to texture analysis. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 661–664, 1992.
- [18] H. H. Chen and J. S. Su. A syntactic approach to shape recognition. In *International Computer Symposium*, Tainan, Taiwan, 1986.
- [19] J. Chen, T. N. Pappas, A. Mojsilovic, and B. Rogowitz. Perceptual multiscale image segmentation based on color and texture features. In *IEEE International Conference on Image Processing (ICIP'04)*, Singapore, Oct. 2004.
- [20] J. Q. Chen, T. N. Pappas, A. Mojsilovic, and B. E. Rogowitz. Adaptive perceptual color-texture image segmentation. *IEEE Transactions on Image Processing*, 14(10):1524–1536, Oct. 2005.
- [21] P. A. Chou. Optimal partitioning for classification and regression trees. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(4):340–354, Apr. 1991.
- [22] V. A. Christopoulos, P. D. Muynck, and J. Cornelis. Colour image segmentation for low bit rate segmented image coding. In *Proc. of the 13th Int'l Conf. on Digital Signal Processing*, pages 861–864, Santorini, Greece, 2–4, July 1997.
- [23] T. M. Cover. The best two independent measurements are not the two best. *IEEE Trans. Systems, Man, Cybernetics*, 4:116–117, 1974.

- [24] T. M. Cover and J. M. Van Campenhout. On the possible orderings in the measurement selection problem. *IEEE Trans. Systems, Man, and Cybernetics*, 7(9):657–661, Sept. 1977.
- [25] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-5(1):25–39, Jan. 1983.
- [26] F. Deravi and S. K. Pal. Gray level thresholding using second-order statistics. *Pattern Recognition Lett.*, 1:417–422, 1983.
- [27] J. C. Devaux, R. K. Kouassi, P. Gouton, and F. Truchetet. Region-based segmentation of color images: application to aerial image cartography. In *Thirty-First Asilomar Conference on Signals, Systems and Computers*, pages 1735–1739, 1997.
- [28] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, London, 1982.
- [29] Prostate Cancer Diagnosis. Retrieved June 10, 2006, from [http : //www.urologychannel.com/prostatecancer/diagnosis.shtml](http://www.urologychannel.com/prostatecancer/diagnosis.shtml).
- [30] H. Digabel and C. Lantuejoul. Iterative algorithms. In *Proc. of the 2nd European Symp. on Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, Caen, France, 1977.
- [31] A. Djouadi and E. Bouktache. A fast algorithm for the nearest-neighbor classifier. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(3):277–282, 1997.
- [32] S. R. Dubios and F. H. Glanz. An autoregressive model approach to two-dimensional shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:55–66, 1986.
- [33] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley, New York, 2001.
- [34] Jonathan I. Epstein. *Prostate biopsy interpretation*. Raven Press, New York, 1995.
- [35] M. Flickner and et al. Query by image and video content: the QBIC system. *IEEE Computer*, 28(9):23–32, 1995.

- [36] K. S. Fu and J. K. Mui. A survey on image segmentation. *Pattern Recognition*, 13:3–16, 1981.
- [37] K. S. Fu and J. K. Mui. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.
- [38] K. Fukunaga. *Introduction to Statistical Pattern Recognition*, second ed. Academic Press, New York, 1990.
- [39] K. Fukunaga and P. R. Hayes. The reduced parzen classifier. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(4):423–425, Apr. 1989.
- [40] K. Fukunaga and D. M. Hummels. Leave-one-out procedures for nonparametric error estimates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(4):421–423, Apr. 1989.
- [41] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov. 1984.
- [42] G.S. Gerber and G.W. Chodak. Routine screening for cancer of the prostate. *J Natl Cancer Inst*, 83:329–335, 1991.
- [43] Y. H. Gong. *Intelligent Image Databases Towards Advanced Image Retrieval*. Kluwer Academic Publishers, 1998.
- [44] R. C. Gonzales and R. C. Woods. *Digital Image Processing*. Addison-Wesley, Reading, MA, 1992.
- [45] R. C. Gonzalez and P. Wintz. *Digital Image Processing*. Addison-Wesley, Reading, Massachusetts, 1987.
- [46] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall International, 2002.
- [47] Tumor Grading. Retrieved June 10, 2006, from [http :
//www.upmccancercenters.com/cancer/prostate/gradingsystems.html](http://www.upmccancercenters.com/cancer/prostate/gradingsystems.html).
- [48] E. L. Hall. *Computer Image Processing and Recognition*. Academic Press, New York, 1979.
- [49] R. M. Haralick and L. G. Shapiro. Image segmentaiton techniques. *Computer Vision, Graphics, and Image processing*, 29(1):100–132, Jan. 1985.

- [50] M. B. Henke-Reed and S. N. C. Cheng. Cloth texture classification using the wavelet transform. *Journal of Imaging Science and Technology*, 37:610–622, Dec. 1993.
- [51] A. Hoover1996 and et al. An experimental comparison of range image segmentation algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(7):673–689, 1996.
- [52] L. Horowitz and T. Pavlidis. Picture segmentation by a tree traversal algorithm. *Journal of the Association for Computing Machinery*, 23(2):368–388, Apr. 1976.
- [53] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification.
- [54] Z. M. Huo and et al. Evaluation of an automated segmentation method based on performances of an automated classification method. *SPIE*, 3981:16–21, 2000.
- [55] B. Jahne. *Digital Image Processing: Concept, Algorithms, and Scientific Applications*. Springer-Verlag, New York, 1991.
- [56] A. K. Jain, R. P. W. Duin, and J. C. Mao. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, Jan. 2000.
- [57] A. K. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(2):153–158, Feb. 1997.
- [58] S. Ji and H. W. Park. Image segmentation of color image based on region coherency. In *Proc. of 1998 Int'l Conf. on Image Processing*, pages 80–83, Chicago, IL, 4–, Oct. 1998.
- [59] A. K. Jian. *Fundamentals of digital image processing*. Prentice Hall International, 1989.
- [60] X. Y. Jiang and H. Bunke. Simple and fast computation of moments. *Pattern Recognition*, 24(8):801–806, 1991.
- [61] T. Jorgensen, K. Yogesan, K.J. Tveter, and et al. Nuclear texture analysis: a new prognostic tool in metastatic prostate cancer. *Cytometry*, 24(277–283), 1996.

- [62] Y. Kanai. Image segmentation using intensity and color information. In *Proc. of the SPIE - The International Society for Optical Engineering Proc. SPIE - Visual Communications and Image Processing '98*, pages 709–720, San Jose, CA, 28-30, Jan. 1998.
- [63] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active countour models. *Int'l Journal of Computer Vision*, 1:321–331, 1988.
- [64] A. Khotanzad and Y. H. Hong. Rotation invariant image recognition using features selected via a systematic method. *Pattern Recognition*, 23(10):1089–1101, 1991.
- [65] W. Kim and R. H. Park. Contour coding based on the decomposition of line segments. *Pattern Recognition Letters*, 11:191–195, 1990.
- [66] J. Kittler and J. Illingworth. Minimum erroe thresholding. *Pattern Recognition*, 19:41–47, 1986.
- [67] T. Kohonen. *Self-Organizing Maps*. Springer Series in Information Sciences vol. 30, Berlin, 1995.
- [68] A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE Trans. Pattern Anal. And Machine Intell.*, 15:1186–1191, Nov. 1993.
- [69] R. Lewis. *Practical Digital Image Processing*. Ellis Honvood, New York, 1990.
- [70] K. C. Liang, X. Wan, and C. C. J. Kuo. Indexing, retrieval and browsing of wavelet compressed imagery data. In *SPIE 9 annual Symposium ET97: Multimedia Processing and Applications*, Feb. 1997.
- [71] W. N. Lie and Y. C. Chen. Shape representation and matching using polar signature. In *International Computer Symposium*, Tainan, Taiwan, 1986.
- [72] C. S. Lin and C. L. Hwang. New forms of shape invariants from elliptic fourier descriptors. *Pattern Recognition*, 20(5):535–545, 1987.
- [73] J. Liu and Y.H. Yang. Multiresolution color image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(7):689–700, 1994.
- [74] X. W. Liu and D. L. Wang. Texture classification using spectral histograms. *IEEE Transactions on Image Processing*, 12(6):661–670, June 2003.

- [75] X. W. Liu and D. L. Wang. Image and texture segmentation using local spectral histograms. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, VOL. , NO. , 15(10):3066–3078, Oct. 2006.
- [76] W. Y. Ma, Y. N. Deng, and B. S. Manjunath. Tools for textures/colour based of images. In *SPIE Int. Conf. 3106, Human Vision and Electronic Imaging II*, Feb.1997.
- [77] A. Madabhushi and D. N. Metaxas. Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. *IEEE Trans. Med. Imag.*, 22(2), Feb. 2003.
- [78] B. S. Manjunath and et al. Color and texture descriptors. *IEEE Tr. CSVT*, pages 703–715, June 2001.
- [79] D. M. Marr and E. Hildreth. Theory of edge detection. In *Proc. R. Soc. Lond.*, pages 187–217, 1980.
- [80] A. D. Marshall and R. R. Martin. *Computer vision, models and inspection*. World Scientific, Singapore, 1992.
- [81] J. McQueen. Some methods for classification and analysis of multivariate observations. In *Proc. of the 5th Berkeley Symp. on Math. Stat. and Prob.*, pages 281–296, 1967.
- [82] A. Mojsilovic. A computational model for color naming and describing color composition of images. *IEEE Transactions on Image Processing*, 14(5):690–699, May 2005.
- [83] J.F. Myrtle. Normal levels of prostate-specific antigen (psa). *Clinical aspects of prostate cancer: assessment of new diagnostic and management procedures*, 183-9, 1989.
- [84] Y. Nakagawa and A. Rosenfeld. Some experiments on variable thresholding. *Pattern Recognition*, 11:191–204, 1979.
- [85] NIH Publication No. 02-4806. Retrieved June 9, 2006, from <http://www.niddk.nih.gov/health/urolog/pubs/prospro/prospro.htm>.
- [86] K. L. Oehler and R. M. Gray. Combining image compression and classification using vector quantization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(5):461–473, 1995.

- [87] Neurologic Complications of Prostate Cancer. Retrieved June 9, 2006, from [http : //www.aafp.org/afp/20020501/1834.html](http://www.aafp.org/afp/20020501/1834.html).
- [88] Anatomy of the Prostate. Retrieved June 9, 2006, from [http : //training.seer.cancer.gov/ss_module02_prostate/unit02_section01_anatomy.html](http://training.seer.cancer.gov/ss_module02_prostate/unit02_section01_anatomy.html).
- [89] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.
- [90] T. Pavlidis. Algorithms for shape analysis of contour and waveforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4), 1980.
- [91] F. Pernus, A. Leonardis, and S. Kovacic. Non-information preserving shape features at multiple resolution. In *11th IAPR International Conference on Pattern Recognition*, 1992.
- [92] L. Priebe. *Lectures on Segmentation*. Institute of Computational Visualistics University Koblenz-Landau, Koblenz, Germany.
- [93] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, Calif., 1993.
- [94] R. Kasturi R. Jain and B.G. Schunck. *Machine Vision*. McGraw-Hill, Inc., New York, 1995.
- [95] S. Raudys. Evolution and generalization of a single neuron: I. single-layer perceptron as seven statistical classifiers. *Neural Networks*, 11(2):283–296, 1998.
- [96] C. Rosenberger and et al. Genetic fusion: application to multi-components image segmentation. In *Proc. ICASSP*, pages 2223–2226, 2000.
- [97] A. Rosenfeld and A. C. Kak. *Digital picture processing*. Academic Press, New York, 1982.
- [98] A. J. Round, A. W. G. Duller, and P. J. Fish. Colour segmentation for lesion classification. In *Proc. of the 19th Annual Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, pages 582–585, Chicago, IL, 30 Oct.-2 Nov. 1997.
- [99] Y. Rui, T. S. Huang, and S. F. Chang. Image retrieval: current techniques, promising directions and open issues. *J. Vis. Comm. Im. Repr.*, 10:39–62, Mar. 1999.

- [100] K. Saarenen. Color image segmentation by a watershed algorithm and region adjacency graph processing. In *Proc. of 1994 Int'l Conf. on Image Processing*, pages 1021–1025, Austin, TX, 13–16, Nov. 1994.
- [101] B. Scholkopf. *Support Vector Learning*. Ph. D. thesis, Berlin Technical University, 1997.
- [102] B. Scholkopf, K. K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik. Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE Trans. Signal Processing*, 45(11):2758–2765, 1997.
- [103] I. K. Sethi and G. P. R. Sarvarayudu. Hierarchical classifier design using mutual information. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1:194–201, Apr. 1979.
- [104] L. Shafarenko, M. Petrou, and J. Kittler. Automatic watershed segmentation of randomly textured color images. *IEEE Trans. on Image Processing*, IP-6(11):1530–1544, Nov. 1997.
- [105] J. R. Smith. Integrated spatial and feature image system: Retrieval, analysis and compression. *Thesis for Ph.D. Columbia University*, 1997.
- [106] I. E. Sobel. *Camera Models and Machine Perception*. Stanford Univ. Press, Stanford, CA, 1970.
- [107] M. S. Stachowicz and D. Lemke. Image segmentation and classification using color features. In *Proc. IEEE VIPromCom*, pages 57–64, 2002.
- [108] T. Taxt, P. J. Flynn, and A. K. Jain. Segmentation of document images. *IEEE Trans. PAMI*, 11(12):1322–1329, 1989.
- [109] Introduction to Prostate Cancer. Retrieved June 9, 2006, from [http : //training.seer.cancer.gov/ss_module02_prostate/unit01_sec01_intro.html](http://training.seer.cancer.gov/ss_module02_prostate/unit01_sec01_intro.html).
- [110] A. Tremeau and N. Borel. A region growing and merging algorithm to color segmentation. *Pattern Recognition*, 30(7):1191–1204, July 1997.
- [111] Prostate under a microscope. Retrieved June 10, 2006, from [http : //cgap - mf.nih.gov/prostateexample/prostatemaandresults/histology/pr2.html](http://cgap-mf.nih.gov/prostateexample/prostatemaandresults/histology/pr2.html).
- [112] V. N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, New York, 1998.

- [113] Y. G. Wang, J. Yang, and Y. C. Chang. Color-texture image segmentation by integrating directional operators into jseg method. *Pattern Recognition Letters*, 27(16):1983–1990, 2006.
- [114] J. S. Weszka and A. Rosenfeld. Threshold evaluation techniques. *IEEE Trans. Syst. Man Cybern*, 8:622–629, 1978.
- [115] S. D. Yanowitz and A. M. Bruckstein. A new method for image segmentation. *Computer Vision Graphics Image Process*, 46:82–95, 1989.
- [116] Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.
- [117] Y. J. Zhang and et al. Objective and quantitative segmentation evaluation and comparison. *Signal Processing*, 39(3):43–54, 1994.
- [118] Y.J. Zhang. *Image Segmentation*. Science Publisher, 2001.