العنوان:	التشخيص الآلي عبر تقسيم غدد أنسجة البروستات
المؤلف الرئيسي:	الحاج صالح، صفاء نعيم
مؤلفين آخرين:	القاضي، عمر سلطان، الزعبى، محمد بلال(مشرف)
التاريخ الميلادي:	2013
موقع:	عمان
الصفحات:	1 - 147
رقم MD:	901717
نوع المحتوى:	رسائل جامعية
اللغة:	English
الدرجة العلمية:	رسالة ماجستير
الجامعة:	الجامعة الاردنية
الكلية:	كلية الدراسات العليا
الدولة:	الاردن
قواعد المعلومات:	Dissertations
مواضيع:	الغدد، أنسجة البروستات، التشريح الاوتوماتيكى
رابط:	http://search.mandumah.com/Record/901717

HISTOPATHOLOGICAL PROSTATE TISSUE GLAND SEGMENTATION FOR AUTOMATED DIAGNOSIS

By
Safa'a Al-Haj Saleh

Supervisor
Dr. Mohammad Belal Al Zoubi, prof.

Co-supervisor
Dr. Omar Sultan Al Kadi

This thesis was submitted in partial fulfillment of the requirements for the Master's Degree in Computer Information Systems

Faculty of Graduate Studies
The University of Jordan

May, 2013

نموذج ترخيص

أنا الطالبة: ~~مهند نعيم حسين الحاج صالح~~ منح الجامعة الأردنية و/ أو من تفوضه ترخيصا غير حصري دون مقابل بنشر و / أو استعمال و / أو استغلال و / أو ترجمة و / أو تصوير و / أو إعادة إنتاج بأي طريقة كانت سواء ورقية و / أو إلكترونية أو غير ذلك رسالة الماجستير / الدكتوراه المقدمة من قبلي وعنوانها:

~~Histopathological prostate tissue gland segmentation for automated diagnosis~~

وذلك لغايات البحث العلمي و / أو التبادل مع المؤسسات التعليمية والجامعات و / أو لأي جهة أخرى تراها الجامعة الأردنية مناسبة، وتمنح الجامعة الحق بالترخيص لغير بجميع أو بعض ما رخصته أنها.

الاسم الكامل: مهند الحاج صالح

التوقيع:

التاريخ: ٢٠١٣/٥/٢٢

COMMITTEE DECISION

This Thesis/ Dissertation (Histopathological Prostate Tissue Gland Segmentation for Automated Diagnosis) was Successfully Defended and Approved on 9/5/2013.

Examination Committee	**Signature**
Dr. Mohammad Belal Al Zoubi **(Supervisor)** Prof. of Data Mining	
Dr. Omar S. Al Kadi **(Co-supervisor)** Assist. Prof. of Image Processing	
Dr. Bassam H. Hammo **(Member)** Assoc. Prof. of Natural Language Processing	
Dr. Thair M. Hamtini **(Member)** Assoc. Prof. of E-learning	
Dr. Amer F. Al-Badarneh **(Member)** Assoc. Prof. of Data Mining (Jordan University of Science and Technology)	

تعتمد كلية الدراسات العليا
هذه النسخة من الرسالة
التوقيع.....التاريخ.....

DEDICATION

I dedicate this thesis to my family and my friends who supported me during each step of the way. Special dedication for my mom that gave me faith that any impossible could simply turn into possible. Thank you all for making this dream true.

ACKNOWLEDGEMENT

I would like to express my gratitude to my main supervisor Prof. Mohammad Belal Al Zoubi and my co-supervisor Dr. Omar Al Kadi for their patience, motivation, enthusiasm, and their great continuous support. I would like also to thank Dr. Fatima Obeidat and Dr. Khalid Wissam from the University of Jordan Hospital and Mr. Ali Dumour from Prince Hamza Hospital for their great help in providing, digitizing, and diagnosing the dataset of prostate tissue images.

Table of Contents

Subject **Page**

List of Figures

Figure Name **Page**

List of Tables

Table Name **Page**

List of Abbreviations

PCa	Prostate Cancer
PSA	Prostate Specific Antigen
MRFs	Markov Random Fields
FCM	Fuzzy C-Means
PCM	Possibilistic C-Means
ACO	Ant Colony Optimization
DCT	Discrete Cosine Transform
FDK	Fisher Discriminant K-means
SVM	Support Vector Machine
AdACM	Adaptive Active Contour Scheme
GAC	Geodesic Active Contour
HNCut	Hierarchical frequency Weighted Mean Shift Normalized Cut
PPMMs	Probabilistic Pairwise Markov Models
RAC	Region-based Active Contour

HISTOPATHOLOGICAL PROSTATE TISSUE GLAND SEGMENTATION FOR AUTOMATED DIAGNOSIS

By
Safa'a Al Haj Saleh

Supervisor
Dr. Mohammad Belal Al Zoubi, prof.

Co-supervisor
Dr. Omar Sultan Al Kadi

ABSTRACT

In this work, we proposed a methodology for segmenting glands automatically in digitized images of histopathological prostate tissue in order to classify those tissues. Developing an automated technique for segmenting glands in prostate tissue leads to accurate classification of those tissues into grades. Gleason grading describes the abnormality of cancer cells and their degree of aggressiveness by using numerical scale from grade 1 that represents benign tissues through grade 5 that represents advanced cancerous tissues. The special characteristics of glands in prostate tissue for each grade play significant role in discriminating Gleason grades. When applied in medical field, this work will help pathologists in the initial grading for prostate tissue images.

To extract required features for grading the tissues, we focused on segmenting lumen objects and glands. Segmenting lumen objects was performed depending on thresholding technique. We were mainly concerned with segmenting inner regions of the glands (that consist of lumen, cytoplasm and inner boundary of nuclei) depending on k-means clustering applied on a* channel of tissue image.

By the end of segmentation process, final segmented lumen objects and glands were labeled to extract statistical and morphological features. Nine features were extracted for segmented lumen objects and glands. Lumen objects features include: average area, maximum area, and average eccentricity. Glands features include: average area, maximum area, average diameter, average perimeter, average eccentricity, and glands

density. Finally, Naive Bayes classifier was used to classify tissue images to get the grade they belong to.

We evaluated the efficiency of the automated segmentation and classification by calculating accuracy, specificity and sensitivity of classification. The classification resulted in accuracy of 91.66%, sensitivity of 96.66 % and specificity of 95.00%. Those results indicate that our automated methodology will have a positive impact when added to the histopathological diagnosis procedure.

CHAPTER I

Introduction

In this chapter we present general overview about prostate cancer. We discuss the traditional diagnosis of prostate cancer, and why were concerned mainly with prostate biopsy test. We present Gleason grading system of prostate cancer, and the importance of automated glands segmentation and classification of prostate tissues based on this grading system.

1.1 Overview of Prostate Cancer

Prostate cancer (PCa) is a major medical problem that has economic effects. It is considered to be the fifth most common cancer among men in the world. It is recognized as one of the important problems facing the males. Around 900,000 new cases were diagnosed in 2008 over the world with the percentage of 14% of cancer cases. In 2007, there were 218,890 new cases of PCa in the Unites States, resulting in 27,050 deaths (around 12.35% of the cases) (Hammerich et al., 2009). In Europe the incident rate of PCa is outnumbering lung and colorectal cancer (Boyle and Ferlay, 2005). In Jordan, 133 new cases of PCa were discovered in 2006 with the percentage of 6.5% of cancer cases (Taraweneh and Nimri, 2006).

1.2 Traditional Diagnosis of Prostate Cancer

Early detection of PCa plays a vital role in the treatment of this disease.To find out the specific prostate problem, doctors ask patients about the symptoms they have. Some tests are needed to determine whether the patient has PCa or not, and to classify the cancer, if found, into grades. The American Cancer Society advises men over the age of

50 to discuss the advantages and the disadvantages of testing with their doctors, so that they can decide if it is the right choice for them. Researchers are working on developing effective methods for diagnosing prostate cancer. The process of diagnosing PCa consists of several tests including digital rectal examination, prostate specific antigen (PSA) test, transrectal ultrasound, gray scale ultrasound, elastography, and transrectal ultrasound-guided prostate biopsy. In this work, we were concerned with prostate biopsy test since it results in digitized prostate tissue image that could be classified using automated approaches.

1.3 Digitized Prostate Biopsy Slides

As mentioned before, doctors use many tests to diagnose PCa cases. Prostate biopsy is one of the important tests used to diagnosis cancer. Prostate biopsy is a sample that is obtained by inserting a thin needle through the rectum or the urethra to collect number of tissue samples from the prostate gland. The biopsy is recommended if results of initial tests show that patient may have PCa. After the prostate biopsy is taken, slides of the tissue are prepared and stained using staining methods like Hematoxylin and Eosin (H&E) staining method. The biopsy is treated and sliced on lab slides. Pathologists use microscope to examine and analyze the biopsy slide. Since digital histopathology emerged recently, the slides are scanned using high resolution scanners to digitize them into images that represent the prostate epithelial tissue.

The resulting digitized images can be viewed with low resolution (10X) or high resolution (40X) for better analysis. The digitized tissue images can be examined instead of the physical slides. By examining those images, pathologists diagnose many prostate problems including cancer. This is based on examining gland units in prostate tissue to check some features such as size, shape, and number of glands in the tissue image.

1.4 Main Components of Glands in Normal Prostate Tissue

After the prostate biopsy is taken, slides of the tissue are prepared and stained using one of the staining methods like H&E staining. H&E is the most widely used staining procedure in medical diagnosis**.** Stained slides are digitized into images that represent the prostate epithelial tissue. Prostate tissue consists mainly of gland units. Prostate gland unit consists of five main components: stroma, lumen, epithelial nuclei, epithelial cytoplasm, and blue mucin. Stroma is the supportive tissue surrounding the gland structure. Lumen is the inside space of gland structure. Epithelial nuclei are large, membrane-bound, and spherical structure within gland structure. Epithelia cytoplasm is the jelly-like substance of the gland between the nuclei and lumen. Mucin is the fine wisps of basophilic material in the gland lumen. It is commonly found in the glands of prostate carcinoma. Those components appear in different colors in a prostate tissue image stained by the H&E method: the stroma is the pink region, the lumen is the white region, the nuclei are the dark blue dots, the cytoplasm is the purple region, and the mucin is the light blue object.

Gland components are arranged as follows: lumen region is surrounded by epithelial cytoplasm. The boundary of the gland (that adjoins the cytoplasm) is formed by layers of epithelial cells that have the epithelial nuclei (Nguyen et al., 2010). The region surrounding the gland is the stroma. In some cancer tissues, blue mucin may be found invading the lumina (Nguyen et al., 2011). It is noticed that both nuclei and cytoplasm gather densely around the gland but scatter sparsely in other areas like stroma (Nguyen et al., 2011).

In normal (also called benign) prostate tissue, stroma is considered the background and the gland units are considered the foreground objects. The gland unit does not have

fixed size or shape. It can be oval, round or branchy (Nguyen et al., 2011). It can also be either small or very large. The gland units are densely-packed and well separated. Each gland has large lumen regions and thick boundary. The nuclei are prominent and form a chain on the gland boundary. The lumen varies in shape; it may be circular, oval, or branchy. Characteristics of glands in other grades are covered in the next section. Figure 1.1 shows sample of benign tissue pattern.

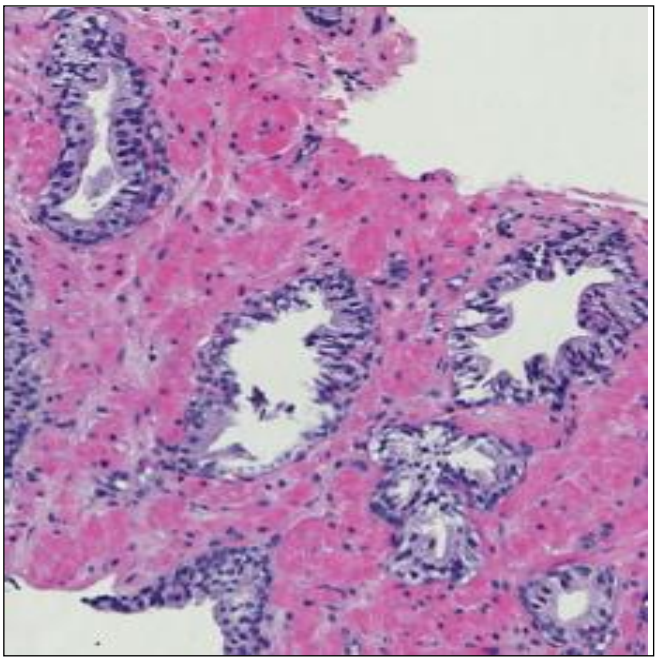

Figure 1.1: Sample of benign tissue pattern (Nguyen et al., 2011)

1.5 Gleason Grading System of Prostate Cancer

When prostate carcinoma is found in a biopsy, it is be graded to estimate how quickly it will grow and spread. Grading describes the appearance of cancer tissue slices observed under microscope. It is also used to describe the abnormality of cancer cells and their degree of aggressiveness. Over 40 histological grading systems for prostate carcinoma have been proposed (Humphrey, 2004). The most widely used grading system is the Gleason grading system, while other grading methods are applied in some laboratories.

The Gleason grading system is the most commonly employed grading system (Tanagho and McAninch, 2008). It is a grading system that relies on the glandular architecture of prostate tissues; describing the degree of loss of the normal glandular tissue architecture (i.e. shape, size and differentiation of the glands). It is a vital determinant of patient's risk of dying due to PCa (O'Dowd et al., 2001). Tumor grading based on Gleason grading helps in deciding the suitable type of treatment for patient. The most prevalent patterns are assigned grades from 1 to 5. The grade of each carcinoma pattern is based on its differentiation (how much of its structure resembles a normal pattern structure). Grade 1 is the most differentiated and grade 5 is the least differentiated. The Gleason score obtains values from 2 to 10; since it is the sum of the two grades. In this work, we care about the grades from 1 to 5. Samples of all grades are illustrated in next chapter.

Gleason grades 1 and 2 patterns are considered the least important grades since they occur in the general population representing healthy prostate. Grades 1 and 2 are called benign patterns (Nguyen et al., 2010). In benign tissues, glands are large, appearing as single separated units, and having large branchy lumen components and thick gland boundaries (since multi layers of epithelial cells are bounding the lumen) with prominent nuclei. There are a lot of variations in lumen shape, from circular to oval or branchy (Nguyen et al., 2011). Nuclei form thick chain on the gland boundary. The glands are also uniformly shaped, and closely packed (Tanagho and McAninch, 2008).

Gleason grade 3 is the most common grade and common case of carcinoma. It is considered to be well differentiated like benign patterns. It is characterized by the invasion of small glands into the stroma (Nguyen et al., 2011). In grade 3, pattern glands are smaller and more circular than in a benign pattern. Blue mucin is commonly

found to mix with lumen. This pattern also has small, circular lumen and thin nuclei boundaries (since only one layer of epithelial cells are bounding the lumen).

In Gleason grade 4**,** the glands start to lose their architecture; we can notice the loss of normal gland unit. The common observation in grade 4 is having incomplete gland formation (Tanagho and McAninch, 2008). It is obvious that this grade does not have well-separated gland unit with separate lumen, and well defined epithelial cell layers on the boundary. Nuclei distribute uniformly instead of forming well-defined boundaries as in benign patterns. Glands are fused with each other and they are poorly defined (Nguyen et al., 2011). Due to the fusion of glands, glandular structures are altered obviously. Multiple glands are mixed together to create a mass of glands containing multiple lumen components.

Gleason Grade 5 is considered to be less common than grade 4. It is rarely found in men whose prostate carcinoma is diagnosed early. This grade shows variety of patterns with no formation of gland units or lumen appearance (Tanagho and McAninch, 2008). This grade is often called undifferentiated; since its features are not significantly distinguishing. For previously mentioned reason, grade 5 is not considered in this research.

The special characteristics of glands in each grade motivated many researchers to extract features for glands and the main components of the glands for automated diagnosis. Classifying prostate tissue images into grades based on extracted features from automatically segmented tissue glands is a challenging machine learning task that consists of three main stages: Segmentation, feature extraction, and classification.

1.6 The Importance of Automated Glands Segmentation and Classification of Prostate Tissues

Developing an automated technique for segmenting glands in prostate tissue leads to accurate classification of those tissues. This automated segmentation and classification leads to faster and accurate diagnosis results. This is because histopathological examination relies on the visual interpretation of a pathologist which may lead to subjectivity in the diagnosis of prostate carcinoma. Automated methods provide objective measures instead of depending on subjective measures. Those methods extract set of mathematical features including statistical and morphological features. Extracting those features is important for classifying the prostate tissue into the suitable grade. In our study, we were concerned with automated segmentation and classification of prostate tissue image into grades based on Gleason grading system. This grading helps doctors in diagnosing cases for choosing the suitable medication for the patient.

1.7 Proposed Glands Segmentation and Classification Methodology

Many approaches were proposed for the segmentation of glands in prostate tissue image. Those approaches mainly depend on algorithms like: Bayesian classifier, active contours, k-means clustering, Markov Random Fields (MRFs), region growing, fuzzy logic rules, watershed algorithm, and Fuzzy C-Means algorithm (FCM).

In this work, we were concerned with segmenting lumen objects and glands in prostate tissue image to extract features for classifying tissue images. Segmenting lumen objects was performed depending on thresholding technique mainly. We focused on segmenting the inner region of the gland (which consists of lumen, cytoplasm and inner boundary of nuclei). Glands segmentation was performed depending on k-means clustering applied on a* channel and other morphological operations. By the end of segmentation process,

final segmented lumen objects and glands were labeled to extract some features. Feature extraction step was important to classify dataset tissue images into the suitable grade using suitable classifier (Naive Bayes classifier). Nine features were extracted for segmented lumen objects and glands. Those features were classified into statistical features and morphological features. Statistical features were related to area, diameter, and perimeter. Morphological feature was related to eccentricity.

While segmenting glands, we were interested in segmenting the inner region of the gland to have good separation of glands since glands in grade 4 tissue tend to fuse. This improved the accuracy of segmenting glands and helped in segmenting glands separately instead of segmenting combined glands especially when glands are close together or when having gland mass that usually occurs in grade 4 tissue images. This will be discussed in later chapters.

The proposed methodology was based on clustering not on model based approaches such as Active Shape model or level sets. This was due to the heterogeneity in prostate tissue image, i.e. the diversity of shape and position of the main parts of a gland (nuclei, cytoplasm and lumen).

1.8 Thesis Organization

This thesis is organized as follows: Chapter 2 views general medical background (Chapter 2 can be skipped, but you can refer to the sections that describe main components of prostate tissue and Gleason grading system for more illustration using samples of images). Chapter 3 focuses on literature review. Chapter 4 discusses the proposed methodology for segmentation and classification. Chapter 5 deals with our experimental results. The thesis is ended with chapter 6 that views conclusion and future work.