Counting Sertoli Cells in Thin Testicular Tissue

Abdolrahim Kadkhodamohammadi
Abstract

Counting Sertoli Cells in Thin Testicular Tissue

Abdolrahim Kadkhodamohammadi

This master thesis develops a novel system to model the tubular structure in thin sections of testicular tissue and count the Sertoli cells. A three-phase method is proposed to model the tubular structure in microscopic images of the tissue, the model is deployed to detect the cells. In the first phase, the germ-mass, which represents the inside layer of tubules, are detected. All cells are detected by radial symmetry transform and then the graph cut algorithm is used to separate the germ cells. Each region covered by a compact set of germ cells is considered as the germ-mass. In the second phase, all bright areas in the image are detected and used to adjust the germ-mass regions. In the last phase, all edges that are line-like are identified and straight lines are fitted to the edges. The lines are later connected to compensate for the broken parts of the tubules' boundaries. The closest cells to the germ-mass are chosen as the Sertoli cell candidates. The approximate boundary of tubules and the angle between the candidate cells are used to detect the Sertoli cells. Our experimental results show that our system is able to detect the tubule and the Sertoli cells with reasonable accuracy. If the method can not find enough edges to approximate the tubule's boundary, detecting Sertoli cells is complicated; the system can report those situations to the experts. Since we use the symmetry attribute of the cells to detect them, the method is quite robust against noise, artifacts, and non-uniform illumination. The method is able to capture all tubules, even tubules that do not have any bright region in the middle (lumen). To the best of my knowledge, no one has proposed a method to model tubular structure without lumen. The border approximation method can work well even for tubules that are partially in the image. It should be mentioned that the proposed method could be applied to model any tubular structure with one or more cells types.

Handledare: Azadeh Fakhrzadeh
Ämnesgranskare: Cris Luengo
Examinator: Anders Berglund
IT 12 017
Tryckt av: Reprocentralen ITC
Acknowledgements

First of all I would to express my deep gratitude to my supervisor Azadeh Fakharzadeh and my reviewer Cris L. Luengo Hendriks at CBA (Centre for Image Analysis, Uppsala, Sweden). Without their continuous valuable support, guidance and encouragement it was impossible to perform this study. Also I would like to thank all the staff in the CBA. You are very kind.

I owe my sincere gratitude and appreciation to my wife Serveh, who has always been supporting me with love. She went through a lot of lonely moments so that I could finish my studies. I would also like to express my sincere appreciation to my loving parents for their dedication and the many years of support during my whole life.
Contents

List of Figures 7

List of algorithms 9

1 Introduction 11

1.1 Motivation and problem description 11

2 Review Related work 17

2.1 Cell detection 17
2.2 Model the tubular structure 19

3 Image Acquisition and Enhancement 21

3.1 Staining process 21
3.2 Image Enhancement 21
3.2.1 Bilateral filter 22
3.2.2 Experimental result 23

4 Modelling Tubules 25

4.1 Construct the germ-mass 25
4.1.1 Cell detection 26
4.1.2 Clustering cells 28
4.1.3 Connect germ cells 33
4.1.4 Experimental result 35
4.2 Detect Lumen 39
4.3 Detect the tubules’ borders 43

5 Detect the Sertoli Cells 47

5.1 Assign non-germ cells to germ-mass 47
5.2 Angle to prune the Sertoli cells 47
5.3 Boundary map to prune the Sertoli cells 49

6 Result and Discussion 51

6.1 Conclusion 55
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Different type cells in testicular tissue.</td>
<td>13</td>
</tr>
<tr>
<td>1.2</td>
<td>Tubules with different type of lumen, a) tubule without lumen, b) tubule with lumen, c) tubule with irregular lumen shape, d) interstitial connective tissue.</td>
<td>14</td>
</tr>
<tr>
<td>1.3</td>
<td>Block-digram of the proposed method.</td>
<td>15</td>
</tr>
<tr>
<td>3.1</td>
<td>(a) the original image, (b) the image after applying bilateral filter.</td>
<td>23</td>
</tr>
<tr>
<td>4.1</td>
<td>The positions of pixels $p_{+ve}(p)$ and $p_{-ve}(p)$ affected by the gradient element $g(p)$ for a range of $n = 2$. The dots represent all the pixels that can be affected by the gradient at $p$ for a radius $n$.</td>
<td>27</td>
</tr>
<tr>
<td>4.2</td>
<td>(a) simple graph, (b) directed graph, arrow shows direction.</td>
<td>29</td>
</tr>
<tr>
<td>4.3</td>
<td>Some regions in three-dimension feature space.</td>
<td>31</td>
</tr>
<tr>
<td>4.4</td>
<td>The sample image used to test the proposed method.</td>
<td>35</td>
</tr>
<tr>
<td>4.5</td>
<td>The saturation channel in HSV color space.</td>
<td>36</td>
</tr>
<tr>
<td>4.6</td>
<td>The fast radial symmetry transform of the sample image.</td>
<td>37</td>
</tr>
<tr>
<td>4.7</td>
<td>The binary image of FRST of the sample image.</td>
<td>37</td>
</tr>
<tr>
<td>4.8</td>
<td>The cell regions after applying morphological operations.</td>
<td>38</td>
</tr>
<tr>
<td>4.9</td>
<td>The cells regions are shown by green color over the original image.</td>
<td>39</td>
</tr>
<tr>
<td>4.10</td>
<td>The cluster of germ regions.</td>
<td>40</td>
</tr>
<tr>
<td>4.11</td>
<td>The result of processing the sample image by algorithm 4.3, (a) the connected germ cluster, (b) fill all fragments, (c) remove big filled fragments from germ-mass, (d) opening with disk structural element of size 10, (e) fill background regions surrounded by one germ-mass, (f) remove small region and medium bright regions.</td>
<td>41</td>
</tr>
</tbody>
</table>
4.12 Processing the bright region in l-channel along with the obtained germ-masses to apply lumen information, (a) l-channel in Lab color space, (b) all bright pixels, (c) the big regions in (b), (d) the regions of (c) after applying level set, (e) the regions that satisfy the size and convexity condition, (f) red regions represent germ-mass and the yellow region is the only region in (e) that fulfill both [4.6] and [4.7].

4.13 Detecting the boundary of tubules, (a) the result of the Canny edge detector, (b) the outside germ-mass regions edges after using [4.8], (c) the approximate lines and their connections where the red lines show approximated lines, and white and green lines show the approximation of lines that divided into two parts and the approximation of each part respectively, and the blue lines represent connecting lines, (d) all lines in (c) is shown over the original image.

4.14 A sample line with slope \( s \) is shown and its neighborhood region is represented by gray color.

5.1 The germ-mass and non-germ regions are shown by green color over the original image. The blue lines connect each cell to its nearest germ-mass.

5.2 Triangle that connect three points, the name of each angle and side are written by capital and small later, respectively.

5.3 The cells with an angle with close neighbors not close to 180 degrees are marked with white star.

5.4 The Sertoli cells are connected to germ-mass. Red lines represent connections that cross more than two edges and stars represent the cell regions that cross at least one edge and are not in straight line with their direct neighbors.

6.1 How an image is processed through the proposed method.

6.2 Choose and combine the lumen candidates with the germ-masses, (a) the formed germ-mass by connecting the germ cells, (b) lumen candidate, (c) the join of the germ-masses and detected lumens.

6.3 Some sample image that suffer from distortion, (a) the original image, (b) output result of the system.

6.4 Border estimation of the tubules, (a) estimation of edge responses with straight line represented with red and green lines and connections that made by our method represented by blue lines, (b) all lines over the original image.
# List of Algorithms

1.1 Build neighborhood graph ........................................... 30
1.2 Define costs to apply graph cuts algorithm .................... 32
1.3 Connect regions that are clustered as germ cell ............... 34
Chapter 1

Introduction

1.1 Motivation and problem description

There is a growing concern about the effect of different substances in our environment on the male reproductive organs of both humans and animals. Using image analysis we can achieve a fast valid method for analyzing testicular histological images and study this effect. Evidence is accumulating that the reduced reproductive function in adult males is related to an increased oestrogen exposure during embryonic development [1]. Oestrogenic substances can be found all around us. Some are industrial chemicals and pesticides with oestrogenic properties and some occur naturally in plants, which are used for both human and animal consumption. In addition, synthetic oestrogens used for medical purposes are excreted in the urine and end up in our water systems [1, 2]. Independent of their origin, substances with estrogenic properties are commonly referred to as xeno-oestrogens.

Oestrogen has traditionally been considered a female hormone but it is now clear that the hormone is also involved in male reproductive function. The exact role of the hormone has not been completely established but it has been detected in testicular tissue and its two main receptors have been localized in several different regions of the male reproductive tract in several species [3, 4, 5, 6]. A number of investigations show that if male animals are exposed to xeno-estrogens during embryonic development or early in life, testicular morphology and function is affected as the animal reaches sexual maturity [7, 8, 9, 10]. To understand the mechanism behind these effects, studies of the microscopic structure of the testis and the process of sperm development are of vital importance. The development of spermatozoa takes place in the epithelium of the seminiferous tubules of the testis and the process is promoted by testosterone secreted from Leydig cells in the interstitial tissue. The epithelium consists of germ cells in various stages of development and the Sertoli cells, which are crucial for normal differentiation of the germ cell and therefore the final output of functional spermatozoa.
The Sertoli cells are highly important for the function of the testes and the development of spermatozoa. They provide support, protection and nutrition to the developing spermatogenic cells, secrete testicular fluid for transport of spermatozoa in the tubular lumen, participate in endocrine regulation, phagocytosis of excess cytoplasm and form the blood-testis barrier, which prevents autoimmune attacks of the spermatogenic cells. Sertoli cells are columnar or pyramidal in shape and adhere to the basal lamina of the seminiferous tubules at the base. Their apical ends extend into the tubular lumen. Each Sertoli cell supports 30-50 developing germ cells and spermatids with tails can be seen protruding from their apical ends. The central and highly important functions of Sertoli cells, and the fact that the number of Sertoli cells limits the amount of spermatogenic cells that can develop into spermatids, make them ideal as marker cells for the study of testicular function. As such, these cells can be a valuable tool in studies of various environmental pollutants known to affect fertility or other effects on testicular tissue after diseases or drug treatment.

Testicular tissue is composed of tightly packed tubules, called seminiferous tubules, separated by connective tissue. Connective tissue contains many large rounded interstitial cells. Immediately surrounding each tubule are flattened myoid cells. Sperm cells are generated in seminiferous tubules at a rate of about $2 \times 10^8$ per day in the adult male. Each tubule is lined with a complex and specialized epithelium called seminiferous epithelium. The seminiferous epithelium consist of two type of cells: supporting or Sertoli cells and germ cells. The function of the germ cells is to produce sperm. Germ cells at all stages of meiosis are found embedded within the epithelium. The Sertoli cells create the environment in which germ cells carry out their function. The number of the Sertoli cells in the tubules is an indirect way of estimating the health of testicular tissue. The Sertoli cells can be distinguished by morphology as well as their location within the tissue.

Figure 1.1 shows a thin section of testicular tissue. Inside each tubule, there is epithelial tissue and a lumen. The lumen is the bright area in the middle of tubule. Generally we see three group of cells. The germ cells are the darker cells that are closer to lumen. The Sertoli cells are the cells that are closer to the border of tubule. Interstitial cells are round cells outside the tubules. The morphology and the texture of the Sertoli and interstitial cells are very similar. A sample of the three types of cells which are indicated by arrows are shown in figure 1.1. A lumen is represented by closed contour. It should be mentioned that in a ideal case when the tissue is cut in a appropriate angle, a clear lumen is seen in the middle of every tubule. However due to the random cut of the tubules, this is not the case here. More sample images can be found in appendix A.

In this project, we are interested in detecting the Sertoli cells which can

\footnote{perpendicular to tubule’s direction}
provide a measure of healthiness. As discussed above, the texture of the Sertoli and interstitial cells are quite similar. The only difference is in their location. The Sertoli cells are inside the tubules and the other one is outside. On the other hand, inside the tubules we have the germ cells as well. But the Sertoli cells are more close to the border of the tubule while the germ cells are surrounding lumen and their texture looks different. Thus we could say that attaching with the inside border of the tubules is the main property that distinguishes the Sertoli cells from others.

A human can detect the Sertoli cells by recognising the tubule’s border and higher order information from the relation between different parts of the tubules. Unfortunately the border of the tubules are ill-defined and they are not detectable via classical line or edge detection operators. Instead we try to use the cells and the lumen to model the tubules. First, we use the texture discrepancy between the germ cells and others to identify them. Then, we try to find epithelial tissue by connecting the germ cells and modelling germ-mass. The germ-mass is defined as the inside regions of each tubule that includes all germ-cells. The germ-mass model is adjusted by lumen and the edge information. Finally, The Sertoli cells can be distinguished, e.g. we can simply define the Sertoli cells as the closest cells surrounding the germ-mass.

Figure 1.2 shows different kinds of tubules. Some of them have lumen and some of them are without lumen. Although finding lumen could be a good starting point to detect germ cells, it is not always reliable. There are some problems with using lumen to detect germ cells:
− There are tubules in which the lumen does not appear, figure 1.2 a.

− We have some bright areas outside the tubules which are very similar to the lumen called interstitial connective tissue, figure 1.2 d.

− Lumens could have any shape (elongated ellipse like thick bended line or roundish region like a circle), figure 1.2 c and b.

− Lumens can be scattered and only appear in some part of the tubules, see figure 1.2 c.

The main building box of the proposed method is given in figure 1.3.

In our method, we try to model the tubules after enhancing images and then detect the Sertoli cells. Image enhancement is described in chapter 3. The modelling tubules and detect the Sertoli cells phases are presented in chapter 4 and 5, respectively.
Figure 1.3: Block-digram of the proposed method
Chapter 2

Review Related work

Applying image processing and computer vision has gained a great popularity in medical applications. The image processing based system can process huge amount of data in different modality\(^1\) and provide a set of re-computable measures for physicians and relieve them from going through whole data. There has been some research on devising automated methods to analyze cells and tubular structure in tissues. Section 2.1 discusses some proposed method for cell detection. A review on methods that model tubular structure is reported in section 2.2.

2.1 Cell detection

The first attempts to detect and segment cells in microscopic images were based on gray value thresholding. Wu et al. [12] segmented cell images by thresholding a parametric image, approximating the original image. They convert the segmentation problem into an optimization problem that try to minimize the error between the original and the constructed image. Petushi et al. [13] applied adaptive thresholding to segment cells and classify them into three classes. Thresholding based methods are sensitive to variation in staining and gray value and they also have problems in dealing with non-uniform illumination.

Clustering algorithm in pixel level is another approach frequently reported in the literature. Fatakdawala et al. [14], used expectation-maximization to cluster pixel into four different classes. A color template was used to choose one cluster as the cancer nuclei and then each separated region in this cluster was applied to initialize the level-set. Since the level-set may connect close regions, the convexity and size of the region used to split them. Finally, some texture features were evaluated to accept or reject the obtained regions. Hafiane et al. [15], applied fuzzy c-mean clustering to

\(^1\)CT, MRI, ECG, EEG, etc.
detect potential nuclei location for initializing level-set and then iterative voting algorithm is used to cluster segmented nuclei.

Veta et al. [16] applied color deconvolution to separate hematoxylin channels, and morphological operations to enhance the images and remove artefacts. The nuclei locations were determined by the orientation of the fast radial symmetry transform. Detected areas were used as a marker for the watershed segmentation method. Then watershed regions were filtered based on their sizes. The morphological watershed transform and Fourier descriptor were applied to segment and detect tumoral cells in breast cancer [17].

A two-layer scheme is reported to count lymphocytes in histopathological images in [18]. In first step, They used morphological operations and thresholding to enhance the image quality and segment cells. In the next step, the segmented regions were filtered using template matching. Duan et al. [19] proposed a method that deploy the morphological operation to search for initial seed points for the level-set. Plissiti, Nikou, and Charchanti [20] used the morphological operation and information from circumference of each nucleus to extract some measures. The performance of the measures are evaluated with both supervised and unsupervised classification techniques in detecting cell nuclei.

Almost all the aforementioned methods deal with one type of cells. They just had one class of nuclei and tried to detect them [15, 12, 20, 18, 17, 16]. Fatakdvala et. al. detected four classes of regions that are the Stroma, lymphocyte nuclei, breast cancer nuclei, and background; by pixel clustering based on color value [14]. Then they have just selected lymphocyte class and extract texture features to detect right ones. Although two kinds of nuclei were available in their images, it seems that the lymphocyte nuclei class is quite distinguishable in color space which made them able to extract the nuclei by pixel clustering.

As mentioned in section 1.1, there are three types of cells in our images and the rest can be considered as background. The sertoli cells and interstitial cells are similar and it is not possible to distinguish them by texture or shape. One main property of all the cells is that the gray level in those regions are symmetric. Therefore, it is better to deploy the property for segmenting the cells. Even though the grey level operation, e.g. gray value thresholding, pixel clustering and morphological operations that is deployed in almost all of the mentioned methods, might somehow reflect the property, directly using the feature of the cells’ region could yield better result.

The symmetric property of the cells can be used to devise a segmentation method. The method could detect regions with high symmetry value as the cells. I believe that using the property of region of interest could have two big impacts on segmentation process. First, it will be less sensitive to noise, artifacts and non-uniform illumination. Second, it is more directed towards our interest points and leads to less false-segmented regions.

Although in [16], radial symmetry is applied to detect nuclei, due to the
low contrast and flat intensity level of nuclei, only orientation information was deployed. The local maximums of radial symmetry transform is later used as foreground and background markers for the watershed. In this study they just have one type of nucleus and the rest is considered as background which are in the same size as nuclei. The size of the nuclei in their images are at least two times bigger than ours. In addition background, which are the lumen and interstitial connective tissue in our case, could be very small or large regions that cause to miss the markers or detect several markers, respectively.

2.2 Model the tubular structure

Apart from the above mentioned method, some methods have been proposed for modelling tubules and inter-cell relations. Taking into account that usually a tubule is composed of a large bright area in the middle as its lumen which is surrounded by layers of cells. Therefore one way of finding the tubules is to segment all white areas and search for those which are in the middle of cells.

Basavanhally et al. [21] clustered pixels via hierarchical normalized cut to detect all bright areas. They detected all nuclei using color deconvolution and the morphological operations and assigned them to the closest bright areas. Finally, they applied domain knowledge and O’Callaghan neighborhoods and extracted 22 features to accept the white areas which are lumen of tubules. Monaco et al. [22] also find all white areas. Then they extracted some morphological features for each region and classified as either malignant or benign via Bayesian classifier. Finally the result served as initial starting point to improve the classification result using Markov random fields.

Niak et al. [23] introduced a three-step approach to segment tubular structure for scoring prostate and breast cancer. In the first step, a Bayesian classifier was trained to assign cytoplasm, lumen, and nuclei likelihood to each pixel. In the second step, level-set and template matching were applied to delineate the border of regions and cluster them. In this step to remove false bright regions that are detected as lumen, they define some constraints based on the size and shape of the white area and the density of nuclei surrounding them.

Nguyen et. al. proposed a method to grade prostate malignancy [24]. They clustered all pixels to detect nuclei and the lumen. All image pixels were clustered and the lumens as well as nuclei were detected. Then tubular structure was defined by considering the structural relations between the lumens and nuclei. Finally they extracted some shape and texture feature for the tubular structure and evaluated the discriminant power of the features using several classification algorithms.
All of these methods \cite{21, 22, 23, 24} which are proposed for detecting the tubules make some strong assumptions which can not be applied in our data set. First, the proposed methods relay on appearance of lumen in the middle of tubules. They find all bright regions in the image and then classify them as tubule’s lumen if there is enough evidence. Second, they have only one type of cells in their images. Third, in their data set only one layer of one type of cell are surrounding the lumen.

As discussed in section \ref{sec:1.1} there are many tubules that appear without a lumen. On the other hand, three different type of cells exist in our images and several layer of cells can encompass one lumen. Therefore we introduce a new method for modelling the tubules which fit the tubular structure in our image better. It allows to have tubule not only without lumen but also with several layers of cells.
Chapter 3

Image Acquisition and Enhancement

The staining process and image enhancement method are described in this chapter.

3.1 Staining process

Testes were dissected from sexually mature mink immediately after euthanization. Slices of testicular tissue were fixed in Modified Davidsson’s fluid for 24 h at 4°C and embedded in paraffin wax. To differentiate Sertoli cells from spermatogenic cells the transcription factor GATA-4 was used as a marker. GATA-4 was localized in 4 µm thin section using a goat polyclonal antibody diluted 1:200 and standard immunohistochemical techniques. The procedure results in brown staining of structures containing the protein GATA-4.

We use this method to take 50 images. The images are used during method development and experiments.

3.2 Image Enhancement

Noise reduction and improving image quality is one the major steps in image processing and computer vision. It helps to remove artifacts and correct image acquisition problems. Noise is considered a high frequency component of the image. Although using low pass filter could decrease noise, it also smooths or might remove edges that convey important information in the image. Bilateral filter is a way to reduce noise in which both spatial-domain and intensity-domain closeness are considered to preserve the edges [25]. Since our images are noisy and their edge' information are too weak, we apply the bilateral filter to enhance the quality of images.
3.2.1 Bilateral filter

An image is modelled as a set of pixels and function \( f \) returns the intensity value at pixel \( x \) over domain \( D \). \( x, \xi \in D \) indicate pixels in the image. A low pass filter is defined as:

\[
h(x) = k_d^{-1}(x) \int_{D} f(\xi) \varsigma(\xi, x) d\xi
\]

(3.1)

where \( \varsigma(\xi, x) \) is the geometric closeness of the centering point \( x \) and neighboring points \( \xi \). The bold font shows that the variable may be multiband. \( k_d \) is computed as:

\[
k_d(x) = \int_{D} \varsigma(\xi, x) d\xi,
\]

that defines a low pass filter to preserve the dc component of low-pass signal. If the filter is shift-invariant, \( \varsigma(\xi, x) \) is only a function of the vector difference \( \xi - x \), and \( k_d \) is constant. Range filtering is similarly defined as:

\[
h(x) = k_r^{-1}(x) \int_{D} f(\xi) s(f(\xi), f(x)) d\xi
\]

(3.2)

except that now \( s(f(\xi), f(x)) \) measures the intensity similarity between the centering point \( x \) and the nearby point \( \xi \). The similarity function \( s \) operate in the range of function \( f \) and closeness function \( \varsigma \) operate in the domain of \( f \). The normalize constant \( k_r \) is defined as

\[
k_r(x) = \int_{D} s(f(\xi), f(x)) d\xi,
\]

called unbiased if it depends only on the difference \( f(\xi) - f(x) \).

Combining both intensity similarity and geometrical closeness is defined as bilateral filter, thereby it enforce both geometric and photometric locality. Bilateral filter is defined as:

\[
h(x) = k^{-1}(x) \int_{D} f(\xi) \varsigma(\xi, x) s(f(\xi), f(x)) d\xi
\]

(3.3)

where the normalizer is defined as:

\[
k(x) = \int_{D} \varsigma(\xi, x) s(f(\xi), f(x)) d\xi.
\]

Therefore, it replace the intensity value of each pixel with the average of nearby pixel with similar value.

The bilateral filter is controlled by two parameters. First, the geometry closeness weights determining the influence of neighboring pixel based on spatial distance. Second, intensity similarity determining the influence of neighboring pixels based on photometric distance [20]. In other words, it smooths each pixel’s intensity by considering its geometric neighbors and their intensity similarity with the current pixel which is actually done by two Gaussian functions.
Two Gaussian functions are applied to weight neighboring pixels based on their geometric and photometric distance. Pixels that are closer to the current pixel will get higher weights. Therefore the two controlling parameters of the bilateral filter are translated as standard deviation of the Gaussian functions applied at spatial and intensity domain.

### 3.2.2 Experimental result

Since the distance between color in RGB (Red, Green, Blue) color space does not reflect their perceived difference, the color space of our images are changed to the Lab (also known as $L^*a^*b^*$ or CIELAB) and mapped into $[0 1]$ interval. The Lab color space approximately model colors as perceived by human vision \[27\]. The L-channel in Lab closely matches human perception of lightness. The distance between colors in Lab approximates their perceived difference.

In order to get a clean and sharp image, the bilateral filter is applied on the Lab color space of the image. Figure 3.1 shows the result of applying the bilateral filter on a sample image. The standard deviation of Gaussian kernel in spatial and intensity domain are set to 3 pixels and 0.2, respectively. As shown in the figure, noise is removed in a good extent and in the other hand, the image contrast gets better and edges are preserved.

![Figure 3.1: (a) the original image, (b) the image after applying bilateral filter.](image-url)
Chapter 4

Modelling Tubules

In this section a scheme is proposed to model tubules. This model is later used to detect the Sertoli cells in chapter 5. Three types of information are combined to model the tubules.

First, the nuclei are identified and clustered into the germ cells and the non-germ cells. The germ cells are very dense and by connecting them we can construct a germ-mass which represents the inside layer of tubules, section 4.1. All cells are detected by fast radial symmetry transform, section 4.1.1. We use textures as features and apply k-means clustering and graph cuts to group the cells, more detail is mentioned in section 4.1.2. Then the germ cells are connected to form the germ-mass which is described in section 4.1.3.

Second, the lumens are segmented. We find the best candidate regions for the lumen and use them for the initialization of level-set to obtain smooth segmentation of the lumen. The lumens are combined with the germ-mass to adjust our model for the tubules. More information can be found in section 4.2.

Third, we extract all edges in the image and exclude the edges which are in the germ-mass region. An algorithm is proposed to find edges that are more similar to line and fit straight lines to the edges. Then those lines are connected to fill the gaps in the boundary of tubules. Section 4.3 describes this step in more detail.

4.1 Construct the germ-mass

Three steps to form the germ-mass are described in this section. Section 4.1.4 contains the results of using the proposed method on a sample image.
4.1.1 Cell detection

In this section a method is proposed to detect all cell regions while the other parts are treated as background. Since the cell regions are in a roundish shape and have symmetry in their gray value, the Fast Radial Symmetry Transform (FRST) is applied to detect them.

**Fast Radial Symmetry Transform (FRST)**

Circular symmetry is one major property of the cells. Loy and Zelinsky [28], proposed fast radial symmetry transform to detect points of interest. They apply this transform to find symmetric points in the image by looking at some specific points along gradient direction. The FRST is computed for a set of one or more radii \(N\) that defines the radii of symmetry feature points. At each radius \(n \in N\), an orientation projection image \(O_n\) and a magnitude projection image \(M_n\) are computed. For each pixel \(p\) positively-affected-pixel \(p_{+ve}(p)\) and negatively-affected-pixel \(p_{-ve}(p)\) are defined as

\[
p_{+ve}(p) = p + \text{round}(g(p) || g(p)||_n),
\]

\[
p_{-ve}(p) = p - \text{round}(g(p) || g(p)||_n)
\]

where \(g(p)\) is the gradient of image, \(||g(p)||\) is the gradient magnitude and \(n\) is the neighborhood distance as shown in figure 4.1. The \(O_n\) is generated by increasing the values of the positively-affected-pixel by one and decreasing the same amount for the negatively-affected-pixel. The \(M_n\) is also computed by increasing the value of the positively-affected-pixel by the gradient magnitude at this point while decreasing the same amount for the negatively-affected-pixel. That is

\[
O_n(p_{+ve}(p)) = O_n(p_{+ve}(p)) + 1,
\]

\[
O_n(p_{-ve}(p)) = O_n(p_{-ve}(p)) - 1,
\]

\[
M_n(p_{+ve}(p)) = M_n(p_{+ve}(p)) + ||g(p)||,
\]

\[
M_n(p_{+ve}(p)) = M_n(p_{+ve}(p)) + ||g(p)||.
\]

The radial symmetry contribution at radius \(n\) is defined as the convolution

\[
S_n = F_n * A_n
\]

where

\[
F_n(p) = \frac{M_n(p)}{k_n} (\tilde{O}_n(p) \frac{1}{k_n} )^\alpha,
\]
and

$$
\tilde{O}_n(p) = \begin{cases} 
O_n(p) & \text{if } O_n(p) < k_n \\
k_n & \text{otherwise}
\end{cases}
$$

$A_n$ is a two dimensional Gaussian, $\alpha$ is radial strictness parameter, and $k_n$ is a scaling factor that normalizes $M_n$ and $O_n$ across different radii.

The average of all symmetry contribution in all radii is considered as full transformation

$$
S = \frac{1}{|N|} \sum_{n \in N} S_n. \quad (4.2)
$$

**Figure 4.1:** The positions of pixels $p_{+ve}(p)$ and $p_{-ve}(p)$ affected by the gradient element $g(p)$ for a range of $n = 2$. The dots represent all the pixels that can be affected by the gradient at $p$ for a radius $n$. This image is taken from [28].

**Applying FRST for cell detection**

Due to the special staining that we have in our data set, it is known that the saturated regions are the cells. Therefore, we have looked at saturation in both Lab and Hue Saturation Value (HSV) color space, and the saturation channel of HSV gives a better representation of the cells. FRST is applied to identify the cell regions in the saturation channel. The important parameters of FRST are:

- the set of radii that define the affected neighboring pixels,
- the radial strictness parameter $\alpha$, and
- the Gaussian kernel $A_n$. 

29
The radii define the spatial distance that are used to search for radial symmetry. It needs to be determined based on the sizes of regions of interest, i.e. the nuclei here. From (4.2), it could be seen that the symmetry contribution at each radius is computed separately and independently. As a result, we could have discrete set of radius that only includes the radii of the nuclei. Furthermore, choosing the radii close to the nuclei’s radius lead to have more precise and accurate radial symmetry transform.

The constant $\alpha$ defines how strictly radial the radial symmetry must be in order to return a high interest value. A Gaussian function is applied to distribute the influence of positively- and negatively-affected pixels. As mentioned in [28], the standard deviation of Gaussian function is chosen to be $0.25n$ where $n$ is the radius or distance of affected pixels, which allows to linearly scale the Gaussian kernel according to the neighborhood distance. We apply the FRST on our images to detect cells as bright symmetric region. The result is shown in section 4.1.4.

Then, we use mathematical morphology to tune the cell regions. The Mathematical morphology operations are set of shape oriented operations that process an image by passing a structuring element over the image. The two basic operations of mathematical morphology are erosion and dilation. Many complex operations are proposed based on these two operations that can be applied for both binary and grayscale images [29]. We filter out noise using image reconstruction by the mathematical morphology. The obtained regions are considered as nuclei. We also apply mathematical morphology operators to fill holes and smooth regions’ boundary.

4.1.2 Clustering cells

Now that we have segmented the cells, the next step is to cluster them into the germ and non-germ classes. We consider the cell regions as graph nodes. Each node has connections with all of its neighboring cells. Then we set up an energy function based on feature similarity and spatial closeness of the cells. The graph cuts algorithm is applied to find the best cuts to cluster the cells. As result, the detected regions in section 4.1.1 are labelled into two groups.

Energy minimization using graph cuts

Energy minimization is a very popular tool in image processing and computer vision. There are many problems in image processing and computer vision that can be formulated as an energy minimization problem. Graph cuts has had a significant impact in this area in the last decade.

To apply graph cuts, we need to formulate the problem as graph and define energy for each node and transition between nodes. A graph is a

\[1\text{such as: segmentation, image registration, image compression, etc}\]
A mathematical tool to model pairwise relation of objects. It is defined by a set of vertices or nodes, and edges that show connection between vertices. If the edges are ordered set, i.e., \(<a,b>\) is different from \(<b,a>\), the graph is called directed graph. Figure 4.2 shows two sample graphs. Circles represent vertices, and lines and directed arrows represent the edges in the simple and the directed graph, respectively.

![Graphs](image)

**Figure 4.2:** (a) simple graph, (b) directed graph, arrow shows direction.

Each node of the graph could be a pixel or a region in an image. The edges show connection between the nodes. Boykov and Kolmogorov reported some experiments that compare the efficiency of min-cut/max flow algorithms for application in computer vision \[30\]. They defined energy (4.3) as the prominent energy function applied in graph based methods. Suppose \(L\) is a set of possible labels and \(\ell\) is a function that assigns label- \(\forall p, \ell_p \in L\). \(\ell_p\) is the label at position \(p\). The energy function is defined as:

\[
E(\ell) = \sum_{p \in P} D_p(\ell_p) + \sum_{pq \in N} V_{pq}(\ell_p, \ell_q). \tag{4.3}
\]

The first term, \(D_p(\cdot)\), is the data cost, the second one, \(V_{pq}\), is the smoothness cost, and \(N\) is a set of all pairs of neighboring regions. The data costs reflect the cost of assigning a specific label to a region. The smoothness costs is a regularizer that try to penalize each \(\ell_p \neq \ell_q\) for neighboring regions \(p, q\) and encourage them to have the same label.

Recently, Delong et al. \[31\] extended the above mentioned energy function (4.3) to consider the cost of label set \(L\).

\[
E(\ell) = \sum_{p \in P} D_p(\ell_p) + \sum_{pq \in N} V_{pq}(\ell_p, \ell_q) + \sum_{L \subseteq L} h_L \cdot \delta_L(\ell) \tag{4.4}
\]

where the indicator function \(\delta_L(\cdot)\) is defined on label subset \(L\) as:

31
\[ \delta_L(\ell) \overset{\text{def}}{=} \begin{cases} 1 & \exists p : \ell_p \in L \\ 0 & \text{otherwise} \end{cases} . \]

In fact, the third term in (4.4) penalize based on the label subset that is used to segment image. In a simple case it could penalize the number of labels.

Graph cuts for clustering cells

To use graph cuts, we need to define the neighborhood graph for the cells. Each cell region is represented as a node in the neighborhood graph. Algorithm 4.1 is applied to define the neighborhood graph. The maximum distance to search for neighboring regions needs to be defined such that it is not too short, producing disconnected cells inside tubules, and also not to long, leading to too many inter-tubule connections. If the defined distance fulfills these constraints, the germ cells of different tubules will be connected through the Sertoli or interstitial cells. Therefore, if we could detect the germ cells, we can have the germ cells of each tubule separately by removing connections with the other cells.

**Algorithm 4.1 Build neighborhood graph**

*Input:* I, a binary image; maxDist, maximum distance integer; considerEdges boolean if true apply edge information; grayImage, a grayscale image with the same size as I

*Output:* An array that define the neighborhood graph of all regions in the binary image I.

```plaintext
for all Region in I do
    NR ← all regions that are closer than maxDist
    for all r in NR do
        if there is no region in-between r and Region then
            Add r to the neighboring list of Region
        end if
        if considerEdges and there is strong edges between r and Region then
            Remove r from the neighboring list of Region
        end if
    end for
end for
```

In the algorithm, edge information could also be applied to remove region connection. The edge information can be computed by any edge detector, e.g. the Canny edge detector, and their length or density of edge, can be considered as their strength. Since the staining process and noise could have
a big impact on this measure, and we need to adjust the edge detector and energy measure per each image, the edge information are not applied for the reported experiments.

In order to collect features, the L-channel in the Lab color space is chosen [27]. The minimum, mean, standard deviation and skewness of L-channel of each cell are used as the features. The features are normalized to have zero means and unit variance. Figure 4.3 shows some manually labeled cells from five different images belong to both groups using three dimension of the feature space. The blue dots represent germ cells and the red crosses represent other types of cells. The cells are shown by three features: minimum, mean and standard deviation. It shows that the features could provide a space to discriminate the germ cells from others.

![Figure 4.3: Some regions in three-dimension feature space](image)

K-means clustering is used to cluster the regions in the feature space. K-means clustering is an unsupervised algorithm that clusters the data into K clusters. It selects K points as the cluster centers and then assigns each observation into one cluster based on its distance to clusters’ centers. There are two policies to update cluster centers: either update the cluster centers when all observation are clustered, or update the cluster centers whenever a new observation is clustered. It repeats assigning observations to cluster centers and calculating new cluster centers. These operations are repeated until the displacement of cluster centers is less than a threshold or reached the maximum number of iterations [32].

Now, the clusters and neighborhood graph could be used to set up costs for the graph cuts algorithm. Because we only have two classes and label costs in (4.4) is constant, the energy function shown in (4.3) is applied. Consequently, we just need to define a proper data costs and smooth costs.

Algorithm 4.2 computes data and smoothness costs by considering the neighborhood graph, regions’ features, and the k-means clustering output. In the algorithm, $\| \cdot \|$ denotes the vector norm and using the colon as a index of an array means all the indices of the dimension of the array, i.e. cluster-
Algorithm 4.2 Define costs to apply graph cuts algorithm

**Input:** NG, neighborhood graph computed by 4.1; clustrCntr, 2D array contains the cluster centers; clusterId, integer array shows the cluster id of each region; features, a 2D array which each row contains the features of one region.

**Output:** dataCosts, 2D array defines data costs; smoothCosts, 2D array defines smooth costs

\[
\text{dataCosts} \leftarrow \text{zeros}(\text{numCluster}, \text{numRegions}) \quad \triangleright \text{zeros create a 2D array of the given size initialized with zero}
\]

\[
\text{smoothCosts} \leftarrow \text{zeros}(\text{numRegions}, \text{numRegions})
\]

\[
\text{NC} \leftarrow \text{zeros}(\text{numCluster}, \text{numRegions})
\]

for \( r \leftarrow 1, \text{numRegions} \) do
  for \( c \leftarrow 1, \text{numCluster} \) do
    \[
    \text{dataCosts}(c, r) \leftarrow \frac{1}{||\text{features}(r, :) - \text{clustrCntr}(c,:)||}
    \]
    \[
    \text{NC}(c, r) \leftarrow \text{length}(\text{clusterId}(\text{NG}(r)) == c)
    \]
  end for
end for

\( \triangleright \text{normalize the number of neighbors per each cluster by total number of neighbors} \)

for \( r \leftarrow 1, \text{numRegions} \) do
  \[
  \text{NC}(:, r) \leftarrow \frac{\text{NC}(:, r)}{\text{sum}(\text{NC}(:, r))}
  \]
end for

for \( r \leftarrow 1, \text{numRegions} \) do
  for all \( n \) in \( \text{NG}(r) \) do
    if \( \text{clusterId}(r) == \text{clusterId}(n) \) then
      \[
      \text{smoothCosts}(r, n) \leftarrow \max(\text{NC}('clusterId(r)), \text{NC}('clusterId(r), n))
      \]
    else
      \[
      \text{smoothCosts}(r, n) \leftarrow \max(\text{NC}('clusterId(n)), \text{NC}('clusterId(r), n))
      \]
    end if
  end for
end for
Cntr(1,:) means all elements of the first row.

In the algorithm, the inverse of the distance of each region’s features with the clusters’ centroids are considered as the data costs. In other words, a region in which features are closer to a cluster centroid is more strongly connected to that centroid. Since the cells of each cluster are mostly surrounded by the cells of the same cluster, we weight the connections between the nodes by considering both their labels and the number of neighbors per each cluster as follows.

We find the direct neighbors of cells. For each cell, the probability of different cluster labels in its neighborhood are calculated, which show the spatial tendencies of the region for the clusters. For example if a cell has three direct neighbors, and two of them have label $a$, the tendency of the cell for the cluster $a$ is $2/3$. For edges that connect two cells with the same cluster label, the maximum value of the tendency of the regions for that cluster is considered as the edge weight. On the other hand for edges that connect two cells with different cluster labels, the maximum value of the tendency of each cell for the cluster of another cell is considered as the edge weight. In fact, if one type of cell densely covers one part of the image, i.e. most nodes of the graph from this part belong to the same type of cell, the edges that make connection with the cells of this class will get high connectivity value. In other words, we encourage all the cells to get the label of the dominant class. Therefore the cluster label of those cells which are weakly connected to their current cluster label, i.e. their data costs is somewhat high, and their current class label is different form the dominant cluster label, will be changed to the dominant one.

4.1.3 Connect germ cells

The darker cluster with less standard deviation is chosen as the germ cluster. Then a mesh is defined over the image using this germ cluster and germ-mass is constructed.

Algorithm 4.3 constructs the germ-mass by starting from the germ cluster. Each region is connected to all neighbors within a certain distance. We know that the distance between the germ cells of different tubules is large. Thus we could expand the neighborhood to include germ cells at larger distance which later helps in constructing the germ-mass. We split the image by connecting each germ cell with its neighbors. By connecting the germ cells together, the lines between the germ cells split the image into some fragments. Therefore the area of fragments in parts of the image which are crossed by more germ cell connections, is small.

Although it is possible to have some connections between the germ cells of different tubules, due to high density of germ cell inside the tubules, the germ cell connections inside each tubule is much more than inter-tubule connections. As a result the area of fragments made by inter-tubule connections
Algorithm 4.3 Connect regions that are clustered as germ cell

**Input:** germCluster, a binary image that contains the germ cluster; lCH, the gray scale image; ND, an integer define neighborhood distance

**Output:** germMasses, a binary image that contains the germ-mass regions

\[ \text{germMasses} \leftarrow \text{germCluster} \]

for all region in germCluster do
    \[ NR \leftarrow \text{All regions that their distance is less than } ND \]
    for all \( r \) in \( NR \) do
        Set pixels of the line connecting \( r \) and region in germMasses
    end for
end for

\[ \text{closedRegion} \leftarrow \text{all background region surrounded by the connecting lines and the germ regions} \]

for all hole in closedRegion do
    if Area(hole) < 2000 then
        Set the hole region in germMasses
    end if
end for

\[ \text{germMasses} \leftarrow \text{Morphological opening of germMasses with disk structural element of size } 15 \]

\[ \text{germMasses} \leftarrow \text{FillHoles(germMasses)} \]

\[ \text{germMassesLbl} \leftarrow \text{All separate regions in germMasses} \]

for all \( r \) in germMassesLbl do
    \[ \text{brightRegions} \leftarrow \text{the size of all bright region in } r \]
    if Area\((r)\) < 1000 then
        Remove \( r \) from germMasses
    else if Area\((r)\) < 8000 and \( 0.5 \times \text{Area}(r) \leq \text{brightRegions} \) then
        Remove \( r \) from germMasses
    end if
end for
are much bigger than those made by intra-tubule connections. Therefore, we construct the germ-mass by combining both the smaller fragments and the germ cells. The morphological opening with disk structural element of radius 10 is applied to disconnect tubule that might get connected via few germ cells or incorrect clustered cells.

In algorithm 4.3, FillHoles fills all background regions that are surrounded by only one germ-mass. The obtained germ-mass regions are filtered by their sizes and the proportion of bright regions. If the proportion of the bright region is high, it is possible to connect the germ cells between tubules. The size of brightRegions is calculated by first finding a threshold level which is greater than the gray level of 70% of all pixels in L-channel of the image; and second computing the area of the regions greater than this threshold inside each germ-mass.

4.1.4 Experimental result

![Figure 4.4: The sample image used to test the proposed method.](image)

In this section we represent the result of the proposed approach to detect germ-mass on a sample image. The sample image is shown in figure 4.4. After applying image enhancement and removing noise, the image is used to detect the cells. The cell regions are detected by FRST in the saturation channel. Figure 4.5 shows the saturation channel of the sample image in HSV color space. As you can see, the more saturated areas shows the cells which make them more distinguishable in this channel.
As discussed, we need to define the set of radii and $\alpha$ for FRST. To evaluate the radial symmetry, the set of radii is defined as $\{4, 5, 6, 7\}$ which are roughly approximate radii of the cells in our data set. As shown in figure 4.5, the cell regions are not completely radially symmetric. Relaxing the radial strictness can help detecting non-radially symmetric regions. Therefore, the radial strictness $\alpha = 1$ was applied.

The FRST of the image is shown in figure 4.6. The output of radial symmetry for the bright pixels is positive and for the dark pixels is negative. In figure 4.6, positive values are shown as light pixels, negative values are shown as dark pixels, and midgray represents zero.

Since cell regions in the saturation channel was bright, the highly positive values of radial symmetry correspond to the cells. The points with positive value greater than one are considered as cells. White pixels in Figure 4.7 show the points with positive FRST greater than one.

Since the smallest radius of the cells is four pixels, morphological filtering with the disk structuring element of radius three is used to remove small regions. Opening followed by closing with the disk structuring element of radius two are applied to smooth the border of the regions, to eliminate holes and to fill gaps.

Figure 4.8 shows the obtained regions from FRST after applying mathematical operations. By comparing figure 4.8 and 4.7, it is clear that the small regions are removed and the regions’ borders are getting smoother. In figure 4.9, the obtained regions are shown in light green over the original
image. The figure shows that this method is able to capture all the Sertoli cells and almost all of the germ cells. It should be mentioned that sometimes small germ cells are missing. We are not interested in counting the number...
of the cells. Also the germ cells which are segmented here will be connected together to form the germ-mass. The small error that we have here will not affect our final result that much. It also shows that the mathematical operations correctly remove most of the miss-detected regions in-between the germ and Sertoli cells.

Then we make the graph in way that nodes represent the detected cells by using algorithm 4.1. The maximum neighborhood distance is 35 pixels which is chosen to be shorter than the closet distance between two germ cells from different tubules. Then the neighborhood graph and the two clusters obtaining form the k-means clustering serve as the input for the algorithm 4.2 to set up costs. The minimum cost is computed by graph cuts. By the costs that we have defined in section 4.1.2 each cluster try to conquer weakly covered nodes of the other one. The weakly covered nodes are those cells for which most neighbors belong to the other cluster.

Figure 4.10 shows the cluster of germ cells in green color over the original image. Even though the algorithm put some cells in an improper cluster, the germ cells inside the tubules are correctly clustered and they densely cover the inside of the tubules.

Finally, we apply the algorithm 4.3 to connect the germ cells and form the germ-mass. Figure 4.11 illustrates how an image is processed by the algorithm. Figure 4.11 a, shows connection between all cells closer than 55 pixels. The neighborhood distance is a bit relaxed compared to previous one when we create the neighborhood graph. There are two reasons for increas-
ing the neighborhood distance. First the Sertoli cells and interstitial cells are excluded and therefore the closest distance between nuclei of different tubules is increased. Second, each germ cell will have more neighbors. There will be more connections and the area of the fragments inside tubules will be smaller. Figure 4.11 b, depicts the germ-mass when all fragments are filled and c shows the obtained germ-mass regions after excluding the big fragments.

Morphological opening is applied to remove inter-tubule connections shown in figure 4.11 d. The background regions encompassed by one germ-mass is filled and small ones are deleted from the germ-mass set shown in e and f respectively.

4.2 Detect Lumen

In section 4.1.1 a method is proposed that detects the germ cells and builds germ-mass to model inside layer of each tubule. As we saw, some of the tubules have a lumen. It is possible to have disconnected germ-mass for tubules and this happens always for tubules that have lumen. Therefore in this section, a method is proposed to detect the lumens and combine them with the obtained germ-mass to improve the result.

As indicated earlier, lumens and some area between tubules are the brightest region, figure 4.12 (a)-(c). Therefore the large bright regions in L-channel of the Lab color space are identified by thresholding. Those re-
Figure 4.10: The cluster of germ regions

regions are used for initializing level set method to delineate the outer border of the regions, figure 4.12 (d).

Level set method is an active contour method that evolve curves by some internal and external forces. Li et. al. [33] defined the level set function $\phi(x,y,t)$ where the evolve curves are embedded as zero iso-contours of the function. The level set evaluation equation is

$$\frac{\partial \phi}{\partial t} = F|\nabla \phi|,$$

where $F$ is the speed function and controls the motion of the contour, and $\nabla$ is the gradient operator. The interior and exterior of $\phi$ in region $\Omega$ are determined by $\{ (x,y) \in \Omega | \phi(x,y,t) < 0 \}$ and $\{ (x,y) \in \Omega | \phi(x,y,t) > 0 \}$, respectively.

The most important task is to set up a level set function that could make a meaningful segmentation of the image. Li et. al. proposed an energy function for level set which consider both the edge and region information. The edge indicator defined as

$$g \sim \frac{1}{1+|\nabla G_\sigma * I|^2};$$

where $I$ is part of image belongs to region $\Omega$ and $G_\sigma$ is the Gaussian distribution with a standard deviation $\sigma$. The proposed energy function for level set function $\phi : \Omega \in \mathcal{R}$ is defined as

$$\epsilon(\phi) = \mu \int_\Omega p(|\nabla \phi|)d\mathbf{x} + \lambda \int_\Omega g\delta(\phi)|\nabla \phi|d\mathbf{x} + \alpha \int_\Omega gH(-\phi)d\mathbf{x}, \quad (4.5)$$

42
Figure 4.11: The result of processing the sample image by algorithm 4.3, (a) the connected germ cluster, (b) fill all fragments, (c) remove big filled fragments from germ-mass, (d) opening with disk structural element of size 10, (e) fill background regions surrounded by one germ-mass, (f) remove small region and medium bright regions.

where $p$ is the energy density function; $\mu$, $\lambda$ and $\alpha$ are the energy coefficients; and $\delta$ and $H$ are the Dirac delta function and the Heaviside function, respectively. Note that the fist part of (4.5) is the distance regularization term for the level set and the second and third term take into account the boundaries and weighted area of the object, respectively. In this experiment the level set algorithm implemented by Li et. al. [33], is used.
Figure 4.12: Processing the bright region in l-channel along with the obtained germ-masses to apply lumen information, (a) l-channel in Lab color space, (b) all bright pixels, (c) the big regions in (b), (d) the regions of (c) after applying level set, (e) the regions that satisfy the size and convexity condition, (f) red regions represent germ-mass and the yellow region is the only region in (e) that fulfil both (4.6) and (4.7).

The area and convexity of regions processed by level set, is used to remove some of the bright areas outside the tubules, figure 4.12 (e). Then potential lumens together with the germ-mass regions are considered and combined if they meet both (4.6) and (4.7) conditions.
\[ \mathcal{P}(\text{PLR} \cap \text{GM}) \geq \gamma \mathcal{P}(\text{GM}), \]  

(4.6)

where \( \mathcal{P} \) is function computing the perimeter of an object; \( \text{PLR} \) and \( \text{GM} \) are the potential lumen region (PLR) and germ-mass, respectively; and \( \gamma \) is a constant. This equation checks whether the boundary of common region between \( \text{PLR} \) and \( \text{GM} \) is \( \gamma \) percent of the \( \text{GM} \). The \( \text{PLR} \) could be either embedded by the \( \text{GM} \) or neighbor of the \( \text{GM} \) one. In the first case, the lumen is already included in the germ-mass, thus we do not need do any further step. For the second case, the \( \text{PLR} \) should touch almost half of the germ-mass’s border to combine them. Otherwise the \( \text{PLR} \) might be a bright area outside tubules, or the germ-mass region belongs to more than one tubule. Since the germ-mass borders are not completely smooth and symmetric, the \( \gamma \) is relaxed to be 0.8.

To check that the \( \text{PLR} \) is actually in the middle of a tubule, the proportion of perimeter of \( \text{PLR} \) and its shared regions with neighboring germ-masses is considered.

\[ \sum_{\text{GM} \in N_{\text{PLR}}} \mathcal{P}(\text{PLR} \cap \text{GM}) \geq \beta \mathcal{P}(\text{PLR}), \]  

(4.7)

where \( N_{\text{PLR}} \) is the set of all germ-masses that share regions with the \( \text{PLR} \). To join disconnected germ-mass, the coefficient \( \beta \) is 0.7. This condition allows to connect separated germ-masses when they belong to one tubule while eliminate non-lumen \( \text{PLR} \)s, figure 4.12 (f).

### 4.3 Detect the tubules’ borders

In section 4.1 and 4.2, a method is presented that models each tubule just by lumen and inside layer of the germ cells. In most of the cases this model could represent tubules, but using this model to detect the Sertoli cells could be error prone due to the following reasons.

First, we did not consider any information about the tubules’ boundary which is their main morphological property. Second, the boundary shape of the obtained germ-mass regions might be completely different from the boundary of their corresponding tubules. Thus it is not trivial to extract tubules’ boundaries from their germ-masses. Even in tubules which we have more than one germ-mass, the situation is getting more complicated. Third it is possible to obtain a germ-mass that connects two separate tubules.

It should be mentioned that since biologists are interested in detecting the Sertoli cells, they tried to apply a staining method that increase the contrast of the nuclei, but also leads to a vague appearance of the tubules’ boundary. To model and find information about the boundaries of tubules, a method is proposed that deploys the germ-mass regions and edge information, which is described below.
To extract edge information, the Canny edge detector is applied. Hence the edge information of the boundaries are too weak and the boundaries are thin. The standard deviation of Gaussian kernel is chosen to be one. The high threshold is chosen to accept 30 percent of pixels with highest gradient and the low threshold is set to 0.4 of the high threshold, figure 4.13 (a). The details of the algorithm can be found in [34].

To remove edges detected inside tubule, the edge information is masked with the obtained germ-mass. Then the shape of edges are analyzed to only keep edges that are line-like. Next, the edges are approximated to straight lines by principal component analysis (PCA). Finally the lines with similar slope get connected to fill gaps in the boundaries.

Podczeck [35] reported some measures based on the Feret diameters, area and perimeter to assess the shape of particles. He proposed equation (4.8) to measure the deviation of a shape from a circle. The deviation of regions
are calculated as

\[ H = \frac{\text{Area}}{(\pi/4)l^2}, \]  

(4.8)

where \( \text{Area} \) is the area of the region in image and \( l \) is the length of region perpendicular to the shortest Feret diameter. Since the area of region is compared with the area of the smallest circle that could encompass the region, the value of \( h \) is low for elongated regions. As a result we only keep edges for which \( H < 0.1 \), that eliminates almost all the cell edges which have more circular shape, figure 4.13 (b).

When the line-like edges have been found, PCA is applied to approximate them with straight lines. PCA is a mathematical procedure that transform data into an orthogonal space using eigen-decomposition of data. Each dimension of the new space is calculated such that the data has largest variance along this direction, and it also perpendicular to the previous dimensions. The eigenvectors represent the base vector of the new space and eigenvalues determine the data variance along their corresponding eigenvector [36].

The eigen-decomposition of the coordinates of the edge points is computed. The eigenvector with largest eigenvalue is selected as the direction of the edge. The two farthest points in this direction are chosen as the end points of the line. The red lines in figure 4.13 (c) depict the approximate lines. Large value for the smallest eigenvalue means a big variance along this direction. This happens when the edge is bended. Therefore approximating the line using only the biggest eigenvector leads to a large error. The edge is split in the middle and the line is fitted for each part separately. We can repeat dividing and edges check the smallest eigenvalue until it is small enough. In this experiment, we only repeat breaking lines with large smallest eigenvalue for one iteration, which is shown by green line in figure 4.13 (c).

\[ \text{Figure 4.14: A sample line with slope s is shown and its neighborhood region is represented by gray color.} \]
Now we have approximated all the edges with straight lines, the slope of the lines are converted into degrees. Then three conditions are verified to connect edges which are represented by the lines now. First, neighborhood distance determines the maximum distance between endpoint of the current line and neighboring lines. The distance specifies the longest gap between edges that could be connected. If this distance is too large, it may connect lines from one side of tubule with another side. On the other hand, if it is too short, it may not be able to compensate discontinuity at the parts of boundaries. Second, angular distance defines the interval of acceptable angles. This constraint allows to only connect lines that are in similar direction. Third, the neighbors of a line must not be in the line plane. The line plane is defined as the region between the perpendicular lines to the line at its endpoints. Figure 4.14 depicts a line with slope $s$ and its endpoints are shown with dotted circle. Gray areas are outside regions of the line plane.

In this experiment, the neighborhood and angular distances are 100 pixels and 20 degrees, respectively. The connecting lines are shown with blue color in figure 4.13 (c) and both approximated and connecting lines are depicted over the original image in figure 4.13 (d). It is shown that even though the edge information on the boundary of tubules is so weak and sparse, the algorithm is able to use the little amount of information to recover major parts of the boundaries. All lines are stored in an image, called boundary map.
Chapter 5

Detect the Sertoli Cells

5.1 Assign non-germ cells to germ-mass

In the previous chapter, a method is proposed to extract some information from testicular tissue images and model the tubules. The model is used to find the Sertoli cells per tubule. As discussed in sections 4.1.2, the detected cell regions are clustered into two germ and non-germ cells. We connect the germ cells to form the germ-mass regions. The germ-masses are used as the indicator of every tubules. The non-germ cells which are closer to the germ-masses are considered as the cells which are inside the tubules. These cells are the Sertoli cells that we want to count them. The closest germ-mass is defined as the germ-mass which is directly connected to the cell and there is no other cells in-between.

The non-germ cell cluster and their closest germ-mass are depicted in figure 5.1. We can see that there are interstitial cells wrongly connected to germ-mass regions. The boundary map is used to remove the false connections as described in section 5.3. It should be mentioned that due to lack of edge information in some images, we are not able to define a boundary map. Therefore the angle between cell and its neighbors are used instead of boundary map. More details can be found in section 5.2.

5.2 Angle to prune the Sertoli cells

Since the curvature of the tubule’s border is not high and the Sertoli cells are almost attached to the borders, each cell and its direct neighbors should be in a straight line. The connecting line of each cell with its corresponding germ-mass divide the image plane into two parts. The closest cells in each part which is connected to the same germ-mass is called the close neighbors of the cell.

We form a triangle by connecting the center of a cell and its close neighbors. Suppose we have a cell A and its close neighbors are B and C. If we
Figure 5.1: The germ-mass and non-germ regions are shown by green color over the original image. The blue lines connect each cells to its nearest germ-mass.

Figure 5.2: Triangle that connect three points, the name of each angle and side are written by capital and small later, respectively.

connect this cells and have the length of each side as shown in figure 5.2, the angle $A$ defined as

$$\angle A = \arccos\left(\frac{b^2 + c^2 - a^2}{2bc}\right).$$

If the angle is close to $180^\circ$, the cells are in same direction.

Figure 5.3 shows all the cells and their connection with the closest germ-mass. The white stars represent cells with $\angle A \neq 180^\circ$. If only a cell is
rejected by the angle condition and its close neighbors are accepted, the cell is correctly rejected by angle constraint. But if other cells among its neighbors are also rejected, it is possible to reject the Sertoli cells and accept wrong ones.

5.3 Boundary map to prune the Sertoli cells

In section 4.3, a boundary map is created by connecting edges in the image. All edges and their connections are approximated by lines in the boundary map. The number of edges that are crossed to connect a cell and germ-mass, is considered as the edge crosses for that connection. If the edge crosses of a connection is greater than two, the connection is removed. This condition is chosen because: first, edges are approximated with lines therefore we have some displacement, second, there are some miss-detected edges.

Figure 5.4 depicts candidate connections that can be removed. All cells that cross more than two edges in the boundary map are shown with red lines. As shown in the figure, almost all of the red connected regions cross the boundary of the tubules. The angle condition is only considered for the rest of cells with two or one edge crosses, and those that do not fulfill the condition are marked with star in the figure.
Figure 5.4: The Sertoli cells are connected to germ-mass. Red lines represent connections that cross more than two edges and stars represent the cell regions that cross at least one edge and are not in straight line with their direct neighbors.
Chapter 6

Result and Discussion

In this project a method is proposed to detect the Sertoli cells in microscopic images of thin testicular tissue. Figure 6.1 demonstrates how our method works. The name of each part is printed over its corresponding output result. The three main parts of the system according to figure 1.3 are indicated by dashed lines. The solid and dashed arrows represent the flow of the system between different parts and inside each part, respectively.

After image enhancement, the output result of each part are shown over the enhanced image. The germ-mass and the boundary map are represented by green and white color, respectively. The detected Sertoli cells are connected to their corresponding germ-mass via blue lines. Since the edge information of the image is too weak and the boundary map is sparse, the method only apply angle condition to find the suspicious Sertoli cells which are marked as white stars.

Since our method does not rely on the lumen to detect the tubule, it can identify tubules for which the lumen does not appear in the image. Whenever we have the lumen for a tubule, it will be added to the germ-mass and applied to adjust the germ-mass. It is possible to have more than one germ-mass for a tubule due to lack of germ cells. But if we have the lumen, one can use this to merge the germ-mass belonging to a tubule.

Because the bright regions outside the tubules and the lumens are very similar, they can incorrectly be considered as lumen. Therefore to combine the detected lumen and germ-mass our method not only considers the regions property but also checks its neighborhood to find enough evidence. The evidences are collected from neighboring germ-mass regions. As shown in figure 6.2 our method is able to correctly detect lumens. In figure 6.2 a, the germ-mass is formed by connecting the germ-mass without considering any information about the lumen. Although we have some false positives for the lumen, our method takes care of this problem and only chooses the correct ones, figure 6.2 b and c. We use the lumen as an indicator of the tubules and connect the germ-mass regions belonging to a tubule, shown in the first...
Figure 6.1: *How an image is processed through the proposed method*

row of the figure. This structure nicely represents the tubules.

Since we consider the symmetry property of cells to segment them, the method is robust against distortion, noise and non-uniform illumination. In non-uniform illumination, the illumination over the whole image is varying, but we can say that the changes in a small neighborhood (like the maximum radius of our cells) is negligible. Hence the FRST computes the symmetry of each point by only considering the gradient and gray value in a certain neighborhood, the non-uniform illumination has little impact on the radial
Figure 6.2: Chose and combine the lumen candidates with the germ-masses, (a) the formed germ-mass by connecting the germ cells, (b) lumen candidate, (c) the join of the germ-masses and detected lumens.

Figure 6.3 depicts some images that suffer from distortion. Even though noise and artifacts have severely destroyed some parts of the images, our method works well. For example as shown in the first images, noise has changed a big part of the image, we can still detect the germ-mass regions and the Sertoli cells. The intensity-based methods such as thresholding, pixel classification, will fail in this situation due to the huge intensity changes in a big part of the image.

Our proposed method for modelling tubule’s border is able to estimate the boundary of tubule even for tubules whose edges are weak and do not have edge responses for a quite large part of their boundaries. Figure 6.4 a, depicts the result of our method in modelling the border of tubules. The red and green lines represent all lines which are computed by approximating the edge responses outside the germ-mass regions. The blue lines show connecting lines that provided by our method to compensate the missing part of the boundary. All lines are shown over the original image with blue color in figure 6.4 a. As you can see in this figure our method works well to estimate the boundary map even in cases that most part of the boundary is missing.

Taking into account that the Sertoli cells are on the borders of the tubules and inside it. If we could collect enough edge information to model the borders of the tubules, the output result is reliable. The tubules can be correctly modeled to detect their Sertoli cells. Otherwise the detected germ-mass regions can be used.

We find the closest non-germ cells neighbors of the germ-mass and check symmetry value of pixels.
angular condition for those cells. If the close neighbors of a cell fulfill the angular condition and the cell does not, this cell is correctly rejected by the condition. For the cells that the cell and one or both of its close neighbors are rejected by the angular condition, it is not a trivial task to decide whether accept or reject them. These cases can be reported to be corrected by experts.
6.1 Conclusion

In this project, a method is proposed to analyze the microscopic images of thin section of testicular tissue. Our method models epithelial tissue using the germ cells and the bright areas as tubules. The model is later applied to detect the Sertoli cells.

Our results show that since we rely on finding cells to model the tubules, the method is able to successfully model tubules even though the lumen does not appear in the images. Regardless of the density of the germ cells, the method always detects the tubules. Our experimental results also show the robustness of the method against non-uniform illumination, noise and artifacts. If the edge information of the tubules’ border could be extracted even partially, it has been combined with germ-mass model to build a reliable and accurate model. Even though the edge information might not be available, the method still produce good results. In this case, since we only use the angular condition to verify the Sertoli cell candidates, it may reject true Sertoli cells. These suspicious cases can be reported to the experts.

All in all, the system can be used to analyze and compute measures for medical images which include one or several type of cells and an spatial relation could be defined between the cells. Our method for approximating the border of tubules can be applied for modeling shapes which only partially appear in image.

6.2 Future work

The method introduced in this project can be further improved and adopted for different application in medical image processing and image processing, For instance,
(I). The local maximum in the *Fast Radial Symmetry Transform* can be used for initializing a level-set algorithm to delineate the cells. This can extract more accurate cell region, if an energy function can be defined which has similar value for all cells.

(II). The neighborhood of cells can be computed by region competitions, e.g. assign each pixel to the closest cells. Different cells that are in competing with the current cell will be the neighbors of the cell. As result the neighborhood distance parameter will be removed.

(III). The connecting lines in boundary map can be adjusted more by using an energy based method such as the snake. The energy function can be computed using line detector algorithm such as the Danielsson line detector [37].

(IV). The germ-mass and cell cluster can be corrected using our boundary map and bright areas. E.g. if a germ-mass is divided by the boundary map, the germ-mass wrongly covers this region and some of the cells are miss-clustered as germ cell. The bright area could also help in such a situation to decide whether the germ-mass cover one tubule or more.

(V). The boundary map can be deployed for approximating different shapes in any image. The precision of the approximation can be adjusted using eigenvalues of the PCA.
References


Appendix A

Sample Images