Automated Region-based Prostate Cancer Cell Nuclei Localization

Part of a Prognostic Modality Tool for Prostate Cancer Patients

Nilgoon Zarei, M.A.Sc., Amir Bakhtiari, M.A.Sc., Jagoda Korbelik, B.S., Anita Carraro, B.S., Mira Keyes, M.D., Martial Guillaud, Ph.D., and Calum MacAulay, Ph.D.

BACKGROUND: Prostate cancer is a disease of disrupted cell genomes. Quantification of DNA from cytology preparations can yield prognostic information about tissue biological behaviors; however, this process is very labor-intensive to perform. Quantitative digital pathology can measure the structural chromatin changes associated with neoplasia and can enable prognostic and predictive assays based on imaging of sectioned prostate tissue.

OBJECTIVE: To design an automated system to recognize and localize cell nuclei in images of stained sectioned tissue (first step in enabling quantitative digital pathology).

STUDY DESIGN: Images of Feulgen-thionin–stained
Prostate cancer tissue microarray constructed from the surgical specimens of 33 prostate cancer patients were acquired for this study. We implemented a new image segmentation technique to overcome tissue complexity, cell clusters, background noise, image and tissue inhomogeneities, and other imaging issues that introduce uncertainties into the segmentation method and developed a fully automated system to localized prostate cell nuclei.

RESULTS: We applied our algorithm on our dataset and obtained a 96.6% true-positive rate and a 12% false-positive rate.

CONCLUSION: In this paper we present a new method to automatically localize thionin-stained prostate cancer cells, enabling the extraction of various nuclear and cell sociology features with high precision. (Anal Quant Cytopathol Histpathol 2016;38:59–69)

Keywords: biomedical imaging, computer-assisted image analysis, computer-assisted image processing, Feulgen stain, Feulgen-thionin stain, image reconstruction, image segmentation, pathology, prostate cancer, quantitative digital pathology, tissue microarray analysis.

Introduction and Background
Prostate Cancer and Quantitative Digital Pathology
Prostate cancer (PCa) is the most frequently diagnosed noncutaneous malignancy in men. PCa is the second and third leading cause of cancer mortality for American and Canadian men, respectively. PCa is a disease of disrupted prostate cell genomes and their corresponding proteins. Can-

tatively. PCa is a disease of disrupted prostate cell genetics and genetic changes which can be phenotypically observed as modifications in cell nuclear structure and tissue architecture. It has been reported that DNA ploidy can serve as a surrogate biomarker to evaluate nuclear morphometric alterations. Further automated analysis of cell suspensions from prostate tissue can be used to recognize aggressive PCa based upon the DNA ploidy status of populations of prostate cells. Many groups have studied nuclear morphometric alteration as a prognostic method for PCa patients. This information can be used to predict the PCa pathologic stages. Although originally this process was performed manually, currently quantitative digital pathology modalities are undergoing a revolution. It has been shown that quantitative digital pathology improves diagnostic accuracy and increases the chance of recognition of early cancer. Using advanced image processing algorithms, the quantitative assessment of nuclear structure using optical microscopy is feasible for a variety of tissues in sectioned material. Computer-assisted imaging systems make it possible to accomplish the measurement of cell morphology semiautomatically or even in a fully automatic manner for well-separated cells. However, current automated diagnostic methods do not provide a sufficiently accurate and reliable means for stratifying risk because they cannot distinguish the aggressive, potentially lethal cases from the indolent, slow-growing PCa cases.

Cell Segmentation/Localization

Analyzing cells and nuclei is the key initial step for various histopathology studies involving cell type recognition, cell count, cellular sociology analysis in digital pathology, etc., which highlight the importance of accurate cell segmentation. There are numerous studies on cell segmentation with different approaches. Meijering, in his review paper, intuitively summarized common cell segmentation methods. One of the well-known and relatively well-used methods is the intensity thresholding approach. This method, although very old, is still being used as one of the underlying segmentation designs. In the intensity thresholding algorithm a threshold is used to divide intensities of an image into foreground and background. This thresholding can be done with or without androgen deprivation therapy and/or brachytherapy or surgery. Transformation of normal epithelium into cancer cells involves epi-


genetic and genetic changes which can be phenotypically observed as modifications in cell nuclear structure and tissue architecture.
either globally or locally.\textsuperscript{25} Feature extraction,\textsuperscript{26} or filtering, is another method used in cell segmentation. Basically, in this method, invariance properties of a particular region are used to localize and extract objects.\textsuperscript{26} The other approach is morphological filtering.\textsuperscript{27} This technique is a nonlinear image analysis that extracts image objects’ information by using different morphological filters.\textsuperscript{27} Various manipulations, such as erosion, dilation, opening, closing, etc., can be used in this approach.\textsuperscript{25} Region accumulation\textsuperscript{25} is an alternative method to the previous approach; the segmentation initiates by choosing a seed point and adding points to the initial point to eventually cover the region.\textsuperscript{25} Watershed transform\textsuperscript{28,29} and region growing\textsuperscript{30} are two approaches that use this methodology; however, this method can result in oversegmentation\textsuperscript{31} and increase computational costs. Deformable model fitting\textsuperscript{32} is another segmentation method in which a deformable model fits into the input image and handles topological changes in the image. Active contours,\textsuperscript{33,34} commonly referred to as “state-of-the-art,”\textsuperscript{35} uses the same methodology, shows high accuracy, and is highly adaptable with image topology\textsuperscript{36,37}; however, this suffers from computational load, sensitivity to initialization, incorrect initialization, and numerical instability.\textsuperscript{35} Cell segmentation is an active research area, and various manual, semiautomated,\textsuperscript{38,39} and automated\textsuperscript{40-42} methodologies have been developed in the course of time. The aforementioned methods, solely or in combination, have been used to enhance segmentation quality.

Hypothesis and Objective

Despite digital pathology’s high diagnostic potential, it is primarily being used as a tool for preliminary diagnosis. We hypothesize that an analytical approach based on the combined measurements of nuclear and cell sociology features can delineate PCa progression/treatment failure risk with greater effectiveness than standard clinical approaches. The first step to achieve this goal requires one to be able to recognize, localize, and segment cell nuclei accurately within stained, sectioned prostate tissue. The segmented nuclei could then be used/accessed as part of an automated pathology system and their characteristics used to monitor and predict prostate cancer progress.

Materials and Methods

Images of Feulgen-thionin–stained PCa tissue microarray (TMA) spots constructed from the surgical specimens of 33 PCa patients were acquired using a whole slide scanner (Pannoramic MIDI scanner [3DHISTECH Ltd.] with ×40 objective); each image is 1024 by 1024 pixels, and the pixel spacing is 0.3 µm in both the x and y directions. The parameters used in this paper were chosen based on the relation between: pixel size and nuclei average size (nucleus size ranged from 20 to 800 pixels for 50 randomly selected nuclei). For each parameter we tried a couple of different values and chose the one that would generate the best result in a small subset (2 images) of our images. The Feulgen-thionin stain was used because it binds in a stoichiometric fashion with DNA, resulting in high-contrast images\textsuperscript{43} and well-defined nuclear features in which light absorption per pixel is directly proportional to DNA amount at the location imaged by the pixel.

Presented here is a novel algorithm (multistep) to automatically localize nuclei using MATLAB image processing functions and code.

We imported the TMA-scanned images of the 33 patients’ TMA cores into our in-house software, denoted D-unit, where 2 experienced technologists (J.K and A.C.) manually located (clicked) each cell, and an automated localization/segmentation was run for each cell; the result for each core was saved. These stored, selected nuclei locations were used as the cell location gold standard against which the algorithm presented herein is compared.

We used MATLAB as the prototyping tool because of its large image processing and machine learning toolboxes. Once an optimized approach has been identified, we plan to implement the methods into JAVA. The reasons for this necessary transition is (1) platform independence and (2) incorporation of the methods into existing object-oriented JAVA software.\textsuperscript{44} Figure 1B shows an image of a thionin-stained prostate core from a TMA. This image is constructed from 16 subimages as acquired by the Pannoramic scanner shown in Figure 1A. To reduce memory usage, we select each subimage separately and then combined the results.

The initial preprocessing applied different filters to the images to reduce photonic and systematic device intensity noise; however, these filters might reduce the image quality and blur the image. In this work we applied a simple Gaussian filter.

After the above filtering step we converted the RGB color image into a greyscale image. Since the
nuclei are stained blue, we constructed the grey image by averaging Red and Green channels (Blue channel does not carry useful information about the nuclei). To construct the binary image from the grey image, 2 different methods were considered. The first approach was to use the gradient weighted thresholding method, this is an effective method that has been applied to many medical images. In this method the gradient image was calculated by using a Sobel filter.

We segmented our images with gradient weight (GW); however, the segmentation performance of this thresholding was reduced due to intensity inhomogeneities between and within nuclei (Figure 2B). To address these shortcomings, we implemented a second thresholding method, the mean gradient (MG). This method is reported by Watanabe for thresholding cytology images. MG is a variant of GW that tries to reduce the effect of isolated high values of the gradient magnitude by dividing the sum of the gradient of image intensity by the number of pixels of that intensity. The threshold calculated from MG histogram is given by:

\[
T = \arg \max_{i} \left\{ \frac{\sum_{(x,y) \in f(x,y)} g(x,y)}{n_i} \right\}
\]

where \( n_i \) is the number of pixels with the intensity of \( i \). We then segmented our images with the MG method, as shown in Figure 2C. From this image one can vividly observe that MG can segment nuclei more accurately than GW.

As the next part of the process we removed small debris from the image by detecting and removing all the small objects (smaller than 10 pixels) and then labeling the remaining connected components, as shown in Figure 3.

Our approach then processes each connected and labeled component individually and proceeds with further object segmentation for each connected component as necessary. The steps for the algorithm are shown in Figure 4.

Now that we have divided each image into multiple regions (connected labeled objects), it is possible to analyze each region independently. Figure 5 compares the histogram of the whole image along with an example histogram from a selected connected object.

Figure 1
Thionin-stained PCa core from a tissue microarray. (A) Sixteen subimages. (B) The whole prostate core image obtained from (A).

Figure 2
Process of image thresholding by mean gradient histogram and gradient weight. (A) Input image, (B) binary image obtained from gradient weight, and (C) binary image obtained from mean gradient.
To further process each labeled object, we applied the Otsu method to the region of interest and subsequently extracted only the object that contains the pixels from thresholding the input image with the MG method. (Since the histogram of the original image, Figure 5B, does not have a clear valley, we expect that the initial segmentation with Otsu’s method will fail, so we did not apply Otsu thresholding for the initial segmentation.)

Although our proposed algorithm was very successful for the localization of well-separated nuclei, it had difficulty with overlapping nuclei in a cluster, as shown in Figure 6A–C.

In order to address the clustering issue, we computed the image skeleton and chose each endpoint of the skeleton as the initial guesses (list of potential centers) for the nuclei centers; then we used a Hough transform to fit an elliptical shape to the edge of the surrounding area, and we chose the ellipse that had the nearest centroid to the initial guess of the object center, and its area (number of pixels) is similar to the average nucleus size. After that we removed the detected object from the cluster (any other endpoint located in the localized cell is also removed from the list of potential cells [Figure 7, middle row]); we then continued with the previous step on the remaining cluster objects, as shown in Figure 8.

We applied this method to the images and the process proceeds as shown in Figure 7. This figure illustrates 3 different object types: (1) isolated nucleus localization (bottom row), (2) single nucleus with a small overlapping neighbor (top row), and (3) cluster of nuclei and localized cells in the first 2 rounds of algorithm (middle row). Final image: estimated centroids of nucleus.

Figure 7 demonstrates the effectiveness of the algorithm for localization of different cases (isolated nucleus, single nucleus with overlapping neighbor, and cluster of nuclei). Figure 9 shows the second part of the algorithm used for the segmentation of both single nucleus and cluster of nuclei.

Figure 10 shows the results of our algorithm on one of our images. In this image red dots represent cell nuclei.

The presented algorithm was successful in localizing both single nucleus and nuclei within cluster, as shown in Figure 11.

The last step after segmentation is saving each nucleus for further analysis. Figure 12 shows a small set of our extracted nucleus images.

A summary of our whole algorithm design is illustrated in Figure 13.

Results

We ran our algorithm on the thionin-stained PCa image dataset from 33 patients and compared the results with human annotation, as shown in Table I.

As seen in Table I, our proposed method missed around 3.4% of nuclei (96.6% true-positive rate) and had a 12% false-positive rate. This rate varied considerably between the individually imaged cores, as can be seen in the true-positive and false-positive rate figures of our 33 TMA cores (Figure 14).
Figure 5
Global thresholding versus local thresholding. (A) Whole image, (B) histogram of the whole image, (C) region of interest, and (D) histogram of (C).

Figure 6
An example of localization failure with simple thresholding. Failure is due to cluster of nuclei. (A) Whole image, (B) cluster of nuclei, (C) mask from (B), and (D) resulting image from our method for comparison (implemented methods are explained in the following sections).

Figure 7
Nuclei localization for three different object types.
Figure 8
Localization method for clustered nuclei. (A) Cluster of nuclei, (B) skeleton cluster of nuclei, and (C) initial guesses.

Figure 9
Final nuclei localization algorithm design.

Input data: Roughly masked nucleus or clusters
Output data: cells centroids’ locations
   1. Apply Otsu thresholding to the roughly masked nucleus to generate a binary image
   2. Clean binary image by removing small objects (smaller than 10 pixels)
   3. Extract the connected component containing initial seeds
   4. If size and shape of connected component is similar to the generalized nuclear size and shape:
      Calculate centroid of the mask and mark it as the center of the nucleus
      Save extracted information and return
   Else:
      Calculate skeleton of the mask
      Find the end-points of computed skeleton
      While there is an end-point:
         Use the end point as initial point
         Generate rough mask around the initial point using a disk (with a radius of 15 pixels)
         Find fitted oval to the edges (Oval center near the initial guess with similar nucleus’s area, 800 pixels)
         Multiply the rough mask with the original mask
         Remove extracted nucleus from the original mask
         Calculate skeleton of the original mask
         Find the end-points of computed skeleton
         Save extracted information about nuclei and return

Figure 10
Cell nuclei detected and shown with red dots.

Figure 11
Nuclei segmentation results for different nucleus arrangement.
Table 1  Processed Nuclei Localization Results with Our Algorithm

<table>
<thead>
<tr>
<th>No. of manually annotated nuclei</th>
<th>No. of nuclei detected by our algorithm</th>
<th>No. of false positives</th>
<th>No. of missed detections</th>
<th>Overall false-positive rate</th>
<th>Overall true-positive rate</th>
<th>False-positive rate per core (mean ± SD)</th>
<th>True-positive rate per core (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>170,761</td>
<td>187,373</td>
<td>22,560</td>
<td>5,948</td>
<td>12%</td>
<td>96.6%</td>
<td>14.44 ± 9.5%</td>
<td>96.53 ± 3.68%</td>
</tr>
</tbody>
</table>
On average, we processed each image in about 2 minutes on a 3-GHz machine with 4 gigabytes of RAM, while the same process done by a human expert required 45 minutes. Figure 15 shows the comparison between human selection and our automated method. The accuracy of our proposed method is comparable with the accuracies reported in other recent publications on nuclei identification methods.46-48

Discussion

The importance of cell analyses for a vast range of biological studies has encouraged development of various cell segmentation methods. Here we compared our methodology with those of 2 recent publications. The method presented by Ali et al48 implemented an adaptive energy selective active contour to segment TMA prostate nuclei. They reported high specificity (1.0) and sensitivity (0.82); however, their method has a higher computational cost (running time of 250 seconds for a 200 × 200 pixel image). Our presented method processes a 1024 × 1024-pixel image in 118 seconds. Also, they reported the results for only 200 randomly selected nuclei, and, as such, these results may not perform as well when generalized over a much larger data set of cell images. In this study we investigated all of the nuclei (>170,000 nuclei) across multiple tissue samples with differing preparation and fixation, increasing the confidence that the reported results would represent the generalized performance of the algorithm. In a recent publication by Dimopoulos et al31 a new automated method is presented in which the cell segmentation

Figure 14  True-positive and false-positive rate results for each individual TMA image dataset of the 33 patients.

Figure 15  Comparison of human annotator (back dots) with our algorithm (red dots).
is done by analyzing the cell membrane/boundaries using directional cross-correlations and considering spatial constraints.\(^{31}\) We applied their online resource on our TMA images (we changed the adjustable parameter and chose those with the best visual results); the running time was very high and the resulted outcome was not as good as the algorithm presented here (Figure 16). We believe the difference in the results is due to the assumptions made by our algorithm, localizing nuclei, whereas they considered the information from the cell boundaries instead. This made our method more applicable to the images under consideration.

**Conclusion**

In this study we presented a new algorithm to accurately locate nuclei from Feulgen-thionin–stained PCa cell images. Unlike other methods,\(^ {46,47}\) our algorithm is not a supervised method (does not require training dataset). Also, the proposed method does not need an accurate description for the shape of nuclei. This algorithm has a low computation cost and produces results similar to methods with a higher computation cost.\(^ {46,47}\)

**References**

1. Canadian Cancer Society: Canadian Cancer Statistics 2012
2. American Cancer Society: Cancer Facts & Figures 2013


