



Unsupervised segmentation method for cuboidal cell nuclei in histological prostate images based on minimum cross entropy

Domingos Lucas Latorre de Oliveira ^{a,*}, Marcelo Zanchetta do Nascimento ^b, Leandro Alves Neves ^c, Moacir Fernandes de Godoy ^d, Pedro Francisco Ferraz de Arruda ^e, Dalisio de Santi Neto ^f

^a Center of Mathematics, Computing and Cognition, Federal University of ABC (UFABC), Santo André, SP, Brazil

^b Faculty of Computing (FACOM), Federal University of Uberlândia (UFU), Uberlândia, MG, Brazil

^c Institute of Biosciences, Letters and Science, Department of Computer Science and Statistics, São Paulo State University (UNESP), São José do Rio Preto, SP, Brazil

^d Interdisciplinary Center for the Study of Chaos and Complexity (NUTTECC), Faculty of Medicine of São José do Rio Preto (FAMERP), São José do Rio Preto, SP, Brazil

^e Surgery Department of Renal Transplantation, Regional Medical Faculty Foundation (FUNFARME), São José do Rio Preto, SP, Brazil

^f Department of Pathology of Base Hospital, Regional Medical Faculty Foundation (FUNFARME), São José do Rio Preto, SP, Brazil



ARTICLE INFO

Keywords:

Prostate cancer

Segmentation of cuboidal cells

Segmentation of nuclei

Minimum cross entropy

ABSTRACT

This paper presents a novel segmentation method for cuboidal cell nuclei in images of prostate tissue stained with hematoxylin and eosin. The proposed method allows segmenting normal, hyperplastic and cancerous prostate images in three steps: pre-processing, segmentation of cuboidal cell nuclei and post-processing. The pre-processing step consists of applying contrast stretching to the red (R) channel to highlight the contrast of cuboidal cell nuclei. The aim of the second step is to apply global thresholding based on minimum cross entropy to generate a binary image with candidate regions for cuboidal cell nuclei. In the post-processing step, false positives are removed using the connected component method. The proposed segmentation method was applied to an image bank with 105 samples and measures of sensitivity, specificity and accuracy were compared with those provided by other segmentation approaches available in the specialized literature. The results are promising and demonstrate that the proposed method allows the segmentation of cuboidal cell nuclei with a mean accuracy of 97%.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Prostate cancer is a serious public health problem with significant disease and mortality rates as well as reduction in quality of life. This disease is the fifth most common type of cancer in the world and the most prevalent in men (Taverna et al., 2009). There is evidence that prostate cancer is the result of the interaction of genetic, hormonal, environmental and nutritional factors. The main risk factors are age (most cases occur after the fifth decade of life), family history, male hormones (high levels of testosterone can increase the risk), diet (excessive red meat and animal fat), ethnicity, a sedentary lifestyle and obesity (de Arruda et al., 2013). The diagnosis is based on clinical (digital rectal exam), laboratorial (PSA) and radiological exams. However, the diagnosis obtained from these methods is controversial, considering the possibility of the indication for treatment without the clinical development of the disease, which can have a significant impact on the quality of life of patients (further details in de Arruda et al., 2013; Hugosson, 2000; INCA, 2011).

The use of microscopes equipped with digital cameras has assisted specialists in the diagnosis of prostate cancer. This technology allows the development and application of techniques for the digital processing of tissue samples (Basavanhally et al., 2013; Doyle, Agner, Madabhushi, Feldman, & Tomaszewski, 2008; Guillaud et al., 2004; Kayser, Riede, Werner, Hufnagl, & Kayser, 2002; Kong et al., 2008; Miedema et al., 2012; Monaco et al., 2010). The result has been a set of image processing methods known as computer-aided diagnosis (CAD) (Gurcan et al., 2009), the aim of which is to highlight, segment, quantify and classify regions of interest in prostate images.

In CAD systems, the segmentation step of the histological components is essential to the prediction of the disease investigated, especially if the process is performed automatically, as manual segmentation is highly dependent on the experience of the specialist and the number of samples examined (He, Long, Antani, & Thoma, 2012). Properly separating the desired signal can ensure the success of the subsequent stages that define a CAD system, such as the extraction of characteristics, classification and the analysis of possible patterns (Gonzalez, Woods, & Eddins, 2003; Pal & Pal, 1993; Sonka, 2000). For color images of prostate tissue stained with hematoxylin and eosin (H&E), the aim of the segmentation step is to separate stroma, lumens and cuboidal cells. The behavior

* Corresponding author. Tel.: +55 1123312705.

E-mail address: domingos.latorre@gmail.com (D.L.L. de Oliveira).

of cuboidal cell nuclei (CCN) is a commonly evaluated characteristic. The analysis of these regions can indicate abnormalities, such as hyperchromasia, prominent nucleoli and apoptotic bodies indicative of cancer, hyperplasia and atrophy (Epstein & Netto, 2007; Gleason, 1966).

Image segmentation methods are based on the assessment of discontinuities or similarities in the desired signals. Approaches founded on discontinuity assessments (detection of edges, lines and points) consider abrupt changes in the brightness of the pixels in the image (Wu, Merchant, & Castleman, 2010). Similarity assessments consider region growing, splitting and merging, clustering and thresholding (Gonzalez et al., 2003; Pratt, 2007; Solomon & Breckon, 2011). Hybrid methods use a combination of both these forms of assessments (Gonzalez et al., 2003). The method of choice depends on the characteristics of the image (Dougherty, 2009). For instance, segmentation methods for prostate images based on dis-

continuity or similarity assessments (specifically region growing and clustering) may provide inconsistent results (Solomon & Breckon, 2011) because prostate cancer alters the architecture of histological components. Thus, regions of CCN are poorly defined and non-standardized, with noise or smoothed (Epstein & Netto, 2007; Humphrey & Andriole, 2010; Zhou & Shah, 2011). On the other hand, methods based on splitting and merging are dependent on the identification of points (generally defined by manual intervention) for the separation of regions of interest. Moreover, noise in demarcated regions can exert a significant effect on the final result of the image investigated (Faruquzzaman, Paiker, Arafat, & Ali, 2008; Freixenet, Muñoz, Raba, Martí, & Cufí, 2002).

Due to the above limitations, specific methods have been developed to address the particularities of different tissue types and images. A semi-automatic method has been developed to segment areas of liver tumors in tomograms (Li, Chui, Chang, & Ong, 2012).

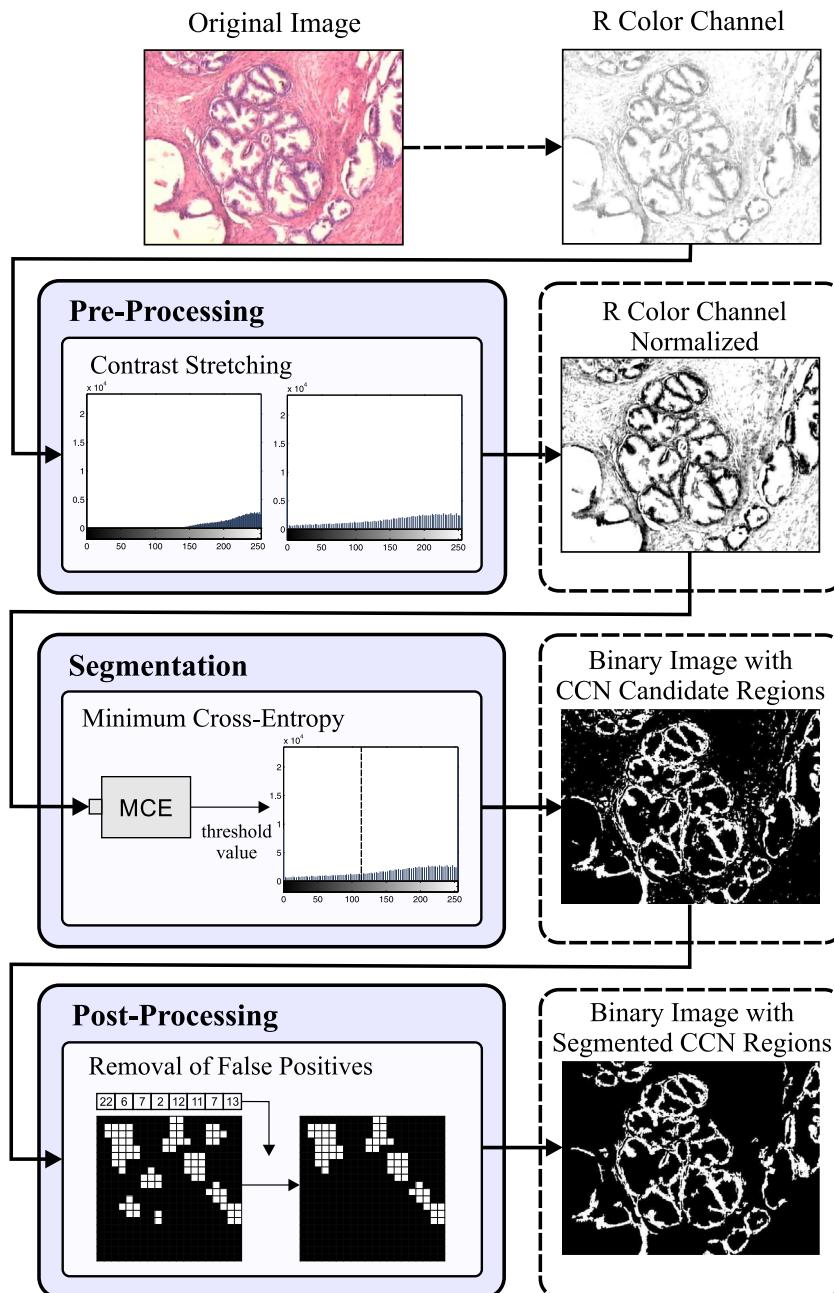


Fig. 1. Flowchart of method proposed for CCN segmentation.

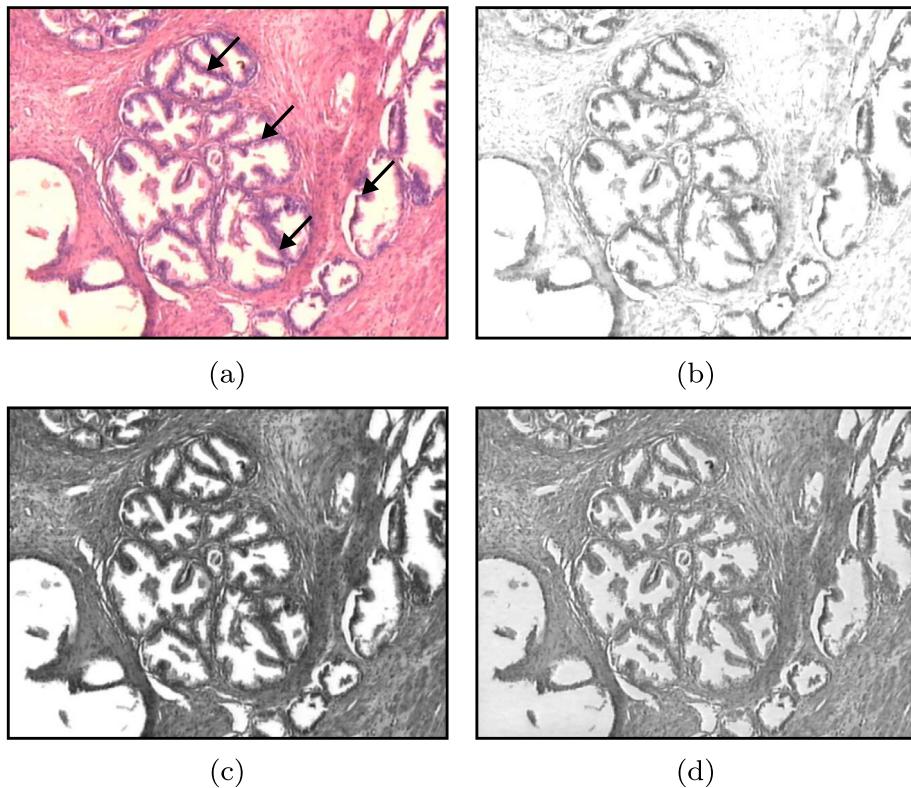


Fig. 2. Prostate tissue sample diagnosed with hyperplasia; arrows indicate CCN (H&E, 40×) (a); images of channels of RGB model: R (b); G (c); B (d).

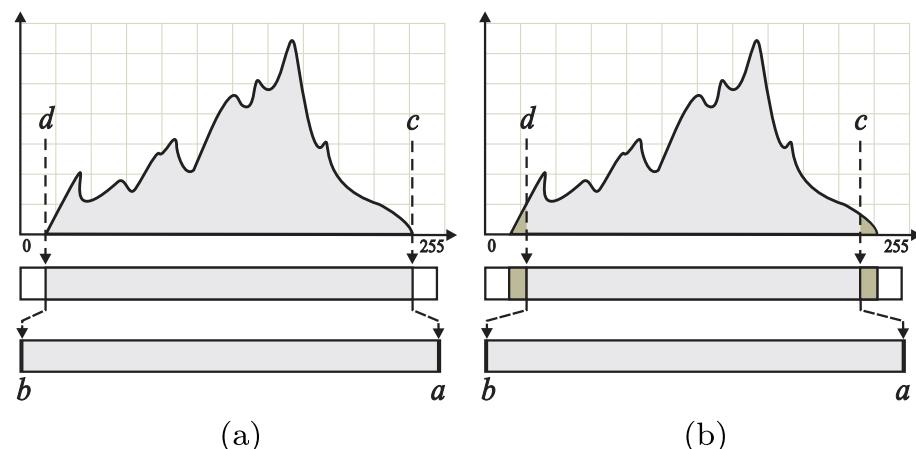


Fig. 3. Representative histograms for choice of output range limits of contrast stretching technique: lesser and greater intensity value (a) and definition ignoring a percentage of pixels (b).

A multi-level thresholding approach based on bio-inspired algorithms and mixture Gaussian functions has been used to segment gray levels in blood smear images (Osuna-Enciso, Cuevas, & Sossa, 2013). An adaptive bacterial foraging method has been proposed for fuzzy entropy optimization applied to the segmentation of gray images (Sanyal, Chatterjee, & Munshi, 2011). In the segmentation of histological images stained with H&E, different methods exemplify specificities: histology slides have been segmented to detect and quantify inflammatory polymorphonuclear leukocytes to aid in the grading of the acute inflammation of the placenta (Thomas, 2010); tissues from images of teratoma derived from human and nonhuman primate embryonic stem cells have been used to determine the performance of a method described by Chen, Ozolek,

Wang, and Rohde (2011); and nuclei, cytoplasm, stroma and lumens have been segmented in the images of breast tissue with lobular carcinoma in situ with the aim of ranking the malignancy level based on the assessment of nuclear morphological properties (Kim, Jeong, Choi, & Kim, 2009). The segmentation method described by Otsu (1979) is commonly employed on H&E stained images, mainly as a comparison and validation standard (Chen et al., 2011; Doyle, Feldman, Tomaszewski, Shih, & Madabhushi, 2011; Kim et al., 2009). This strategy is also employed in the present study.

The segmentation of CCN regions in images of prostate tissue stained with H&E is not an easy task. The difficulties reside in the separation of CCN in regions that are recognizably of stromal

tissue and considering the variations in CCN that stem from abnormalities. Prostate tissue in an advanced stage of cancer is characterized by infiltrations and CCN regions with no defined pattern, whereas hyperplastic images are characterized by an increase in cuboidal cells. Moreover, the shapes and colors of CCN and the nuclei of stromal cells are similar, which can lead to incorrectly segmented regions and errors in the CAD.

This paper presents a novel unsupervised CCN segmentation method for histological images of prostate tissue. The proposed method is based on thresholding and aims to separate normal, hyperplastic and cancerous images. For such, the first step consists of selecting the R color channel of a prostate image given as input. In the pre-processing step, contrast stretching is applied to the pixels in the selected channel in order to enhance the contrast of the CCN. In the segmentation step, minimum cross entropy (Li & Lee, 1993) is applied to identify candidate regions. This method was chosen because it offers consistent results in unimodal histograms with the distribution of identical or similar intensity between the background and object (Tsai, 1995). This unimodal distribution is found in the R color channel of prostate images. Moreover, it is not necessary to adjust the parameters or have a priori knowledge on the input image. In the post-processing step, false positives are removed using the connected component method. The proposed segmentation method was applied to an image bank of 105 samples. The methods proposed by Thomas (2010) and Otsu (1979) were applied for qualitative and quantitative evaluation of the proposed method.

This paper is organized as follows: Section 2 presents the image databank and the methods used in the pre-processing, segmentation, post-processing and qualitative/quantitative analysis of the method. Section 3 presents the results and discussion on the qualitative/quantitative evaluation. Section 4 presents the conclusions.

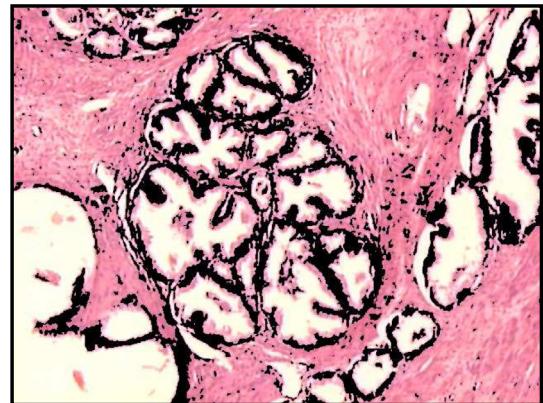


Fig. 5. Candidate CCN regions (marked in black) determined by segmentation.

2. Materials and methods

The proposed method was organized in three steps: pre-processing, segmentation and post-processing. Each step is described in detail in the following subsections. Fig. 1 displays the flow chart of the method.

2.1. Image bank

The image bank used in the present study was made up of histological slides from the prostates of 35 patients aged 50 to 75 years (mean age: 64.4 ± 5.9 years). The patients were treated at the Urology Service of the Base Hospital in São José do Rio Preto, state of São Paulo, Brazil, between 2007 and 2008. This study re-

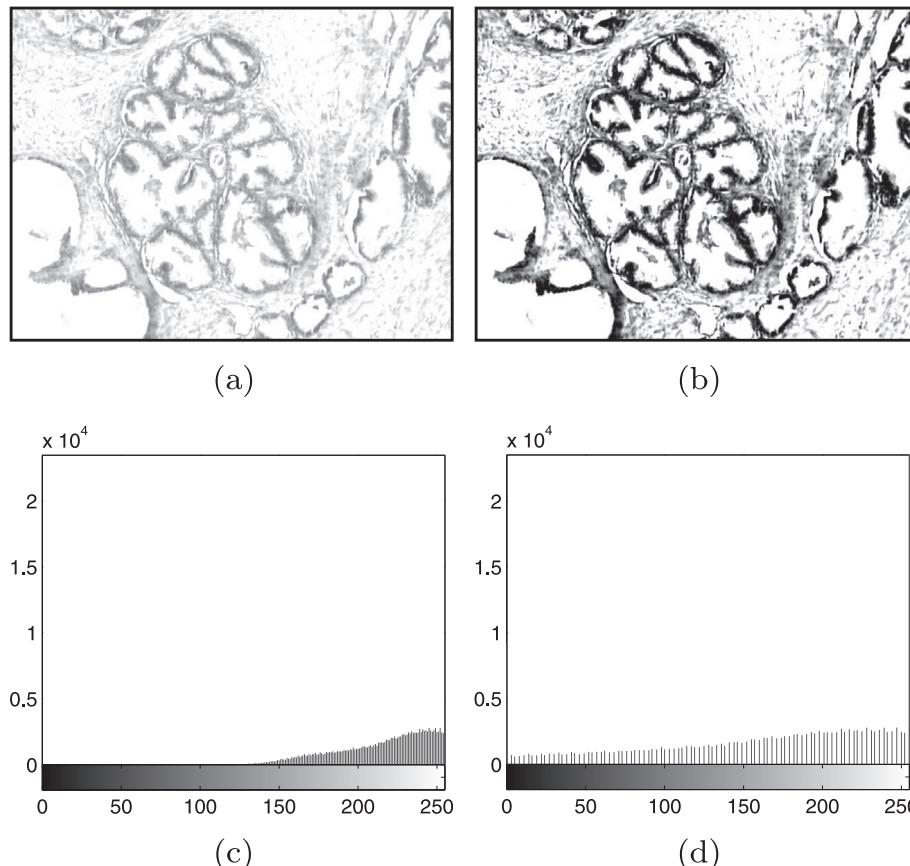


Fig. 4. R channel images before (a) and after (b) contrast stretching; R channel histogram before (c) and after (d) contrast stretching.

ceived approval from the Human Research Ethics Committee of the São José do Rio Preto School of Medicine (Brazil) (Hugosson, 2000).

The extraction of each prostate allowed the production of histological slides from fragments removed from normal, hyperplastic and cancerous areas. The material was fixed in a 10% formalin solution and embedded in paraffin. Serial cuts measuring 5 to 6 μm in thickness were made. The fragments were then stained with H&E. The microscopic images were acquired using a trinocular microscope (Olympus BX41) with a 10 \times plan achromatic objective coupled to a digital camera (Color Samsung SCC-131 with an Olympus U-TV1X-2 adapter). Each slide was photographed at a magnification of 40 \times and the image was saved in bitmap format using the RGB model, 1200 \times 900 pixels and 24 bits of quantization for subsequent analysis. The sample was made up of 105 images, which were divided into three groups of 35 images: normal, hyperplastic and cancerous.

2.2. Pre-processing

The contrast of an image is defined as the difference in brightness among the pixels. Enhancing the contrast of a histological component improves its quality and makes it more appropriate for the segmentation step (Sahidan, Mashor, Wahab, Salleh, & Jafar, 2008). Thus, the aim of the pre-processing step is to enhance the contrast of CCN in relation to the adjacent histological components. Fig. 2 displays the contrast in each channel of the RGB model

for an example of a prostate image. In this and all other images of the databank, greater contrast between the CCN and adjacent regions was found in the R channel.

The contrast in the R channel (Fig. 2(b)) is due to the characteristics of the H&E stain used in the tissue pigmentation process. Hematoxylin is attracted by nucleic acids in cell nuclei, which are subsequently stained a violet color. Eosin is attracted by basic elements in the cytoplasm of the cells, which are subsequently stained pink to red (Mescher, 2010). Thus, contrast stretching in the R channels was applied to enhance the differences in brightness between the CCN and adjacent regions. This technique is denominated normalization and allows reorganizing the range of intensity levels in an image or in a single color channel. Normalization within a single brightness interval of the histogram allows enhancing the contrast in relation to the other elements that make up the image (Gonzalez et al., 2003), as defined in Eq. (1):

$$P_{out} = (P_{in} - c) \left(\frac{a - b}{c - d} \right) + a, \quad (1)$$

in which, P_{out} is the intensity of the output pixel; P_{in} is the intensity of the input pixel; a and b are the upper and lower limits of the output range; and c and d are the upper and lower limits of the input range. Parameters a and b are respectively defined as brightness levels of 0 and 255. Parameters c and d respectively represent the highest and lowest intensity levels of the image. However, if the value of a pixel is located in a region near the brightness limits of the

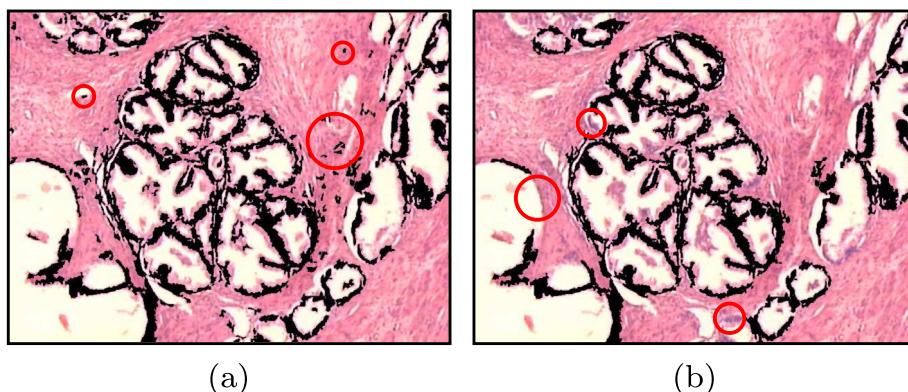


Fig. 6. Region of CCN marked in black following post-processing with n value set at 20 (a) and 250 (b); areas of false positive indicated by red circles in (a); areas of true positive are indicated by red circles in (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

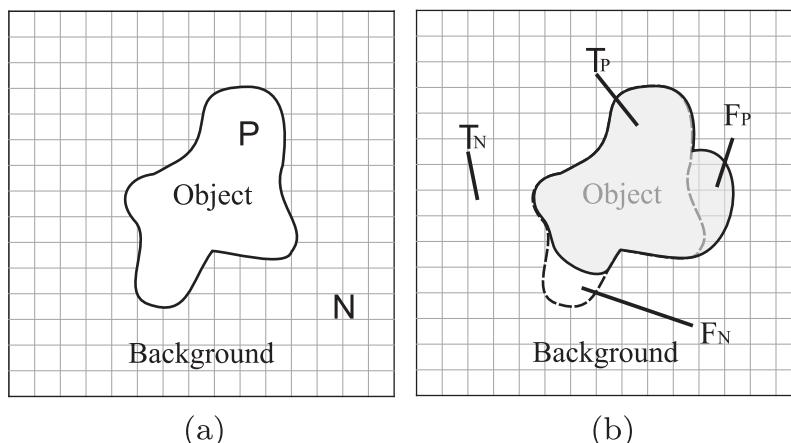


Fig. 7. Illustration of reference image showing segmentation performed by specialist with indications of P and N (a); illustration of overlaid segmented regions with indications of T_P , T_N , F_P and F_N (b).

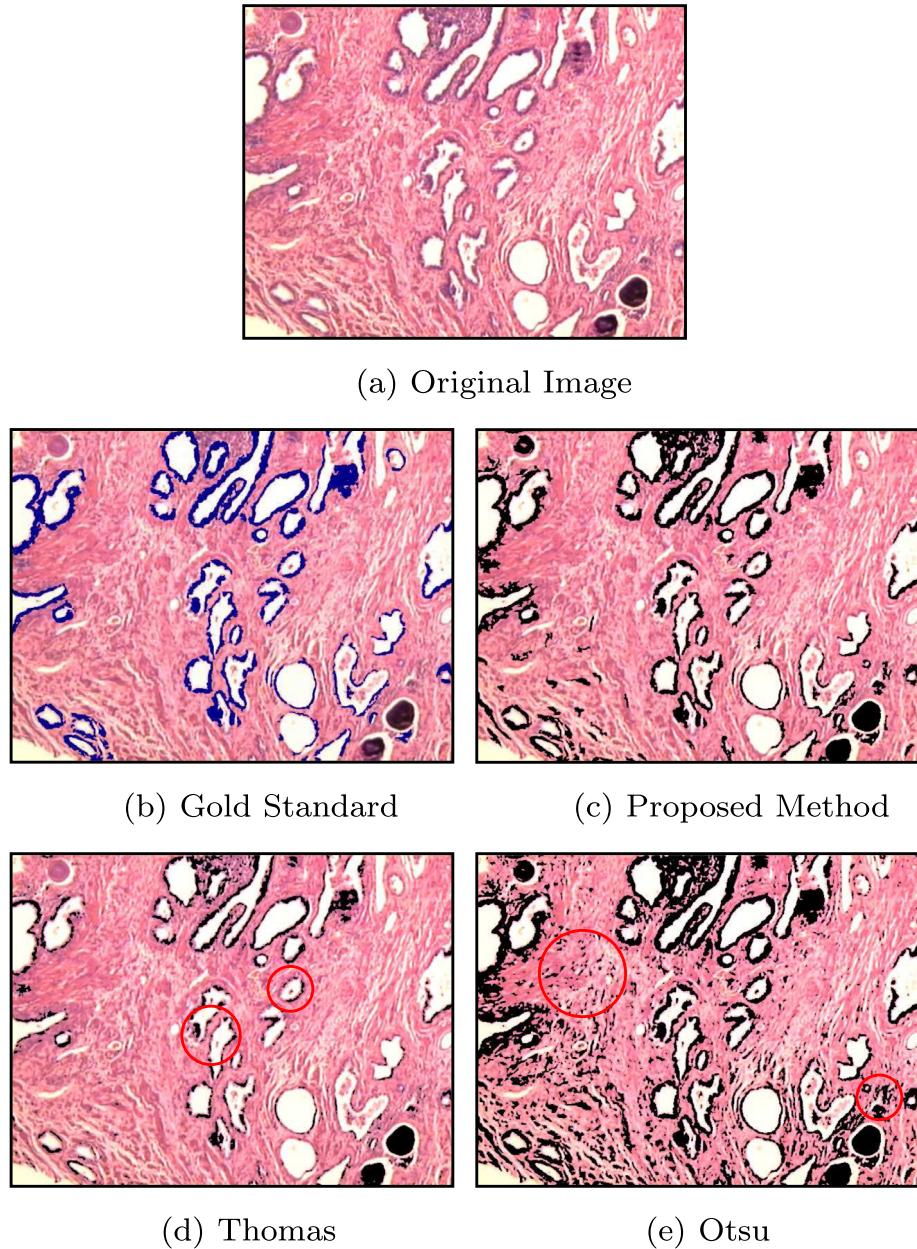


Fig. 8. Image of prostate sample diagnosed as healthy (H&E, 40×) (a); regions in blue manually selected by specialist (gold standard) (b); segmentation results achieved with proposed method (c), Thomas (d) and Otsu (e). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

histogram (0 or 255), which is defined as an outlier, the pixel is selected and the application of the method will not result in an image with enhanced contrast (Burger & Burge, 2009). This problem is illustrated in Fig. 3(a).

To ensure that the values of c and d represent the intensity information of the image, pixels near the lower and upper limits of brightness (range of 2%) are discarded. Thus, the 98th and 2nd percentiles of the R channels are chosen for c and d , respectively. This decision was based on the analysis of the histograms of the images investigated in the present study and is illustrated in Fig. 3(b).

Fig. 4 displays the images and histograms in the R channel before and after the contrast stretching technique.

The intensity level range was reorganized for the entire 0-to-255 interval (Fig. 4(d)), resulting in enhanced contrast of CCN (Fig. 4(b)).

2.3. Segmentation of cuboidal cell nuclei

In the set of images investigated in the present study, the behavior of the histograms of the R channel exhibited unimodal distribution. Minimum cross entropy (MCE) is widely used to define thresholds in histograms of this type (Li & Lee, 1993; Tsai, 1995) and has the advantage of not requiring user-provided parameters. This method is given by:

$$\eta(t) = \sum_{j=1}^{j=t-1} j h_j \log \left(\frac{j}{\mu_1(t)} \right) + \sum_{j=t}^{j=L} j h_j \log \left(\frac{j}{\mu_2(t)} \right) \quad (2)$$

with μ_1 and μ_2 given as:

$$\mu_1(t) = \frac{\sum_{j=1}^{j=t-1} j h_j}{\sum_{j=1}^{j=t-1} h_j}, \mu_2(t) = \frac{\sum_{j=t}^{j=L} j h_j}{\sum_{j=t}^{j=L} h_j} \quad (3)$$

in which $\eta(t)$ is the value of cross entropy for level of intensity t ; h is the probability of level of intensity j ; and L is the upper limit of intensity (255). The application of Eq. (2) yields threshold t_0 , which given by:

$$t_0 = \min(\eta(t)) \quad (4)$$

Fig. 5 illustrates the application of global thresholding (Gonzalez et al., 2003) with the threshold based on MCE. Candidate CCN regions are marked in black.

2.4. Post-processing

Cell nuclei are found in prostate tissue (cuboidal cells) and stromal tissue. The intensity of the pixels in these regions is similar due to the aforementioned properties of the H&E stain. Thus,

regions of cell nuclei in stromal tissue may be incorrectly detected. **Fig. 6(a)** displays regions marked in red with incorrectly identified areas. Two sub-steps are needed to remove these false positives.

In the first sub-step, clusters of pixels in the binary image are identified using the connected component labeling technique. This procedure attributes unique identifiers for each cluster of pixels in a binary image (Solomon & Breckon, 2011) and allows the individual analysis of each cluster. The algorithm used in the present study follows the general procedure described in Haralick and Shapiro (1992).

After identification of the clusters of pixels, it is necessary to remove those with an area smaller than a set n value. However, a set value may furnish inconsistent results for normal, hyperplastic and cancerous prostate images, since alterations stemming from the different processes that characterize hyperplasia and cancer need

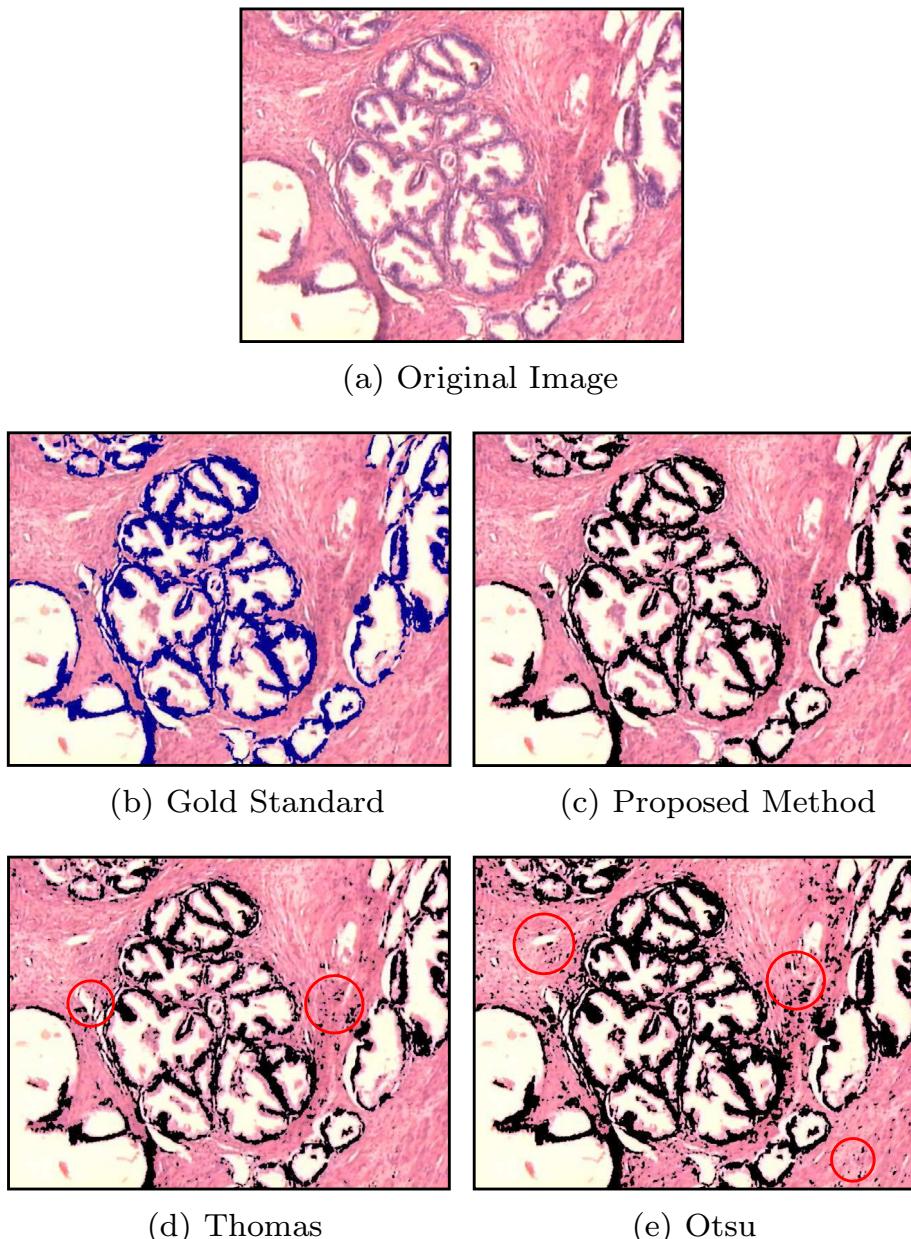


Fig. 9. Image of prostate sample diagnosed with hyperplasia (H&E, 40×) (a); regions in blue manually selected by specialist (gold standard) (b); segmentation results achieved with proposed method (c), Thomas (d) and Otsu (e). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

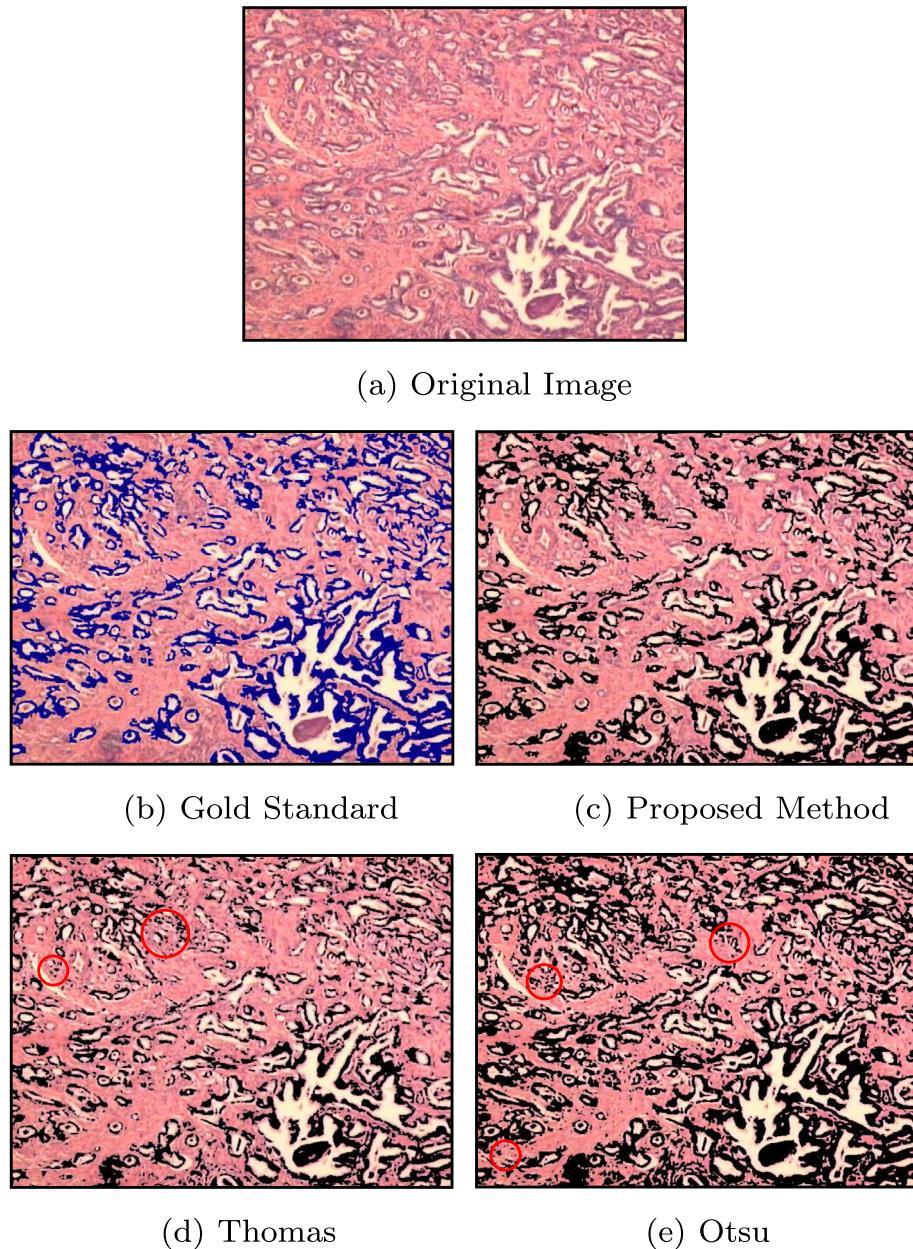


Fig. 10. Image of prostate sample diagnosed with cancer (H&E, 40×) (a); regions in blue manually selected by specialist (gold standard) (b); segmentation results achieved with proposed method (c), Thomas (d) and Otsu (e). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Percentage of true positives and false negatives.

| Class | True positives | | | False negatives | | |
|---------------------|----------------|-------------|-----------------|-----------------|-------------|-----------------|
| | Otsu | Thomas | Proposed method | Otsu | Thomas | Proposed method |
| Normal tissue | 0.98 ± 0.02 | 0.89 ± 0.05 | 0.95 ± 0.05 | 0.02 ± 0.02 | 0.11 ± 0.05 | 0.05 ± 0.05 |
| Hyperplastic tissue | 0.99 ± 0.01 | 0.81 ± 0.09 | 0.90 ± 0.10 | 0.01 ± 0.01 | 0.19 ± 0.09 | 0.10 ± 0.10 |
| Cancerous tissue | 0.91 ± 0.12 | 0.76 ± 0.14 | 0.83 ± 0.18 | 0.09 ± 0.12 | 0.24 ± 0.14 | 0.17 ± 0.18 |
| Mean total | 0.96 ± 0.05 | 0.82 ± 0.09 | 0.89 ± 0.11 | 0.04 ± 0.05 | 0.18 ± 0.09 | 0.11 ± 0.11 |

to be considered. Areas of false positives can remain in the image if the n value is too low (Fig. 6(a), red circles) and areas of true positives may be removed if the value is too high (Fig. 6(b), red circles). To solve this problem, the number of pixels for each cluster is determined and the mean is calculated. Only clusters of pixels with an area smaller than the mean are removed from the binary image.

2.5. Quantitative evaluation

To evaluate the proposed method, 30 prostate images were randomly chosen and first segmented by specialists (considered the gold standard). For each image segmented by the proposed method, the regions of interest were laid over the regions of interest in

Table 2

Percentage of true negatives and false positives.

| Class | True negatives | | | False positives | | |
|---------------------|----------------|-------------|-----------------|-----------------|-------------|-----------------|
| | Otsu | Thomas | Proposed method | Otsu | Thomas | Proposed method |
| Normal tissue | 0.87 ± 0.07 | 0.97 ± 0.02 | 0.97 ± 0.02 | 0.13 ± 0.07 | 0.03 ± 0.02 | 0.03 ± 0.02 |
| Hyperplastic tissue | 0.87 ± 0.11 | 0.97 ± 0.06 | 0.97 ± 0.07 | 0.13 ± 0.11 | 0.03 ± 0.06 | 0.03 ± 0.07 |
| Cancerous tissue | 0.90 ± 0.06 | 0.98 ± 0.02 | 0.99 ± 0.01 | 0.10 ± 0.06 | 0.02 ± 0.02 | 0.01 ± 0.01 |
| Mean total | 0.88 ± 0.08 | 0.97 ± 0.03 | 0.98 ± 0.03 | 0.12 ± 0.08 | 0.03 ± 0.03 | 0.02 ± 0.03 |

Table 3

Sensitivity and specificity of segmentation methods.

| Class | Sensitivity | | | Specificity | | |
|---------------------|-------------|-------------|-----------------|-------------|-------------|-----------------|
| | Otsu | Thomas | Proposed method | Otsu | Thomas | Proposed method |
| Normal tissue | 0.99 ± 0.03 | 0.89 ± 0.05 | 0.95 ± 0.05 | 0.87 ± 0.07 | 0.97 ± 0.02 | 0.97 ± 0.02 |
| Hyperplastic tissue | 0.99 ± 0.01 | 0.81 ± 0.09 | 0.90 ± 0.10 | 0.87 ± 0.11 | 0.97 ± 0.06 | 0.97 ± 0.07 |
| Cancerous tissue | 0.91 ± 0.12 | 0.76 ± 0.14 | 0.83 ± 0.18 | 0.90 ± 0.06 | 0.98 ± 0.02 | 0.99 ± 0.01 |
| Mean total | 0.96 ± 0.05 | 0.82 ± 0.09 | 0.90 ± 0.11 | 0.88 ± 0.08 | 0.97 ± 0.03 | 0.98 ± 0.03 |

Table 4

Accuracy of segmentation methods.

| Class | Accuracy | | |
|---------------------|-------------|-------------|-----------------|
| | Otsu | Thomas | Proposed method |
| Normal tissue | 0.89 ± 0.06 | 0.96 ± 0.02 | 0.97 ± 0.02 |
| Hyperplastic tissue | 0.89 ± 0.09 | 0.94 ± 0.04 | 0.96 ± 0.05 |
| Cancerous tissue | 0.91 ± 0.03 | 0.94 ± 0.03 | 0.96 ± 0.02 |
| Mean total | 0.90 ± 0.06 | 0.94 ± 0.03 | 0.97 ± 0.03 |

the corresponding image segmented by the specialists, which is a common procedure described in the specialized literature (Estrada & Jepson, 2009). The next step was to calculate sensitivity (S_E), specificity (S_P) and accuracy (A_{CC}) (Sonka, 2000). S_E measures the percentage of true positives that are correctly identified and is defined as:

$$S_E = \frac{T_p}{T_p + F_N} \quad (5)$$

in which T_p is the number of true positives (represented by correctly marked object pixels) (Fig. 7(b), T_p) and F_N is the number of false negatives (represented by incorrectly marked background pixels) (Fig. 7(b), F_N).

S_E alone is insufficient to assess segmentation quality, as it only considers the percentage of true positives identified. S_P is determined to quantify the percentage of correctly defined true negatives and is calculated as follows:

$$S_P = \frac{T_N}{T_N + F_p} \quad (6)$$

in which T_N is the number of true negative (represented by correctly marked background pixels) (Fig. 7(b), T_N) and F_p is the number of false positives (represented by incorrectly marked object pixels) (Fig. 7(b), F_p).

The strategy employed also allows calculating accuracy, which quantifies the proportion of correctly detected true positives and true negatives in relation to all positive (P) and all negatives (N):

$$A_{CC} = \frac{T_p + T_N}{P + N} \quad (7)$$

in which P and N are respectively the number of positive (Fig. 7(a), P) and negative (Fig. 7(a), N) pixels in the reference image.

The performance measures for the proposed method were compared to those calculated for the methods by Thomas (2010) and

Otsu (1979). It should be stressed that the method by Otsu (1979) is widely used for the comparison of histological component segmentation methods (Chen et al., 2011; Doyle et al., 2011; Kim et al., 2009). Using the algorithm by Otsu (1979), the optimum threshold value is calculated by the minimization of intra-class variance, but the method does not furnish consistent results when the histogram of the image is unimodal or nearly unimodal. This type of histogram is often seen in color channels of prostate CCN images.

3. Results and discussion

Figs. 8(a), 9(a) and 10(a) respectively display normal, hyperplastic and cancerous prostate images. Figs. 8(b), 9(b) and 10(b) respectively display the reference images manually segmented by a pathologist (gold standard). Figs. 8(c), 9(c) and 10(c) respectively display the results of the CNN segmentation by the proposed method. Figs. 8(d), 9(d) and 10(d) respectively display the results obtained using the method described by Thomas (2010). Figs. 8(e), 9(e) and 10(e) respectively display the results obtained using the method described by Otsu (1979).

In the results obtained with the proposed method for the separation of CCN regions in the images representative of normal, hyperplastic and cancerous groups (Figs. 8(c), 9(c) and 10(c), respectively), there was no significant over-segmentation or under-segmentation. Moreover, the proposed method successfully achieved segmentation results very similar to those of the gold standard (Figs. 8(b), 9(b) and 10(b)). In contrast, the Thomas and Otsu methods achieved poorer results than those obtained with the proposed method, as demonstrated in the images representative of the normal group (Fig. 8(a)), in which tissue structures and CCN do not exhibit alterations caused by disease, and in the hyperplastic (Fig. 9(a)) and cancerous (Fig. 10(a)) groups, in which CNN areas were not segmented correctly (regions marked by red circles in Fig. 8(d) and (e)).

The results achieved with the Thomas method (Figs. 8(d), 9(d) and 10(d)) exhibit under-segmentation, incorrectly defining the CCN regions. This under-segmentation can result in an incorrect analysis and compromise subsequent steps of the CAD. The Otsu method exhibited over-segmentation (Figs. 8(e), 9(e) and 10(e)), segmenting regions of nuclei of the stroma as CCN, which compromises the subsequent steps of the CAD, such as the extraction of characteristics and CCN pattern recognition.

The performance of the proposed method is also presented for the group of images that constituted the gold standard: 30 randomly chosen images segmented by specialists. Table 1 displays the quantitative results with the percentages of T_p and F_N . Table 2 displays the percentages of T_N and F_p . Tables 3 and 4 respectively display the measures of S_E , S_P and A_{CC} .

Table 3 shows that the proposed method achieved S_E and S_P values of 0.90 ± 0.11 and 0.98 ± 0.03 , respectively. This balance of values demonstrates appropriate CCN segmentation. With the Otsu method, the high S_E (0.96 ± 0.05) and low S_P (0.88 ± 0.08) confirm the over-segmentation observed in the qualitative evaluation. Moreover, the low S_E (0.82 ± 0.09) and high S_P (0.97 ± 0.03) achieved with the Thomas method confirm the under-segmentation observed in the qualitative evaluation.

With regard to ACC, the proposed method (0.97 ± 0.03) also performed better than the Thomas (0.94 ± 0.03) and Otsu (0.90 ± 0.06) methods (Table 4), demonstrating superior automatic CCN segmentation. In comparison to Otsu, the proposed method offers the advantage of defining the threshold limit with greater precision and detecting CCN regions without over-segmentation. The ACC results achieved with the proposed method are important to the diagnosis and prognosis of cancer as well as the extraction of characteristics and image classification methods.

4. Conclusion

This paper presented an unsupervised method for the automatic segmentation of cuboidal cell nuclei in images of prostate tissue stained with H&E at a magnification of $40\times$. The unsupervised method does not require the configuration of parameters or a priori knowledge on the input image. Moreover, the proposed method is effective for the segmentation of images of normal, hyperplastic and cancerous prostate tissue stained with H&E. In future studies, this method will be applied to images of prostate tissues with different degrees of magnification ($100\times$ and $400\times$) to determine the most appropriate magnification for the segmentation of regions of interest.

References

- Basavanhally, A., Ganesan, S., Feldman, M., Shih, N., Mies, C., Tomaszewski, J., et al. (2013). Multi-field-of-view framework for distinguishing tumor grade in ER+ breast cancer from entire histopathology slides. *IEEE Transactions on Biomedical Engineering*, 60(8), 2089–2099.
- Burger, W., & Burge, M. (2009). *Principles of digital image processing: Fundamental techniques*. Springer.
- Chen, C., Ozolek, J., Wang, W., & Rohde, G. (2011). A general system for automatic biomedical image segmentation using intensity neighborhoods. *Journal of Biomedical Imaging*, 2011, 8.
- de Arruda, P. F. F., Gatti, M., Junior, F. N. F., de Arruda, J. G. F., Moreira, R. D., Murta, L. O., et al. (2013). Quantification of fractal dimension and Shannon's entropy in histological diagnosis of prostate cancer. *BMC Clinical Pathology*, 13(1), 6.
- Dougherty, G. (2009). Digital image processing for medical applications. *Recherche*, 67, 02.
- Doyle, S., Agner, S., Madabhushi, A., Feldman, M., & Tomaszewski, J. (2008). Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In *Fifth IEEE international symposium on biomedical imaging: From nano to macro, ISBI 2008* (pp. 496–499). IEEE.
- Doyle, S., Feldman, M., Tomaszewski, J., Shih, N., & Madabhushi, A. (2011). Cascaded multi-class pairwise classifier (CascaMPa) for normal, cancerous, and cancer confounder classes in prostate histology. In *2011 IEEE international symposium on biomedical imaging: From nano to macro* (pp. 715–718). IEEE.
- Epstein, J., & Netto, G. (2007). *Biopsy interpretation of the prostate*. Lippincott Williams & Wilkins.
- Estrada, F., & Jepson, A. (2009). Benchmarking image segmentation algorithms. *International Journal of Computer Vision*, 85(2), 167–181.
- Faruquzzaman, A., Paiker, N. R., Arifat, J., & Ali, M. A. (2008). A survey report on image segmentation based on split and merge algorithm. *IETECH Journal of Advanced Computations*, 2(2), 86–101.
- Freixenet, J., Muñoz, X., Raba, D., Martí, J., & Cuffi, X. (2002). Yet another survey on image segmentation: Region and boundary information integration. In *Computer vision ECCV 2002* (pp. 408–422). Springer.
- Gleason, D. (1966). Classification of prostatic carcinomas. *Cancer Chemotherapy Reports, Part 1*, 50(3), 125.
- Gonzalez, R., Woods, R., & Eddins, S. (2003). *Digital image processing using MATLAB*. Pearson Prentice Hall.
- Guillaud, M., Cox, D., Malpica, A., Staerkel, G., Matisic, J., Niekrirk, D. V., et al. (2004). Quantitative histopathological analysis of cervical intra-epithelial neoplasia sections: Methodological issues. *Cellular Oncology*, 26(1–2), 31–44.
- Gurcan, M., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., & Yener, B. (2009). Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2, 147–171.
- Haralick, R., & Shapiro, L. (1992). *Computer and robot vision* (Vol. 1). Addison-Wesley.
- He, L., Long, L. R., Antani, S., & Thoma, G. R. (2012). Histology image analysis for carcinoma detection and grading. *Computer methods and programs in biomedicine*, 107(3), 538–556.
- Hugosson, J. (2000). Early diagnosis: State of the art in clinical routine and screening studies. *Renal, Bladder, Prostate and Testicular Cancer: An update* (pp. 91–95).
- Humphrey, P., & Andriole, G. (2010). Prostate cancer diagnosis. *Missouri Medicine*, 107(2), 107.
- INCA. (2011). Estimativa 2012: Incidencia de CGncer no Brazil, Tech. rep., Instituto Nacional de CGncer.
- Kayser, G., Riede, U., Werner, M., Hufnagl, P., & Kayser, K. (2002). Towards an automated morphological classification of histological images of common lung carcinomas. *Electronic Journal of Pathology and Histology*, 8(2), 22–30.
- Kim, S., Jeong, H., Choi, H., & Kim, D. (2009). Automatic histologic grading for lobular carcinoma in situ. In *World Congress on medical physics and biomedical engineering, September 7–12, 2009, Munich, Germany* (pp. 654–657). Springer.
- Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H., & Gurcan, M. N. (2008). A multi-resolution image analysis system for computer-assisted grading of neuroblastoma differentiation. In *Medical imaging: International Society for Optics and Photonics* (pp. 69151T–69151T).
- Li, B. N., Chui, C. K., Chang, S., & Ong, S. H. (2012). A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images. *Expert Systems with Applications*, 39(10), 9661–9668.
- Li, C., & Lee, C. (1993). Minimum cross entropy thresholding. *Pattern Recognition*, 26(4), 617–625.
- Mescher, A. (2010). *Junqueira's basic histology: Text & atlas*. McGraw-Hill Medical.
- Miedema, J., Marron, J. S., Niethammer, M., Borland, D., Woosley, J., Coposky, J., et al. (2012). Image and statistical analysis of melanocytic histology. *Histopathology*, 61(3), 436–444.
- Monaco, J. P., Tomaszewski, J. E., Feldman, M. D., Hagemann, I., Moradi, M., Mousavi, P., et al. (2010). High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. *Medical Image Analysis*, 14(4), 617.
- Osuna-Enciso, V., Cuevas, E., & Sossa, H. (2013). A comparison of nature inspired algorithms for multi-threshold image segmentation. *Expert Systems with Applications*, 40, 1213–1219.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1), 62–66.
- Pal, N., & Pal, S. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26(9), 1277–1294.
- Pratt, W. K. (2007). *Digital image processing: PIKS scientific inside*. Wiley-Interscience.
- Sahidan, S., Mashor, M., Wahab, A., Salleh, Z., & Jaafar, H. (2008). Local and global contrast stretching for color contrast enhancement on Zieli-Neelsen tissue section slide images. In *4th Kuala Lumpur international conference on biomedical engineering 2008* (pp. 583–586). Springer.
- Sanyal, N., Chatterjee, A., & Munshi, S. (2011). An adaptive bacterial foraging algorithm for fuzzy entropy based image segmentation. *Expert Systems with Applications*, 38(12), 15489–15498.
- Solomon, C., & Breckon, T. (2011). *Fundamentals of digital image processing: A practical approach with examples in Matlab*. Wiley.
- Sonka, M. (2000). *Handbook of medical imaging: Medical image processing and analysis* (Vol. 2). Society of Photo Optical.
- Taverna, G., Colombo, P., Grizzi, F., Franceschini, B., Ceva-Grimaldi, G., Seveso, M., et al. (2009). Fractal analysis of two-dimensional vascularity in primary prostate cancer and surrounding non-tumoral parenchyma. *Pathology – Research and Practice*, 205(7), 438–444.
- Thomas, K. A. (2010). Image processing as applied to medical diagnostics, Master's thesis, University of Oregon, Eugene, OR, United States.
- Tsai, D. (1995). A fast thresholding selection procedure for multimodal and unimodal histograms. *Pattern Recognition Letters*, 16(6), 653–666.
- Wu, Q., Merchant, F., & Castleman, K. (2010). *Microscope image processing*. Academic Press.
- Zhou, M., & Shah, R. (2011). *Prostate biopsy interpretation: An illustrated guide*. Springer Verlag.