International Journal of Engineering Research and Technology. ISSN 0974-3154 Volume 10, Number 1 (2017)
© International Research Publication House http://www.irphouse.com

Automatic Detection of Prostate Cancer by Computer Aided Diagnosis

Shashidhar Gurav
Assistant Professor Department of computer science and engineering
Sharad institute of Technology College of Engineering yadrav –ichalkaranji-416121
Dr Kshama Kulhalli
Vice-principal & HOD IT Department D.Y.Patil College of Engineering and Technology,Kolhapur-416006

Abstract:

Prostatic adenocarcinoma is the second most commonly occurring cancer among men, and it has also become major cause of death. In traditional prostate cancer diagnosis pathologist examines biopsies mainly based on cell morphology and tissue distribution. However, this is subjective and often leads to considerable variability.This paper presents a computational steps in automated prostate cancer diagnosis based on histopathology. Prostate tissue characteristics play an important role in processing the recurrence of prostate cancer. In step 1: Image preprocessing to determine focal areas, focal area determination is usually processed by noise reduction to improve its success; this step also comprises nucleus/cell segmentation. In step 2: Feature extraction to quantity the properties of the focal areas mainly automated cancer diagnosis relies on capturing deviation in the cell structure and the changes in the given tissue .The features are extracted to quantity these changes in a given tissue, and feature selection is mean to select the relevant important features. In step 3: Classifying the focal areas as malignant or not or identifying there malignancy levels. Automated diagnostic systems that operation quantitative measures are designed. After the design, this step also estimates the accuracy of the system.

***Keywords**: computer-aided detection, computer-aided diagnosis, prostate cancer,*

INTRODUCTION:

Treatment of the prostate cancer is effective only if it is detected at an early stage. Fortunately, the recent advances in medicine have significantly increased the possibility of curing prostate cancer. However, the chance of curing prostate cancer primarily relies on its early diagnosis and the selection of its treatment depends on its malignancy level. Therefore, it is critical for us to detect cancer, distinguish cancerous structures from the benign and healthy ones and identify its malignancy level. The digital image processing techniques are giving new insights to medical image processing research and other application domains in recent years. Pathologists make a diagnostic decision by viewing a specimen and measuring various diagnostically important attribute of isolation in images and it is a complex process. In recent years, computer-aided image processing has played a significant role in quantitative pathology. The specimens of biopsy was increasing because of the general trend in clinical diagnosis practice towards using smaller, less invasive sampling, often generating very small amounts of tissue. To overcome these problems and to meet the higher medical needs the image processing and image analysis techniques are used.. The tissue sample is removed from the body and then prepared for viewing under the microscope by placing it in a fixative, which stabilizes the tissue to prevent decay. For the sake of visualizing under the microscope, different components of the tissue are dyed with different stains. Then, different staining techniques are applied to reveal specific tissue components under the microscope.The pathologist plays a central role in therapeutic (healthful) decision making [1],[2].

Automated nuclei segmentation and classification is a recurring task, particularly difficult on pathology images. Indeed, the detection and segmentation of nuclei in cytopathology images are generally facilitated due to the well-separated nuclei and the absence of complicated tissue structures. In contrast, the segmentation of nuclei on histopathological images (tissue preserving its original structure) is more difficult since most of the nuclei are often part of histological structures presenting complex and irregular visual aspects. In this paper, we present a systematic survey on the computational steps to automatically diagnose prostate cancer by using histopathological images. In each step, we explain the techniques, address the challenges, and discuss the remedies offered by these techniques to overcome the challenges.

II - PREPROCESSING STEP

The main aim of the preprocessing step is to determine the focal areas in the image. Due to a considerable amount of noise that arises from the staining process, it is usually necessary to reduce the noise prior to the focal area identification**.**

Noise reduction and artifacts elimination can also be performed prior to detection and segmentation. Additionally, region of interest (ROI) detection can also be performed in order to reduce the processing time.

A. Noise reduction

The most trivial method for noise reduction is thresholding the pixels of an image, which usually follows background correction and filtering. Background correction standardizes the images by making use of an emptyimage. Thus, it lessens the effects of different image acquisition conditions such as different lighting conditions [23]. In filtering, the value of a pixel is transformed to a new value which is computed as a function of the values of pixels located in a selected neighborhood around this particular pixel. Filters especially reduce the random noise and, hence, improves the results of thresholding [24], [25], [26].

1) *Illumination Normalization:* The illumination can be corrected either by using white shading correction or by estimating the illumination pattern from a series of images. In white shading correction, a blank (empty) image is captured and used to correct images pixel by pixel [73].

2) A common equation is

$$\text{Transmittance} = \frac{(\text{White Reference value}) - (\text{Background value})}{(\text{Specimen value}) - (\text{Background value})}$$

A downside of this method is that a blank image must be acquired for each lens magnification whenever the microscope illumination settings are altered. An alternative normalization method is based upon the intrinsic properties of the image which are revealed through Gaussian .

STAINING AND IMAGE ODALITIES IN DIGITAL PATHOLOGY

Digital pathology is the microscopic investigation of a biopsy or surgical specimen that is chemically processed and sectioned onto glass slides to study cancer expression, genetic progression,and cellular morphology for cancer diagnosis and prognosis. For tissue components visualization under a microscope, the sections are dyed with one or more stains. H&E staining is a widespread staining protocol in pathology.H&E staining has been used by pathologists for over a hundred years [31] and is still widely used for observing morphological features of the tissue under white light microscopes. Hematoxylin stains nuclei in dark blue color, while eosin stains other structures (cytoplasm, stroma, etc.) with a pink color [see Fig. 1(a)]. Nuclei are susceptible to exhibit a wide variety of patterns (related to the distribution of chromatin, prominent nucleolus) that are diagnostically significant. IHC is a technique used or malignant and for determining the stage of a tumor. By revealing the presence or absence of specific proteins in the observed tissue origin of a tumor. According to the proteins revealed by IHC,specific therapeutic treatments adapted for this type of cancer are selected. Fig. 1(b) shows an example of IHC under light microscopy. After staining, fast slide scanners are used to generate digital images that contain relevant information about the specimen at a microscopic level. They embark one or multiple lenses to magnify the sample and capture digital images with a camera.

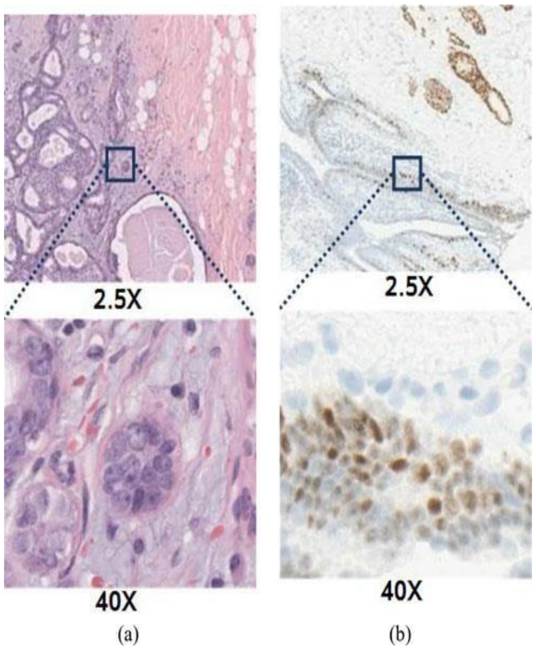

Fig:1: staining methods

B. Cell segmentation

The region-based approach is based on determining whether a pixel belongs to a cell or not, whereas the boundary based approach is based on finding the boundary points of a cell. The next subsections explain these approaches in detail.

1)* The region-based approach:* Thresholding is also used for the purpose of cell segmentation. Although thresholding separates the cells from the background, it does not separate the overlapping cells from each other. The mathematical morphology [28] and watershed algorithms [29] offer solution to this problem. Successive application of opening and closing operators is used to find the separate centers of gravity of the cells. Watershed algorithms are useful to detect the boundary lines between the touching cells [30], [31], [32]. An important limitation of the watershed algorithms is over-segmentation. Applying low-pass filters prior to segmentation and letting water rise only from the marked seeds lessens the over-segmentation problem. After thresholding the pixels, deleting small objects and large areas and filling small holes improve the segmentation results [31], [24]. For that, it is also possible to apply the morphological operators with different types of structuring elements such as square [27] and octagonal [25] ones. Here, selecting the size of the structuring element is important. The size should be smaller than the minimum size of a cell that will be determined; all the objects smaller than the structuring element are eliminated. On the other hand, it must be large enough to eliminate the noisy areas.

2) The boundary-based approach: The trivial method to determine the boundary points is manual segmentation [17], [18]. A number of points are taken from the user and a closed curve is approximated from these points. For example, Einstein et al. approximate a closed curve by joining the arcs characterized by every three successive points. This approach leads to successful segmentation since it requires extensive user interaction. However, its large scale use (for a large number of cells) is not feasible.

III.NUCLEI DETECTION, SEGMENTATION AND CLASSIFICATION

Nuclei detection and segmentation are important steps in cancer diagnosis and grading. The aspect of nuclei is critical for evaluating the existence of disease and its severity. For example,infiltration of LN in breast cancer is related to patient survival and outcome [31]. Similarly, nuclei pleomorphism has diagnostic value for cancer grading. Furthermore, mitotic count is also an important prognostic parameter in cancer grading [24]. Survey introduce the most commonly used image processing methods. Nuclei detectation and segmentation is carried out by image processing method and iterative segmentation approach.

A. Image Processing Methods

1)**Thresholding**: Thresholding is a method used for converting intensity image *I* into a binary image *I_* by assigning all pixels to the value one or zero if their intensity is above or below some threshold *T*. Threshold *T* can be global or local.

2) Morphology*:* Morphology is a set-theoretic approach that considers an image as the elements of a set and process images as geometrical shapes . The basic idea is to probe an ima with a simple, predefined shape, drawing conclusions on how this shape fits or misses the shapes in the image. This simple probe is called the structuring element and is a subset of the image. The typically used binary structuring elements are crosses, squares, and open disks.

The two basic morphological operators are the erosion _ and the dilation.

3) **Region Growing**: Region growing [28] is an image segmentation method consisting of two steps. The first step is the selection of seed points and the second step is a classification of neighboring pixels to determine whether those pixels should be added to the region or not by minimizing a cost function.

4)Active Contour Models and Level sets: Active contour models (ACMs) or deformable models, widely used in image segmentation, are deformable splines that can be used to depict the contour of objects in an image using gradient information by seeking to minimize an energy function . In case of nuclei segmentation, the contour points that yield the minimum energy level form the boundary of nuclei. The energy function is often defined to penalize discontinuity in the curve shape and graylevel discontinuity along the contour [12].

5) Probabilistic Models: Probabilistic models can be viewed as an extension of *K*-means clustering. Gaussian mixture models (GMMs) are a popular parametric probabilistic model represented as weighted sum of Gaussian cluster densities. The image is modeled according to the probability distribution.

B) Iterative segmentation approach

Our approach to the segmentation process begins by first localizing the nuclei of the DNA channel, which is the focus of this survey. The information obtained from this step is used as a prior localization tool for the cells in the other two channels.

Initial Segmentation. In MAGIC™ images are first segmented into small groups of contiguous pixels known In our application, the size of objects is controlled to be 10-20 pixels at the finest level. Based on this level, subsequent higher and coarser levels are built by forming larger objects from the smaller ones in the lower level. The component modules of the prostate tissue image analysis system are shown as a flow chart in Figure.

Background Extraction. The background extraction stage segments the TMA tissue core from the background (transparent region of the slide) using intensity threshold and convex hull.

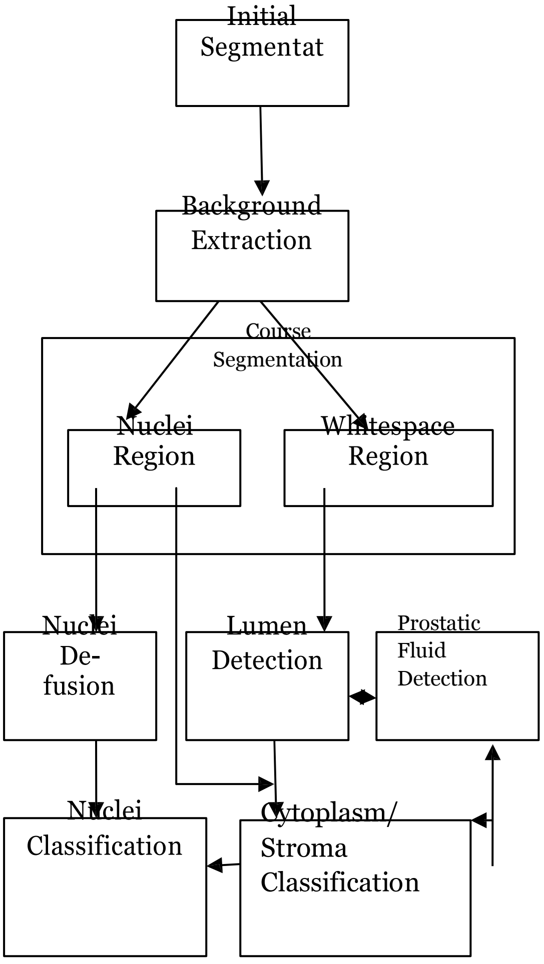

Fig:2 segmentation steps

Coarse Segmentation. The foreground (TMA core) is then re-segmented into rough regions corresponding to nuclei and white spaces. The main characterizing feature of nuclei in H&E stained images is that they are stained blue compared to the rest of the pathological objects.

White Space Classification. In the stage of coarse segmentation, the white space regions may correspond to both lumen (pathological object) and artifacts (broken tissue areas) in the image. The smaller white space objects (area less than 100 pixels) are usually artifacts. Area filter is applied to classify them as artifacts.

Nuclei De-fusion and Classification. In the stage of coarse segmentation, the nuclei area is often Obtained as contiguous fused regions that encompass several real nuclei. Moreover, the nuclei region might also include surrounding misclassified cytoplasm. These fused nuclei areas need to be de-fused in order to obtain individual nuclei.

Feature Extraction and Selection

Automated cancer diagnosis relies on capturing (i) the deviations in the cell structures (*cellular-level*), and (ii) the changes in the cell distribution across the tissue (*tissue-level*). The features are extracted to quantify these changes in a given tissue. To measure the deviations at the cellular-level, morphological, textural, fractal, and/or intensity-based features can be used. In extracting such kinds of the cells should be determined beforehand. On the other hand, to measure the changes at the tissue-level, textural, fractal, and/or topological features can be extracted. The use of these features at the tissue-level does not always necessitate cell segmentation (e.g., textural feature extraction at the tissue level). Even if the segmentation is necessary, determining the coarse locations of the cells is usually sufficient (e.g., topological feature extraction).

Morphological features

The morphological features provide information about the size and shape of a cell. The size is expressed by the radius, area, and perimeter of the cell. On the other hand, the shape is expressed by the compactness, roundness, smoothness, length of the major and minor axes, symmetry, concavity, and perimeter [20].

Suppose that $S=\{s_1 .s_2 .....s_n\}$ a set of the boundary points of a segmented cell/nucleus and *C* is the centroid of these boundary points; a sample of a nucleus with its boundary points and centroid is illustrated in Fig.3(a). The morphological features defined on the set of the boundary points, *S*, are given as follows:

^{2} ***Radius*** *r* is defined as the average length of the radial lines towards every boundary points. Mathematically

$$r = \frac{\sum_{i=1}^{n} |s_i C|}{n}$$

Area is the number of pixels within the boundary.
Compactness is the ratio of the square of the perimeter to the area
(*compactness* = *perimeter*^{2}/*area*).
Roundness is defined as *compactness*/4*π* [37].
Smoothness of the boundary is the sum of the smoothness of each boundary points. For a boundary point si, the smoothness is defined as the difference between the radial line [siC] and the average length of the radial lines surrounding si.
Major axis is the longest chord that goes through the center and *minor axis* is the line that is perpendicular to the major axis and that goes through the center.
symmetry *S* is quantified by measuring the length difference between the line segments in opposite directions that are defined to a boundary point and that are perpendicular to the major axis.
Concavity is quantified by drawing chords between non-adjacent boundary points and checking whether or not the boundary points lie inside these chords.

B. Textural features

Texture is a connected set of pixels that occurs repeatedly in an image. At the tissue level, textural features are computed for an entire image or its sub-images [18], [19], [50]. For a single image/sub-image or a cell, different textural features with different sets of parameters (e.g., with different values of *d* and *μ* in the case of the co-occurrence matrix) are first computed. Then, either a subset of features is selected or the same type of features are combined by using statistics measures (e.g., averaging, selecting the maximum, etc.).

CTopological features

The **topological** features provide information on the structure of a tissue by quantifying the spatial distribution of its cells. For that, this approach encodes the spatial interdependency of the cells prior to the feature extraction.

IV. Evaluation of the classification system

In general, a classification system should have two stages: (i) training the classifier to learn the system parameters and (ii) testing the system to evaluate the success of the classifier. Since there is a limited amount of available data in training, it is very important to test the system with extra data. However, it is an issue how to use this limited amount of data in both training and testing. For a given sample, a diagnostic system can lead to one of the four possible categories: True positive (TP):the diagnostic system yields positive test result for the sample and the sample actually has the disease,^{2} False positive (FP): the diagnostic the sample does not actually have the disease, True negative (TN): the diagnostic system yields negative test result for the sample and the sample does not actually have the disease, False negative (FN): the diagnostic system yields negative test result for the sample but the sample actually has the disease,By using the number of samples that fall into these categories, *sensitivity* and *specificity* are defined to assess the success of the diagnostic system. Sensitivity is the probability of a positive diagnosis test among persons that have
the disease and it is defined as,

$$\text{sensitivity} = \frac{\text{number of TP}}{\text{number of TP} + \text{number of FN}}$$

Specificity is the probability of a negative diagnosis test among persons that do not have the disease and it is defined as,

$$\text{specificity} = \frac{\text{number of TN}}{\text{number of TN} + \text{number of FP}}$$

V. CONCLUSION

In this review, we investigate the computational steps to automatically diagnose cancer by making use of histopathological images. There are primarily three steps: preprocessing, feature extraction, and diagnosis.In the preprocessing step, the focal areas determined. This comprises to eliminate the noise and improve the image quality. Although, different techniques such as filtering and mathematical transformations show different levels of success in noise reduction, the problem of noise elimination has not been entirely solved. In the case of diagnosis at the cellular-level, this step also includes cell segmentation. The segmentation is achieved either by finding their boundary points or by differentiating the pixels of cells. The important challenge in segmentation is to separate the touching cells from each other. Next is the feature extraction step. This step quantifies the properties of the biological structures of interest, extracting features either at the cellular-level or at the tissue-level. While cellular-level features focus on capturing the deviations in the cell structures, tissue-level features focus on capturing the changes in the cell distribution across the tissue. The features can be grouped into five based on the information they provide; the list of these features are given below. The important challenge in this step is to find the most proper cell/tissue representation(s) and select a subset of the features extracted from this representation(s).

[2] The **morphological** features provide information about the size and the shape of a nucleus/cell.The **textural** features provide information about the variation in the intensity of a surface and quantify properties such as smoothness, coarseness, and regularity. The **fractal-based** features provide information on the regularity and complexity of a cell/tissue by quantifying its self-similarity level. The **topological** features provide information on the cellular structure of a tissue by quantifying the spatial distribution of its cells. The **intensity-based** features provide information on the intensity (gray-level or color) histogram of the pixels located in a nucleus/cell.

After feature extraction, the next step is to distinguish benign and malignant structures as well as to classify the malignancy level. The important challenge is to evaluate the reliability of the designed diagnostic systems because of a limited amount of available data. This limited amount of data should be used both to learn the system parameters and to estimate the system reliability. The improper use of an evaluation method, however, may lead to biased and misleading results. For example, if the system performance evaluation is not done by using independent samples, overoptimistic results might be obtained. The numerical comparison of different studies is important to identify and avoid such biased and misleading results. For that, it is essential to form a benchmark of data sets that include biopsy samples taken from a large number of patients and examined by different pathologists.

REFERENCES

[1] American Cancer Society, “Prostate Cancer Overview,2015.http://www.cancer.org/acs/groups/cid/documents/webcont ent/003134-pdf.pdf

[2] P. A. Ganz, “Survivorship: Adult Cancer Survivors,” Primary Care, Vol. 36, No. 4, 2009, pp. 721-741.

[3] Cigdem Demir And B’Ulent Yener, “Automated Cancer Diagnosis Based On Histopathological Images: A Systematic Survey”, Technical Report, Rensselaer Polytechnic Institute, Department Of Computer Science, Tr-05-09. 1

[4] H. Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, M. Feldman, J. E. Tomaszewski, and A. Madabhushi, “Expectation maximization driven geodesic active contour with overlap resolution (emagacor): application to lymphocyte segmentation on breast cancer histopathology,” .

[7] Jean-Romain Dalle, Wee Kheng Leow, Daniel Racoceanu, Adina Eunice Tutac, Thomas C. Putti, “Automatic Breast Cancer Grading of Histopathological Images”, 30th A

[8] Baochuan Pang, Yi Zhang, Qianqing Chen, Zhifan Gao, Qinmu Peng, Xinge You,”Cell Nucleus Segmentation in Color Histopathological Imagery Using Convolutional Networks”, 978-1-4244- 7210-9/10, 2010 IEEE

[9] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, “Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology,” Histopathology Using Spectral Clustering with Textural and Architectural Image Features”, 978-1-4244-2003-2/08, 2008 IEEE.

[10] Metin N. Gurcan, Tony Pan, Hiro Shimada, and Joel Saltz, “Image Analysis for Neuroblastoma Classification: Segmentation of Cell Nuclei”, Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006

[11] Xiaobo Zhou, Fuhai Li, Jun Yan, and Stephen T. C. Wong “A Novel Cell Segmentation Method and Cell Phase Identification Using Markov Model”,

[12] Cigdem Gunduz-Demir, Melih Kandemir, Akif Burak Tosun, Cenk Sokmensuer, “Automatic

segmentation of colon glands using object-graphs", Medical Image Analysis 14(2010) 1-12

[13] Yousef Al-Kofahi, Wiem Lassoued, William Lee, and Badrinath Roysam "Improved Automatic Detection and Segmentation of Cell Nuclei in histopathology Images," IEEE Transactions On Biomedical Engineering, Vol. 57, No. 4, April 2010 Pp 841-850

[14] Maciej Hrebień, Jozef Korbicz, and Andrzej Obuchowicz" Hough Transform, Search Strategy and Watershed Algorithm in segmentation of Cytological Images":Computer Recognition Systems 2, ASC 45, pp. 550–557,Springer-Verlag Berlin Heidelberg, 2007

[15] Omar S. Al-Kadi, "Texture measures combination for improved meningioma classification of histopathological images", Pattern Recognition 43(2010) 2043-2053

[16] T.S. Subashini, V. Ramalingam, S. Palanivel, "Breast mass classification based on cytological patterns using RBFNN and SVM", Expert Systems with Applications,2008

[17] Olcay Sertel • Jun Kong • Umit V. Catalyurek • Gerard Lozanski • Joel H. Saltz • Metin N. Gurcan , J Sign Process Syst , "Histopathological Image Analysis Using Model-Based Intermediate Representations and Color Texture: Follicular Lymphoma Grading", DOI 10.1007/s11265-008-0201

[18] P. S. Umesh Adiga, B. B. Chaudhari, "Region based techniques for segmentation of volumetric histopathological images", Computer Methods & Programs in Biomedicine 61(2000) 23-47

[19] Pornchai Phukpattaranont and Pleumjit Boonyaphiphat, Non-members "Color Based Segmentation of Nuclear Stained Breast Cancer Cell Images", ECTI Transactions On Electrical Eng., Electronics and Communications Vol.5, No.2 August 2007 [22] rishnan,.Mousumi Pal,

[20]R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2013,"CA: A CancerJournal for Clinicians, vol. 63, pp. 11–30, 2013.

[21]S. Naik, S. Doyle, A. Madabhushi, J. E. Tomaszewski, and M. D. Feldman, "Auto-mated gland segmentation and Gleason grading of prostate histology by integratinglow-, high-level and domain specific information,"MIAAB workshop, Piscataway, NJ 2007.

[21] J. Monaco, et al., "High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models,"Medical Image Analysis vol. 14, pp. 617–629, 2010.

[22] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classifica-tion," IEEE Trans. Syst. Man Cybern. SMC, vol. 3, pp. 610–621, 1973.

[23]"Prostate cancer world map,"http://globocan.iarc.fr/factsheets/cancers/prostate.

[24]"The prostate in male body,"http://byhealth.com/prostate-cancer

[25] N. Howlader, et al., "Seer cancer statistics review, 1975-2009 (vintage 2009 popula-tions)," Bethesda, MD,http://seer.cancer.gov/csr/1975_2009_pops09 based on November 2011 SEER data submission, posted to the SEER web site, April 2012.

[26] "Prostate needle biopsy, ultrasound,CT scan, MRI scan,"http://ayie-disease.blogspot.com/2011/01/prostate cancer-pictures-slideshow.html

[27] "Microtome machine," http://pathology.utscavma.org/?page_id=3

[28] "Glass slide with tissue,"http://histology-slides database.blogspot.com/ 2011/08/glass-slide-with-tissue-piece.html.

[29] "Stained Tissue slide,"http://www5.pbrc.hawaii.edu/ccrt/taras/site/histology.html

[30] D. Gleason, "Histologic grading and clinical staging of prostatic carcinoma," Urologic Pathology: The Prostate, M. Tannenbaum, Ed. Philadelphia, PA: Lea and Febiger, pp.171–198, 1977.173

[31]"Panoptiq tissue slide scanner,"http://innovative-imaging.net/panoptiq_ panorama.html.

[32] "Leica scn400 tissue slide scanner,"http://www.leicabiosystems.com/products digital pathology/scan/details/product/leica-scn400-2/.