Automated Analysis of PIN-4 Stained Prostate Needle Biopsies

Bikash Sabata, Boris Babenko, Robert Monroe, and Chukka Srinivas

BioImagene Inc., 919 Hermosa Court, Sunnyvale, CA 94085
Bikash.Sabata@bioimagene.com
http://www.bioimagene.com

Abstract. Prostate Needle biopsies are stained with the PIN-4 marker cocktail to help the pathologist distinguish between HGPIN and adenocarcinoma. The correct interpretation of multiple IHC markers can be challenging. Therefore we propose the use of computer aided diagnosis algorithms for the identification and classification of glands in a whole slide image of prostate needle biopsy. The paper presents the different issues related to the automated analysis of prostate needle biopsies and the approach taken by BioImagene in its first generation algorithms.

Keywords: Computer Aided Diagnostics (CAD), Prostate Analysis, Medical Imaging, Histopathology Image Analysis.

1 Introduction

Several immunohistochemistry (IHC) markers are routinely used by pathologists in the interpretation of prostate biopsies, including P504S (racemace), p63, and high molecular weight (HMW) cytokeratins (CK5 and CK14) [1]. P504S is a protein preferentially expressed in the cytoplasm of prostatic adenocarcinoma as well as high-grade prostatic intraepithelial neoplasia (HGPIN). p63 and the HMW cytokeratins are expressed in the nucleus and cytoplasm respectively of prostatic basal cells surrounding benign prostatic glands, but not in the secretory cells of these glands [2]. The combination of these markers in the PIN-4 antibody cocktail (Biocare) is useful to the pathologist in the distinction between adenocarcinoma, HGPIN, and benign glands, particularly in cases with limited tissue [3]. However, correct interpretation of multiple IHC markers staining different subcellular compartments of different cell types can be challenging. Computer aided image analysis (CAD) algorithms are therefore required to assist the pathologist in the interpretation of prostatic tissue stained with the PIN-4 cocktail.

The workflow within the clinical labs is optimized to maximize the number of cases a pathologist can sign out without compromising the quality of diagnosis. Digital pathology promises to create the transformation to pathology practice that increases the overall quality and quantity of pathology diagnosis. The pathology slides are scanned using the whole slide scanner such as the iScan device from BioImagene. The images generated are managed within a workflow

A. Madabhushi et al. (Eds.): Prostate Cancer Imaging 2010, LNCS 6367, pp. 89–100, 2010.

software such as Virtuoso. The pathologist reviews the slides and selects regions in the images for analysis. The CAD algorithms for interpretation of PIN-4 marker cocktail in prostate glands consist of two steps. In the first step we detect and identify the glands within the needle biopsy and in the second step we classify each of the detected glands.

The paper describes the challenges associated with the segmentation, identification and classification of prostate glands in PIN-4 stained needle biopsies. We also present some experiments conducted to compare the CAD classification with the manual classification on whole slide images.

2 Approach

Whole slide images generated for Digital Pathology are typically Gigapixel images that require a scalable computational infrastructure to support the data volume. We have developed a pathology image analysis platform called *iAnalytics* that is able to handle such data volumes. In addition, the *iAnalytics* framework has a layered component based architecture that allows for the rapid development of scalable algorithms for pathology CAD. The system is developed in C++ and has interfaces to other languages such as Java, C#, Python and Matlab.

The components in *iAnalytics* are arranged in 3 layers (figure 1). At the bottom layer, interfaces to external low level imaging libraries are maintained through wrappers that allow these libraries to be plugged in or replaced. We have specifically integrated with the Intel IPP, Intel MKL, FreeImage [4] and OpenCV [5] libraries. The Imaging layer above this consists of the basic classes

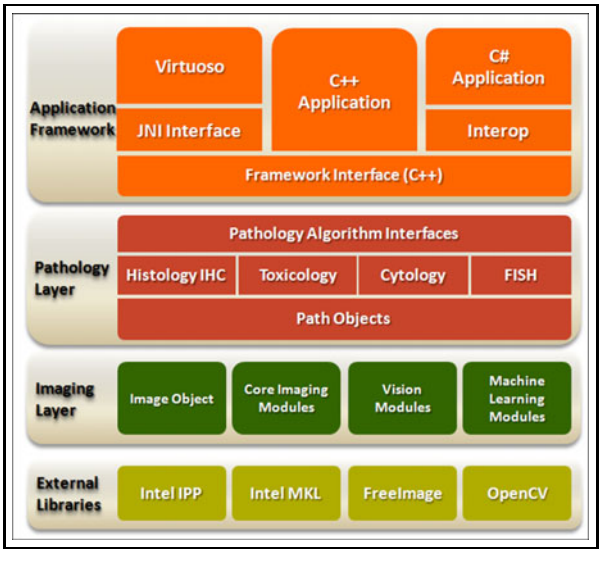

Fig. 1. iAnalytics Component Based Layered Architecture

needed for creating imaging objects and processing them. Image processing functions that are available in the off-the shelf commercial and open source systems are used with wrapper classes. In addition, we have implemented a large collection of basic imaging modules that include color processing, texture processing, and morphological operations. Vision modules that compute different types of features are also implemented in this layer. Most of the vision modules are high performance implementations of well known algorithms. Proprietary techniques have also been developed and included in the framework for the pathology domain that specialize well known techniques to the domain. Finally, classes implementing different types of Machine Learning techniques are included in the Imaging layer. The topmost layer of iAnalytics implements the components associated with the pathology CAD algorithms. High level pathology objects such as cells, nuclei, membrane, tissue regions, and glands are implemented as compositions of the lower layer objects. The pathology objects are combined together to define a CAD algorithm. All algorithms are accessed through a common interface. New algorithms that are developed can easily be included within existing applications that implement the interface framework. The *iAnalytics* component architecture allows the development of new algorithms as a composition of existing components.

2.1 Workflows

The automated analysis of PIN-4 stained prostate needle biopsies is supported within two different types of workflows. In the first workflow, the pathologist follows the steps enumerated below:

1. Select the case to be reviewed
2. Select the PIN-4 slide within the case to be reviewed
3. Review the virtual whole slide image at different resolutions and identify potential regions of tumor
4. Select the region encompassing the tumor using the FOV (Field of View) selection tool
5. Invoke the analysis algorithm on the FOV
6. The analysis algorithm automatically segments and identifies the individual glands within the FOV
7. The algorithm then classifies the individual glands
8. Finally, the algorithm in addition to the gland class, reports on some of the measurements on the gland (tumor area, median intensity).

In the above workflow the identification of the region is carried out by the pathologist. The onus of identifying all regions with tumor and carrying out the analysis is left to the pathologist. The second workflow is more complex but makes the job of the pathologist easier. The algorithm analyses the whole slide image and identifies all glands in the biopsy and classifies them. The identified glands are then presented to the pathologist in a sorted order that is diagnostically relevant. The steps of the workflow are

1. Select the case to be reviewed
2. Select the PIN-4 slide to be reviewed
3. The system does a whole slide analysis of the PIN-4 slide.
4. The system presents all the glands that have been identified and in an order that allows the pathologist to review all glands very efficiently
5. The system presents the gland class during review and also presents additional measurements on the glands and tumor.

3 Segmentation, Identification and Classification

The key phases of the analysis in Prostate PIN-4 CAD algorithm are

1. Segment the glands from the other stromal areas
2. Identify glands such that care is taken to make sure glands from two different classes are not merged and identified as a single gland
3. Classify the gland as adenocarcinoma, HGPIN, and Benign

3.1 Gland Segmentation

Segmenting glands out of microscopy/biopsy images is a challenging problem for several reasons. Segmentation of the glands when the glands are well formed is relatively easy. However, the regions of interest are cancerous tumor areas. Such areas rarely have well formed and separated glands. Glands come in a wide variety of irregular shapes and sizes. Furthermore, the images are typically full of distracting structures and have varying background that can result in poor segmentation. Sometimes the manual identification of individual glands is also challenging. We explored two different segmentation approaches:

1. Unsupervised Color Segmentation
2. Learning based Classifier

Traditional segmentation algorithms are fully unsupervised and are therefore in general unable to classify a segmented region as being a gland or other tissue. It has been observed that segmentation performance is very poor if higher level information is not taken into account. On the other hand, statistical learning methods from the object detection literature, such as the Viola-Jones face detector [6], cannot be readily applied to this problem because the object of interest are deformable and come in a wide variety of shapes and sizes. Nevertheless, we would like to take advantage of the vast advances made in this field in recent years (e.g. pedestrian detection accuracy has improved dramatically over the last decade [7]).

Unsupervised Color Segmentation. Our approach to segmentation and identification follows an iterative multiphase process. In the first phase, we segment and extract candidate gland regions using just the color staining. The brown regions are typically the basal layer cells. However, regions with non-specific brown staining can also occur. The deep blue regions are the epithelial

cells within the glands but also the stromal cells nuclei. The red regions appear within the cytoplasm of the epithelial cells of the glands. Non-specific red staining also appears in some areas of the biopsy. The light regions within the tissue regions are typically are the candidate lumen regions of the gland.

Color segmentation is done in a CIE L*a*b space. A constrained K-means clustering technique in the three dimensional color space is used to segment the image into clusters that broadly correspond to *Brown*, *Blue*, *Red*, and *Light* colored regions. Initial three-dimensional color vector seed values for *Brown, Blue, Red* and *Light* are specified. The convergent color thresholds are constrained to be around the specified initial cluster seeds. We determine the seed color vectors through a calibration step during the training phase of the algorithm. Sample images with known ground truth are used to determine the seed color vectors of the *Red, Brown, Blue* and *Light* clusters. This model is fairly robust within a single lab setting because of the operations followed within the lab quality process. Color segmentation is followed with a connected component blob analysis to extract multiple connected regions. One of the weaknesses of pixel-level clustering methods is that the context and any top-level information about the shape priors and size are ignored in the segmentation process.

Learning based Classifier. Our learning classifier is motivated by the advances in object detection. We turned to a recently published state-of-the-art pedestrian system [7], and applied a similar method for gland segmentation. At a high level, our learning approach can be summarized as follows. We begin by sliding an image patch classifier densely over the input image. In the first step, the patch classifier outputs the likelihood that the pixel in the center of a given patch belongs to a gland (see figure 2). In the next step, the resulting classifier response map is passed into an efficient segmentation algorithm.

The first component of our system involves an image patch classifier. To train this classifier, we labeled 8 training images by hand. Labeling was done by outlining all glands in these images. From this, we extracted 5,000 positive image patches (i.e. patches extracted from the gland regions), and 15,000 negative image patches (i.e. patches extracted from the non-gland regions). The patch size was fixed to 61x61. The classifier we trained was AdaBoost with decision stumps

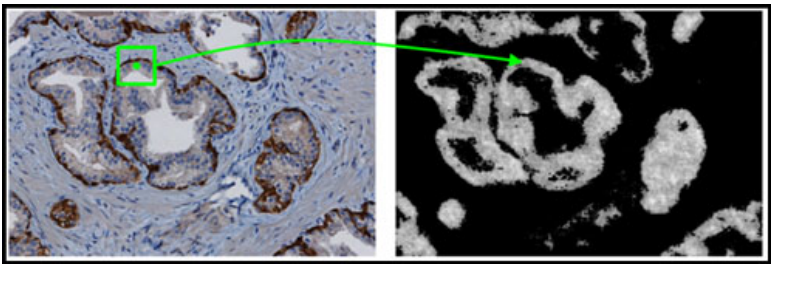

Fig. 2. A sliding window classifier is used to predict the likelihood of belonging to a gland for each pixel. Left image is the input image, and the right image is the resulting response map

as the weak classifiers. The features used were Haar-like features computed over the following channels: LUV color channels, gradient magnitude, histogram of gradient orientation channels, and a local standard deviation channel. For more details we refer the reader to [7].

Graph-Based Segmentation. The result of applying the trained classifier to a novel input image is a response map where for each pixel we have a likelihood that that pixel belongs to a gland (e.g. the right image figure 2). In the second phase of our system, we pass this response image into an efficient graph based segmentation algorithm described in [8]. Note that segmenting the response image is much easier than segmenting the original input image. Furthermore, the class (i.e. gland or non-gland) of each segment can subsequently be determined using the response image.

Qualitative results on novel test images are shown in figure 3. The trained classifier works well on picking up gland regions even though there is a high variability of appearances (colors and textures). Notice that, unlike pixel color clustering approach, this approach does not rely on a particular stain or color - glands that are not are surrounded by brown staining are picked up just as well as glands that are. One of the challenges we need to address is that the method sometimes merges two glands together incorrectly.

3.2 Gland Identification and Classification

In this step, the objective is to either to merge or split the regions output by the segmentation process into gland objects.

Adjacency analysis: The non-connected candidate clusters are ranked by its probability to be part of a gland. Candidate regions are analyzed within the context of the adjacent regions. Regions are combined by associating the candidate regions with a gland using the geometric relations between the different parts of the gland. The step results in candidate glands.

Merging and Splitting: Candidate glands are split into two or merged into a single gland. The splitting and merging criterion is based on identification of the basal layer and intervening stromal regions. Further, we do a risk based analysis for the sensitivity of splitting or merging. If the class of two candidate regions are the same then merging them carries no risk. Similarly if splitting does not change the class of either one of the candidate regions then there is no risk associated with the split. For merges and splits where the classification changes, additional analysis is done to improve the probability of the gland identification. This is done by the improved detection of stromal and basal regions.

Classification: Once the identification is completed the classification of the glands is not too complex. The classification criterion is

1. If gland has only the brown basal staining then the tumor is benign
2. If gland has both the red Racemace and the brown basal staining then it is classified as HGPIN
3. If gland has only the red Racemace then it is classified as adinocarcinoma.

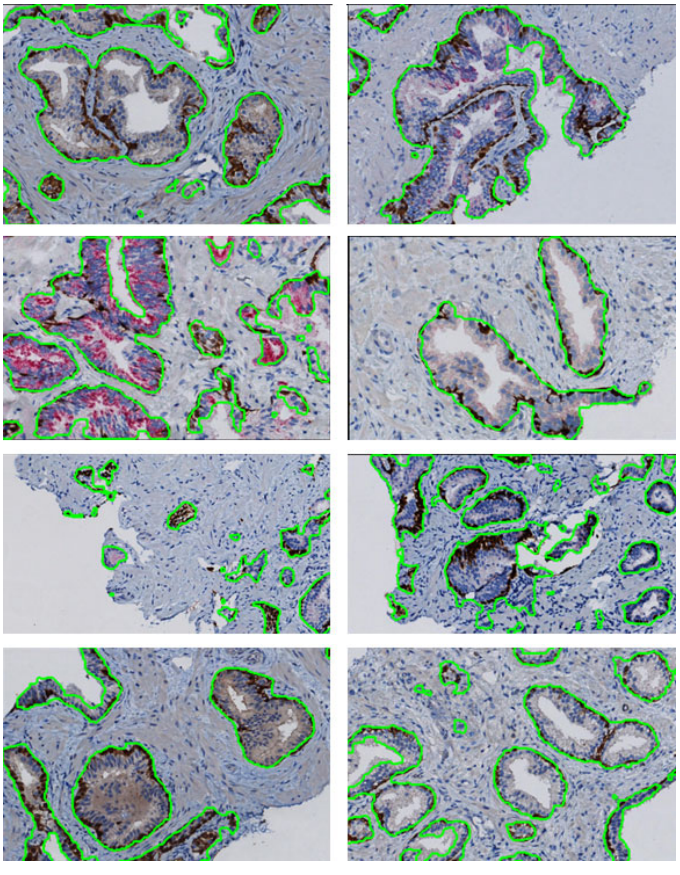

Fig. 3. Above Figure shows example results of our system on test (novel) images. See text for details.

4 Experiments

We have conducted some experiments with pathologists in the loop to evaluate the effectiveness of our automated CAD algorithm for Prostate PIN4 analysis. Fifty formalin-fixed, paraffin-embedded prostate biopsy cases, each consisting of corresponding H&E and PIN-4 stained slides were selected. These cases were part of a routine workflow in a pathology practice. This ensures that there is no selection bias for the study. DAB chromagen was used to visualize the p63 and HMW cytokeratin antibodies, and AEC chromagen was used to visualize the P504S antibody. Slides were scanned at 20x magnification on the BioImagene iScan Slide Scanner. Manual interpretation (manual digital read) was performed on a computer monitor that allowed the pathologist to view whole slide images at magnifications from 1x to 40x. After a one-week wash-out period, the same cases were reviewed using the PIN-4 image analysis algorithm and BioImagene

Virtuoso software for selected regions of interest. For manual scoring, cases were categorized as benign, HGPIN, atypical/ASAP or adenocarcinoma. For automated analysis, the algorithm categorized cases as benign, HGPIN or adenocarcinoma. Some of the representative samples are presented in figures 4-6 below. We present samples from each one of the classes. Note that the H&E image and the IHC image are not of the same tissue section but are from the same tissue block (serial sections) therefore there is a gross gland level correspondence but no cell level correspondence.

In Figure 4 we present the case of a benign tumor. The pathologist examines the H&E image and suspects a cancer tumor based on the gland morphology and structural arrangements. The PIN-4 test is ordered that is carried out on a serial section from the same block from which the H&E was generated. In 4.b the algorithm has segmented and identified the individual glands. All glands have been classified as benign (therefore outlined with green). Figure 5 presents the case of a high grade prostatic intraepithelial neoplasia (HGPIN) which are suspected to be the precursors to adenocarcinoma. The automated algorithm identifies the glands and correctly classifies them as HGPIN (therefore marking them with a yellow outline). Finally, Figure 6 presents the case of an adenocarcinoma. The CAD algorithm has correctly segmented and identified the individual glands. The classification is that of adenocarcinoma (therefore outlined with red). Note

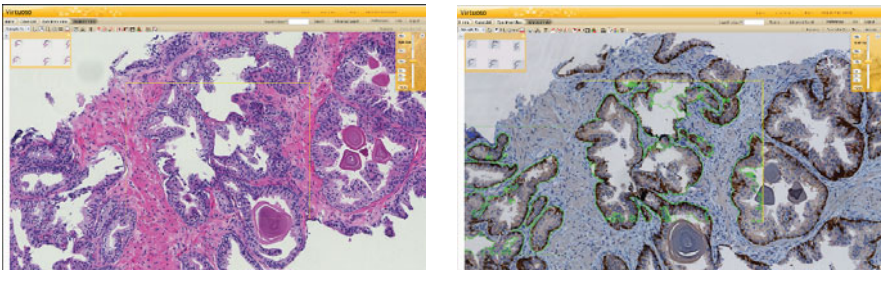

Fig. 4. Examples of benign prostate biopsies stained with H&E with corresponding area in PIN-4 IHC. The benign glands are outlined in green in the PIN-4 stained image as a result of the CAD analysis.

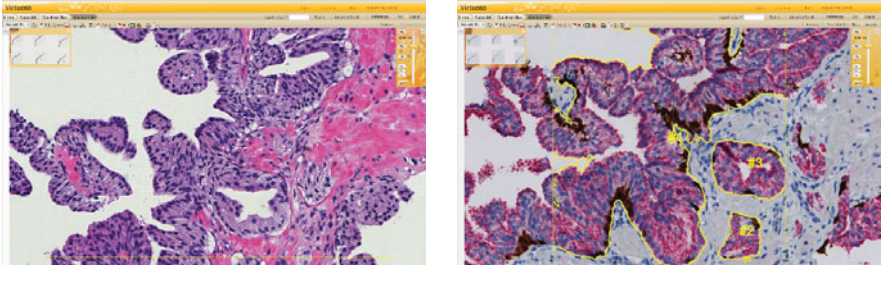

Fig. 5. Example of HGPIN stained with H&E and corresponding PIN-4 IHC. The HGPIN glands detected and classified by the CAD algorithm are outlined in yellow.

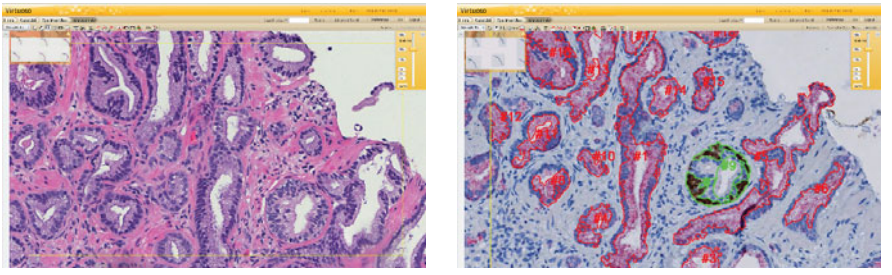

Fig. 6. Example of adenocarcinoma stained with H&E and corresponding PIN-4 IHC. The adenocarcinoma glands detected and classified by the CAD algorithm are outlined in red.

that there is one gland in the mix that is actually benign. It is important for the CAD algorithm to not merge this gland with the surrounding glands or the diagnosis will be incorrect.

5 Results

We present the results of our experiments that compare the final automated classification of the whole slide v/s the manual classification. The comparative results are based only on gland segmentation produced by the unsupervised color segmentation followed with gland identification and classification. In a future work, we plan to do a similar experiment using the gland segmentation results generated from the image patch classifier approach.

Although comparison of only the final diagnostic results hides a large number of misclassifications, mis-identifications and erroneous segmentation, this is the metric that is of final relevance to the final practice of pathology. We are in the process of developing a large ground truth dataset that includes the manual segmentation and gland identification for the whole slide images. With such datasets we will be able to provide quantitative accuracy results of the gland identification and segmentation procedures.

The first set of tabulated results (Table 1) shows comparison of manual read v/s automated scoring of benign cases and abnormal cases. Benign cases include cases that are classified as Benign and HGPIN. Abnormal cases are adenocarcinoma and atypical/ASAP. Note that the automated algorithm did not classify ASAPs but the manual reads did use the ASAP category. As the table indicates, there was only one case where the manual classification was Abnormal while the CAD analysis algorithm reported Benign. This is the case of ASAP that was not recognized by the CAD algorithm (figure 7). In the future we plan to include another class for ASAP. We have not included ASAP in the current experiment as they are relatively rare and difficult to get sufficient training samples. As seen in the result below we are seeing a concordance of 98% between the manual and the automated classifications.

Table 1. Manual digital read vs. CAD analysis for interpretation of benign versus abnormal PIN-4 IHC staining. Benign classification includes benign and HGPIN cases. Abnormal classification includes atypical/ASAP and adenocarcinoma cases. Concordance = 98%.

		Image Analysis	
		Benign	Abnormal
Manual Digital Read	Benign	29	0
	Abnormal	1	20

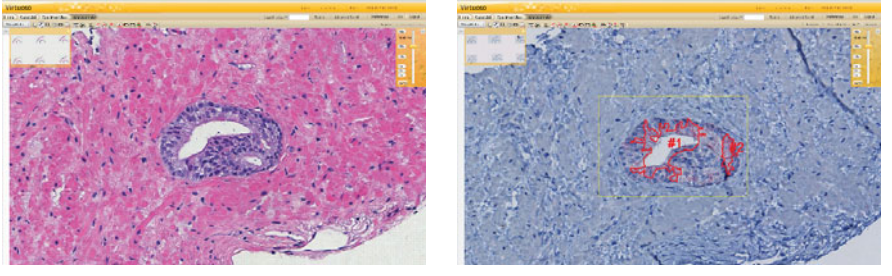

Fig. 7. Example of an atypical focus (ASAP) in a prostate biopsy stained with H&E and the corresponding detection in PIN4. The analysis result was adenocarcinoma.

Table 2. Manual digital read vs. CAD for interpretation of benign versus benign/HGPIN PIN-4 IHC staining. Benign classification includes benign cases while benign/HGPIN classification includes benign cases with HGPIN. Concordance = 90%.

		Image Analysis	
		Benign	Benign+HGPIN
Manual Digital Read	Benign	16	1
	Benign+HGPIN	2	10

In table 2 below we present the results of the comparison of manual read v/s automated scoring of Benign cases and HGPIN cases. The concordance between the manual read and the automated CAD is lower in this case. We have a 90% concordance. The reason is that the racemace staining is difficult to use and therefore there is a big variation in the staining intensity. When the glands are small, as in the case of HGPIN, the staining variation causes the recognition of the red staining to be error prone.

In summary:

- There is close agreement between manual digital reading and image analysis for interpretation of PIN-4 IHC staining
- CAD image analysis correctly categorizes glands into benign, HGPIN and malignant categories in most cases

- There is a 98% concordance between manual and automated classification for benign (benign and HGPIN) vs. abnormal diagnoses (adenocarcinoma and atypical) (refer to Table 1)
- There is a 90% concordance between manual and automated classification for benign vs. benign/HGPIN (benign and HGPIN) diagnoses (refer to Table 2)
- The CAD image analysis algorithm does not currently have an atypical category for small numbers of glands lacking basal cell staining. This explains the handful of cases where the manual interpretation was benign or atypical, but the IA interpretation was adenocarcinoma.

6 Discussions

Image analysis based CAD is a useful adjunctive tool to aid the pathologist in the interpretation of PIN-4 IHC studies. The PIN-4 algorithm can identify glands in three categories, and there is a high degree of concordance between manual interpretation and automated image analysis. To our knowledge, this is the first example of an algorithm to incorporate three color image analysis of an IHC cocktail.

Planned future versions of the PIN-4 algorithm will include an atypical/ASAP category, perform more accurate analysis based on integrating with H&E and calibration slides.

As digital pathology is increasingly adopted within the pathology practice, CAD algorithms for assisting the pathologist will enable a fundamental change to the practice of pathology. In the early phases of the adoption we observe that only a limited number of CAD algorithms provide the pathologist the additional information that helps them with certainty of diagnosis. An area of increasing utility is the optimization of workflow through the use of CAD algorithms. If the system is able to preprocess the images for the pathologist and presents the information in the order that is diagnostically relevant, the time to derive an accurate and correct diagnosis will be significantly reduced. Further, the directed review of the pathology slides will result in lower fatigue which will improve the quality of the pathology practice.

The PIN-4 algorithm developed by BioImagene is one such algorithm that has been adopted by the practicing pathologist to help in the diagnosis of adenocarcinoma in prostate needle biopsies. The usefulness of PIN-4 is still being studied and a final equivocal decision has not been arrived at in the community. However, we feel that the use of PIN-4 in the preprocessing of the whole slide image has the potential to dramatically alter the protocol of prostate screening process. The discussion of the usefulness of PIN-4 at that time will take a new dimension that will have to take into account the automated screening aspect of the test.

References

1. Weng Leong Ng, V., et al.: Is Triple Immunostaining With 34E12, p63, and Racemase in Prostate Cancer Advantageous? Am. J. Clin. Pathol. 127, 248–253 (2007)
2. Ayala, A.G., et al.: Prostatic Intraepithelial Neoplasia: Recent Advances. Arch. Pathol. Lab. Med. 131(8), 1257–1266

3. Molinie, V., et al.: Diagnostic utility of a p63/alpha-methyl-CoA-racemase (p504s) cocktail in atypical foci in the prostate. Mod. Pathol. 17(10), 1180–1190 (2004)
4. Freeimage Open Source Project: `http://freeimage.sourceforge.net/`
5. OpenCV Library Open Source Project: `http://sourceforge.net/projects/opencvlibrary/`
6. Viola, P., Jones, M.: Robust Realtime Face Recognition. International Journal of Computer Vision 57(2), 137–154 (2004)
7. P. Dollar, Z. Tu, P. Perona and S. Belongie: Integral Channel Features. In: BMVC 2009 (2009)
8. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient Graph-Based Image Segmentation. International Journal on Computer Vision 59(2), 167–181 (2004)