Selective Scanning for faster Prostate Pathology

Mayank Kabra
ECE, U. C. San Diego
mkabra@ucsd.edu

Yoav Freund
CSE, U. C. San Diego
yfreund@ucsd.edu

Steve Baird
School of Medicine, U. C. San Deigo
sbaird@ucsd.edu

Abstract

Whole slide scanning microscopes are now commercially available. These microscopes combine a digital microscope with a scanning mechanism to produce high magnification and high resolution images of complete specimens. These digital images have the potential of revolutionizing histopathology. For example, instead of sending glass slides through the mail, pathologists at different locations can consult each other by sharing image files through the Internet.

One of the main obstacles on the way to realizing this possibility is the very large size of the image files that are generated. A single image scanned at 40× resolution requires around 20GB before compression and around 1GB after compression using *jpeg2000*. As a result, a terabyte hard disk can store only a thousand or so images and sharing such large files over the Internet can result large delays that would make collaborative consultation impractical.

Pathologists, however, hardly ever view the whole image at high resolution but concentrate on few relevant parts. Relevant parts, if stored at higher resolution can save space, and if transmitted earlier over network can reduce delay.

We use a combination of computer vision and machine learning techniques to analyze low magnification images and select those parts of the specimen that should be scanned at high magnification. This is a similar problem to that of detecting cancer, however, the required accuracy is significantly lower. Instead of using computer vision to produce a diagnosis we use computer vision to eliminate those parts of the slide which are unlikely to hold useful information for human diagnostician.

Examples of prostate tissue sections are shown in the figure 1. Areas of interest marked by an experienced pathologist at low resolution are used as training set for learning. Using these markings, a classifier is built using AdaBoost, which classifies each pixel by using a histogram of blue and red chroma values in a patch surrounding it. We choose histograms as different regions have a specific mix of stroma(pink), nuclei(blue) and cavities(white) which is explicit in histograms.

The classifier built using histogram is able to separate connective tissue and healthy glands from cancerous regions. Using the classifier, around 80-90% of image can be discarded while still keeping all the parts relevant for diagnosis. This decreases the space required for storage and time for transmission by a factor of ten.

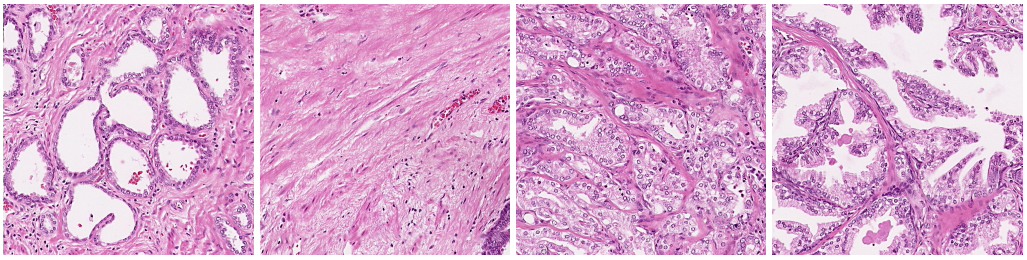

Figure 1: **a**: Glands in a prostrate tissue have glandular cells (blue) neatly arranged on the edge of gland cavity. **b**: Connective tissue are featureless and have very few nuclei. **c**: Region with prostate cancer. Cancerous glandular cells divide faster, destroy the structure of glands and invade the cavity. **d**: Hyperplasia, where normal growth in glandular cells leads to in-folding while still maintaining the gland structure. Hyperplasia is benign and also needs to be separated from cancer.