ARTICLE

Analysis of colonic histopathological images using pixel intensities and Hough Transform

Laurence A. Gan Lim[1,2,*], Raouf N.G. Naguib[1], Elmer P. Dadios[3], Jose Maria C. Avila[4]

[1] Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Coventry University, Coventry, U.K.

[2] Mechanical Engineering Department, De La Salle University, Manila, Philippines

[3] Manufacturing Engineering and Management Department, De La Salle University, Manila, Philippines

[4] Department of Pathology, University of the Philippines, Manila, Philippines

Cancer of the colon is among the leading types of cancer. Worldwide, colorectal cancer is considered the third most common neoplasm. Similar to other types of cancers, early detection of colon cancer is the key to a successful treatment. Traditionally, pathologists use microscope to examine histopathological images of biopsy samples taken from patients and make judgments based on their professional expertise. Typically, a pathologist would make observations on some key features in the image and subsequently be able to classify whether or not the tissue under examination contains abnormality. Since this procedure is performed by a human expert, it is therefore subject to inconsistencies due to factors that might affect human performance. To overcome this problem, it has been proposed to use computers to aid in the analysis of medical images, such as histopathological images of colonic tissues. One of the strongest signs of abnormal cellular growth is abundance of DNA material in the nuclei, a condition known as hyperchromasia. Another sign is a 'chaotic' appearance of tissue usually observed during histological viewing. Cancer is known to be characterized by abnormal and excessive proliferation of cells and larger-than-usual nuclei. These effects are visually manifested by a darkening of the regions with excess DNA due to reaction to staining and a loss of structural order in the tissue. In this paper, pixel intensity and the presence of circular formations are examined separately as discriminating image features to distinguish between normal and abnormal samples. The images used in this study were derived from slides and cases randomly chosen from surgical pathology files of a hospital, previously diagnosed by pathologists as colonic adenocarcinoma, adenomatous polyps from the colon, as well as resection planes of the colonic resections without tumor to serve as controls. All samples were stained with hematoxylin and eosin and were taken at 400x magnification. The results showed very promising results.

*Corresponding author
Email Address: ganliml@uni.coventry.ac.uk
Submitted: October 29, 2009
Revised: May 17, 2010
Accepted May 17, 2010
Published: June 17, 2010
Editor-in-charge: Eduardo A. Padlan

KEYWORDS

Automatic Image Classification, Colonic Histopathological Images, Hough Transform, Pixel Intensity, Colon Cancer, Hyperchromasia

INTRODUCTION

It has been reported that in the Philippines, cancer ranks third among the leading causes of morbidity and mortality (Ngelangel and Wang 2002). Cancer of the colon is among the leading types of cancer. Worldwide, colorectal cancer is considered the third most common neoplasm (Shuttleworth 2005). Similar to other types of cancers, early detection of cancer of the colon is key to a successful treatment. Traditionally, pathologists use microscope to examine histopathological images of biopsy samples taken from patients and make judgments based on their professional expertise. Typically, a pathologist would make observations on some key features in the image and subsequently be able to classify whether or not the tissue under examination contains abnormality. Since this procedure is performed by a human expert, it is therefore subject to inconsistencies due to factors that might affect human performance. To overcome this problem, it has been proposed to use mathematics and computers in the analysis of medical images, such as histopathological images of colonic tissues.

Considerable research has been undertaken over the past two decades in an effort to automate cancer diagnosis (Demir and Yener 2005). Based on previous research in this area, it appears that image texture has been a popular choice as basis in choosing discriminating features. Research has shown that textural features derived from grey-level co-occurrence matrices (GLCMs) are very useful. Esgiar *et al.* (1999) analyzed 44 normal and 58 cancer images captured on a computer via microscope with a CCD camera. Atlamazoglou *et al.* (2001) used GLCMs to extract features from a total of 70 fluorescence microscope images of colonic tissue sections stained with a novel selective fluoroprobe. In 2002, Shuttleworth *et al.* proposed to use color texture analysis in classifying colon cancer images, and in 2003 used a multiresolution color texture-based approach to classify images of colorectal tissue into five levels of dysplastic severity. Esgiar *et al.* (2002) proposed the use of fractal-based features in the detection of colon cancer images. Although there was a slight increase in the classification success, it was not significant enough compared to the results of previous methods. Nyowe *et al.* (2006) used shape and texture descriptors calculated from spectral analysis and greyscale statistical co-occurrence matrix analysis of microscopic cell images.

Despite some promising results, most (if not all) of the ideas presented in this area so far have been purely derived from an image processing point of view. This paper suggests that the inclusion of some principles of pathology in developing an automated medical image classifier is vital in its success. It is because of this idea in mind that the pixel intensity and the presence of circular formations in the images have been chosen and examined separately as discriminating image features. To represent hyperchromasia, the mean pixel intensity has been selected as its equivalent feature since darker regions of an image should produce lower pixel intensity values in the RGB color space. On the other hand, the chaotic appearance of an abnormal tissue image sample has been determined based on the character of its accumulator space following an application of Hough Transform to detect circular structures in the image formed by the sectioned glands.

The next sections of this paper will discuss the preparation of the digital images that were used, followed by discussions regarding the use of the two features mentioned in the previous paragraph, and finally the conclusion. Sample images are shown in the next section and a short review of the Hough Transform is included below.

MATERIALS AND METHODS

Preparation of Images

The images used in this study were derived from slides and cases randomly chosen from the 2007 and 2008 surgical pathology files of Medical Center Manila Hospital, previously diagnosed as colonic adenocarcinoma, adenomatous polyps from the colon, as well as resection planes of the colonic resections without tumor to serve as controls. These slides were routinely processed using a Sakura tissue processor and cut at 8 micra using a standard microtome. All were stained with hematoxylin and eosin. All images were taken at 400x magnification using an Olympus DP20 digital photomicrography apparatus mounted on an Olympus microscope (trinocular) at 1200x1800 dpi resolution. There were a total of 75 1600x1200-pixel-images (25 for each diagnosed case) used in this study. Shown on Fig. 1, Fig. 2, and Fig.3 are samples of these images.

Image Classification Using Pixel Intensities

Malignant neoplasms characteristically contain cells with nuclei that have an abundance of DNA and are extremely dark staining or hyperchromatic (Kumar et al. 2005). Cancer is also known to be characterized by abnormal and excessive proliferation of cells and larger-than-usual nuclei. These observations invite one to assume that colonic histopathological images that have cancer must contain darker pixels than images of a normal colon. A simple feature of an image that can reflect this property is the pixel intensity average of the image itself. Images with hyperchromasia can be expected to have lower pixel intensity averages owing to darker staining. The average pixel intensity of a digital image is therefore used here as a discriminating feature in automated classification of colon histopathological images. The classification problem involved 3 categories or cases (normal, adenomatous polyp, and cancerous) and the images used were divided into 2 sets: the training set and testing set. As shown on Table 1, 15 out of the 25 images in each case were devoted for the training of the algorithm while 10 images were used for testing or actual classification. The success and failure in the classification are presented and summarized in a confusion matrix. The images from the training set were used to establish and visualize clustering trends using the image pixel average intensity as the sole discriminating

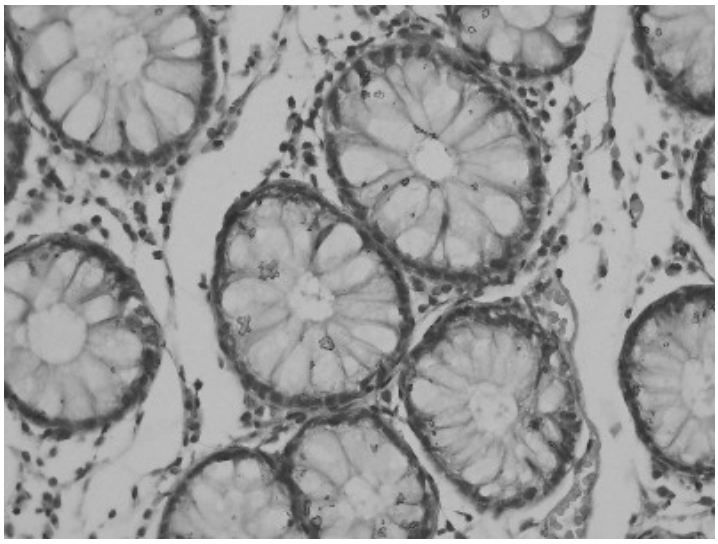

Figure 1. Normal Colon Sample Image

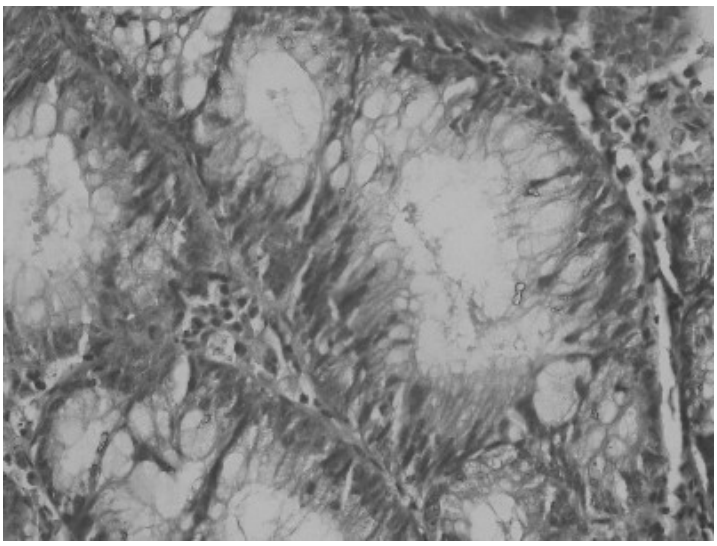

Figure 2. Adenomatous Polyp Sample Image

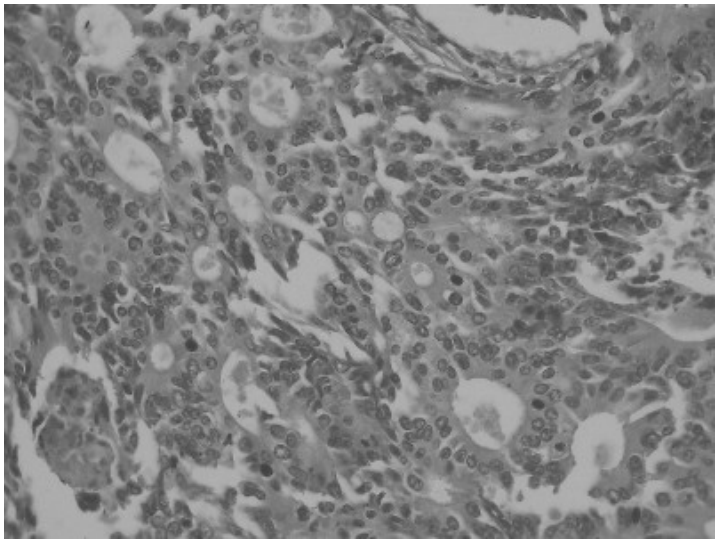

Figure 3. Colon Cancer Sample Image

Table 1. Groupings of the images used in the classification using pixel intensities.

	normal	Adenomatous polyp	cancer
No. of Training samples	15	15	15
No. of Testing Samples	10	10	10
TOTAL per category	25	25	25

Table 2: Confusion matrix of the testing set of images resulting from using pixel intensities. The rows represent the 'true' classification, while the columns represent the classification using pixel intensities.

	'Normal'	*'Polyp'*	*'Cancer'*
Normal	10	0	0
Polyp	2	8	0
Cancer	0	3	7

feature. The effectiveness of using such feature was evaluated with the use of the images from the testing set and a 1D version of the nearest neighbor algorithm to facilitate image classification. The best classification performance was heuristically achieved by using a distance radius of 10 pixel units. As with a conventional nearest neighbor algorithm, each image from the testing set was given a classification by examining the number of data points from the training set that fall within the chosen distance radius, with the image being classified as the center or reference point. Decision of classification is given to the case or category with the most number of training data points that fell within the distance radius. This meant that previously classified images that belonged to the testing set were not used in classifying 'new' images. Errors in classification were then isolated in each data point and not spread throughout the subjects. Actual implementation was done using MATLAB.

Image Classification Using Hough Transform (HT)

Hough transform (HT) is a shape-finding technique that utilizes a mapping of the image points to points in a corresponding accumulator space or Hough space. The accumulator space is formed by interchanging the roles of the variables and the parameters of the equation of a line or a curve in question.

In a binary image, the equation of the pixels forming a circle can be expressed as

$$(x - x_o)^2 + (y - y_o)^2 = r^2.$$

The equation above defines a locus of pixels (x, y) having point (x_o, y_o) as center and radius r. To form the accumulator space, the same equation needs to be interpreted in a different way as a locus of points (x_o, y_o) having point (x, y) as center and radius r. This dual interpretation of the circle equation allows for the identification of the center of the circle in the image by observing the most common intersection point in the accumulator space. Fig.5 and Fig.6 illustrate this principle. The intersection points in the accumulator space are monitored by assigning votes to each point every time a circular curve passes through it. Obviously, intersection points would get high number of votes making them distinct from ordinary points. The center of the circle is ultimately identified by determining the point in the accumulator space with the highest number of votes; which essentially is the point with the most number of circular curves passing through it.

One problem in using HT to find circular shapes is the value of the radius to be used. If the radius is known *a priori*, then all that one needs to do is to plug-in its value to the Hough transform algorithm and everything else will follow. However, in most cases, the circular shapes to be detected do not have known radii to begin with. One way to solve the radius problem is to try several values of the radius and evaluate the votes in the resulting accumulator space. This was the approach adopted in this research.

RESULTS AND DISCUSSIONS

Image Classification Using Pixel Intensities

Surprisingly, the very simple approach used here performed quite successfully compared to other related studies that utilized more elaborate image features. Since the image feature that was chosen was only 1-dimensional, calculation speed was never an issue. Fig. 4 illustrates how the two sets of images were clustered based on the average image pixel intensity, reflecting its usefulness in representing hyperchromasia in cancerous image samples. It can be observed from Fig. 4 that the individual categories are well separated from each other; except for the overlap between the polyp and cancer categories in the training set of images. This overlap between the polyp and cancerous cases was expected since both possess above-normal cellular growths and therefore exhibit higher amounts of DNA material. This naturally translated to lower pixel values due to darker staining of hyperchromatic regions of the tissue. In the testing phase however, the categories are all well separated from each other. Another important observation that can be made from Fig. 4 is that all the images from the polyp case have clustered between the normal and cancerous cases. This is an important observation since it agrees with the fact that among the three categories under consideration, the adenomatous polyp case can be considered to be a 'middle-ground'. Table 2 summarizes in a confusion matrix the classification performance of using average pixel intensity as image feature. It is clear from Table 2 that the

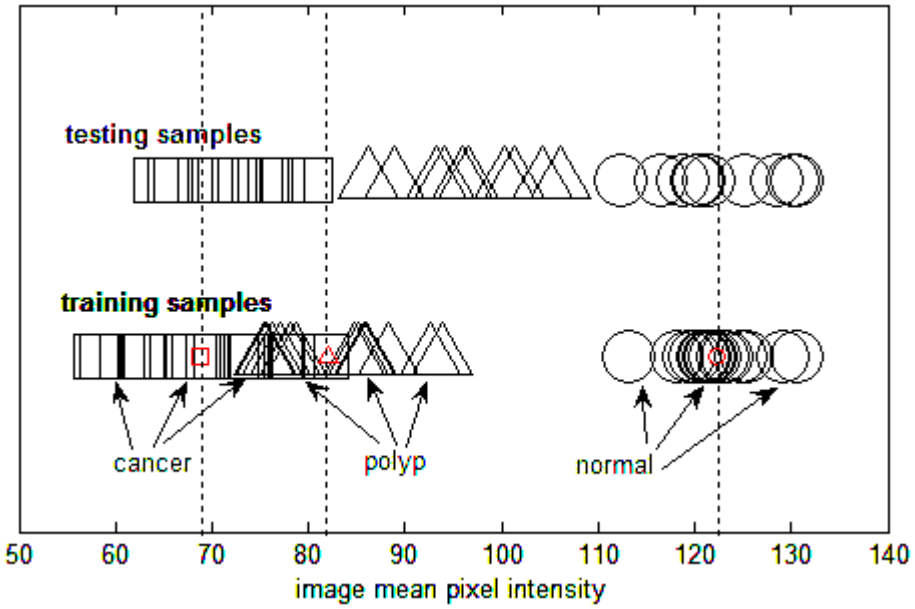

Figure 4. Clusters of image samples in the application of mean pixel intensity as discriminating feature. Circles, triangles, and squares represent 'true' classification: normal, adenomatous polyp, and cancerous cases, respectively. The three vertical lines represent the mean values of the clusters in the training samples.

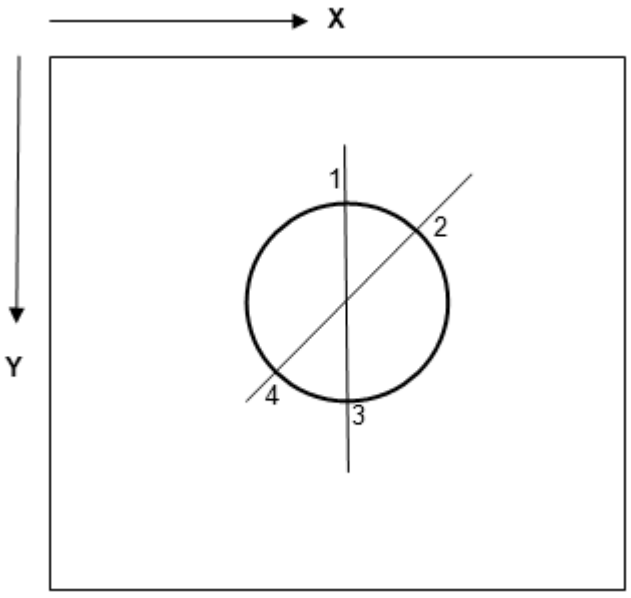

Figure 5. Example of an image containing a circle

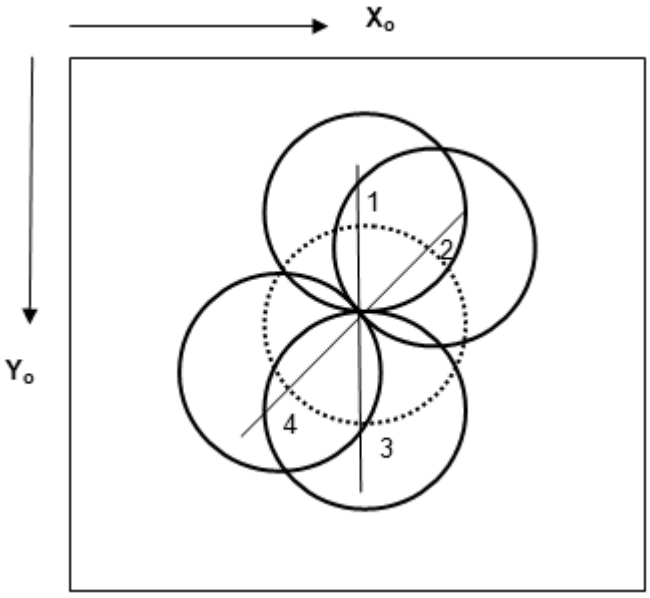

Figure 6. Accumulator space of image in Fig. 5

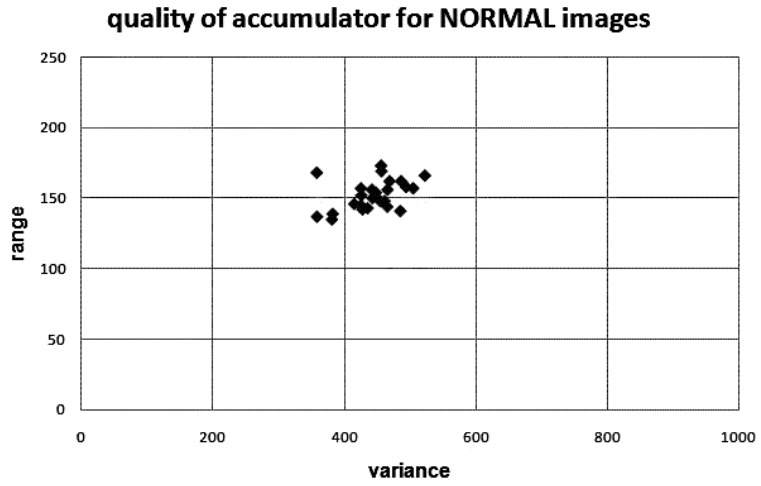

Figure 7. Range-variance scatter of the votes in the HT accumulator space for the NORMAL images

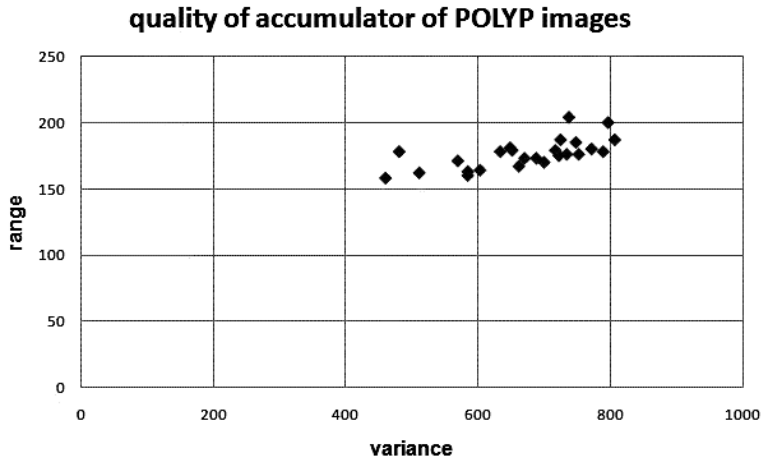

Figure 8. Range-variance scatter of the votes in the HT accumulator space for the A. POLYP images

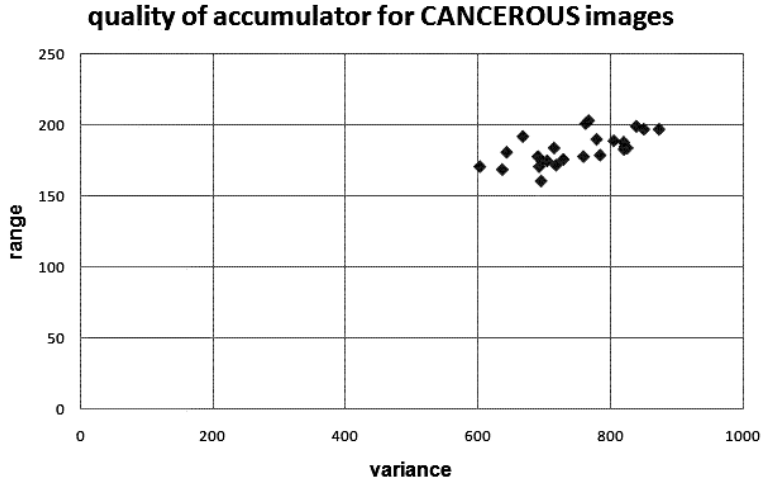

Figure 9. Range-variance scatter of the votes in the HT accumulator space for the CANCEROUS images

identification of the normal images was without mistake (10 correct classifications out of 10 images). Keep in mind that in Table 2, the rows represent the distribution of images as diagnosed by a human pathologist while the columns represent the classification done by the computer using the average pixel intensity as discriminating feature. Ideally for a confusion matrix such as Table 2, there should be no off-diagonal elements since these elements only signify differences between the classification schemes used in the rows and in the columns. Table 2 also shows that misclassification problems affected only the distinction between the polyp and cancerous cases. Of the 10 actual polyp images, 2 were mislabeled as normal, while 3 were mislabeled as polyp cases out of the 10 actual cancerous images. More importantly, there were no actual cancerous images that were mislabeled as normal and vice versa. This is certainly good news as far as the results of this experiment is concerned since no cancerous conditions were misdiagnosed as normal (false negatives). All the images in this study were used 'as-is' and no prior image processing was done. Some image processing schemes have been known to destroy some important image features.

The relatively low numbers in the off-diagonal of the confusion matrix in Table 2 clearly show the great potential of using the average pixel intensity as an ideal discriminating feature in automated classification of colonic histopathological images. It can be used together with other image properties such as texture. Future studies will focus on developing image processing algorithms that will allow a more controlled application of the average image pixel intensity to regions where hyperchromasia are more likely to be observable instead of applying it to the entire image. This will probably enable a more refined distinction between the polyp and cancerous cases as non-relevant regions of the tissue under consideration will not be included in the automated decision-making process.

Image Classification Using Hough Transform (HT)

To improve execution time, all the images were resized to 560 x 420 pixels, 8-bit gray-scale. This step was not expected to affect the outcome of the classification since the profile of the glands remained clearly visible in each image. The following steps were taken to cluster each colonic image sample:

- Read image into memory.
- Utilize the Canny edge detection algorithm to generate a binary image. (The Sobel technique was initially applied but did not provide satisfactory results.)
- Implement Hough transform on the binary image to produce accumulator space with votes.
- Analyze accumulator space by calculating the variance and range of the votes.

To evaluate each accumulator space produced, the properties of the vote profile in each accumulator space were used to represent the quality of the accumulator space itself. Similar to any application of HT to detect lines or curves, it is expected that images with clear circular shapes would produce accumulator spaces with distinct areas having distinct spikes indicating locations of center points of detected circles. A good way to represent these spikes is to use the statistical variance and range of the votes in each accumulator space. To address the problem of choosing a radius value for each image, accumulator spaces were generated for each image using different values of the radius starting with 20 pixels up to 200 pixels with increments of 20 pixels. The radius value that produced an accumulator space with the highest sum of variance and range was ultimately selected as the optimum radius value corresponding to a particular image. It turned out that for all the images examined here, the selected optimum radius values fell within the range between 160 pixels and 200 pixels. Figures 7, 8, and 9 illustrate the scatter plot of the different image categories with the variance and range of the votes in the accumulator spaces forming the axes. It is clear from these plots that normal and cancerous cases clustered in different regions and the distribution of the 'polyp' data points is consistent with the fact that the adenomatous polyp case can be considered to be a middle ground between the normal and cancerous cases.

CONCLUSION

The effectiveness of using pixel intensity and the quality of the Hough Transform accumulator space as discriminating features in the classification of colonic histopathological images have been shown to be very promising. As mentioned earlier, lower mean pixel intensity of an image tends to be associated with hyperchromatic regions. Fig.4 is strong evidence that hyperchromasia can be conveniently represented by this feature. Meanwhile, the fact that there was clear clustering in the variance-range plots of the HT application (Fig. 7, Fig. 8, and Fig 9) only tells us that the character or quality of an accumulator space has strong relation with tissue abnormality. Unfortunately, this part of the study did not go as far as using another set of images to serve as a testing group similar to what was done with mean pixel intensity in the first part of this paper. Nevertheless, the results showed the promise of using HT. Future work will try to combine the two procedures that were presented here and consider adding texture and soft computing algorithms such as fuzzy logic, neural networks, and genetic algorithms.

ACKNOWLEDGEMENT

The authors wish to give acknowledgement to the University Research Coordination Office (URCO) of De La Salle University for providing financial and administrative support

CONFLICTS OF INTERESTS

The authors certify that there is no conflict of interest arising from this study including the publication of this paper.

CONTRIBUTION OF INDIVIDUAL AUTHORS

Laurence A. Gan Lim

The main author and researcher in this paper. This study is part of his PhD dissertation.

Raouf N.G. Naguib and Elmer P. Dadios

Guided the main author in the formulation of the algorithms.

Jose Maria C. Avila

Provided the microscopic images used in this study and discussed some key points of Pathology with the main author.

REFERENCES

Atlamazoglou V, Yova D, Kavantzas N, and Loukas S. Texture Analysis of Fluorescence Microscopic Images of Colonic Tissue Sections. Medical and Biological Engineering & Computing, 2001; 39: 145-151.

Demir C and Yener B. Automated cancer diagnosis based on histopathological images: a systematic survey. Technical report, Department of Computer Science, Rensselaer Polytechnic Institute TR-05-09 2005. Available from http://www.cs.rpi.edu/research/pdf/05-09.pdf [October 29, 2009]

Esgiar AN, Naguib RNG, Sharif BS, Bennett MK, and Murray A.Texture Descriptions and Classification for Pathological Analysis of Cancerous Colonic Mucosa. IEE Proc., Int. Conf. on Image Processing and its applications, University of Manchester, UK 1999; 1: 335-338.

Esgiar AN, Naguib RN G, Sharif BS, Bennett MK, and Murray A. Fractal Analysis in the Detection of Colonic Cancer Images. IEEE Transactions on Information Technology on Biomedicine 2002; 6: 54-58.

Kumar V, Abbas A K, Fausto N (eds.) Robbins and Cotran Pathologic Basis of Disease. 7th edition. Elsevier Inc., 2005.

Ngelangel CA and Wang EH. Cancer and Philippine Cancer Control Program. Japanese Journal of Clinical Oncology 2002; 32: S52-S61.

Nwoye E, Khor L C, Dlay S S, Woo W L. Spectral and Statistical Features in Fuzzy Neural Expert Machine for Colorectal Adenomas Adenocarcinoma Classification. Proc. of Communication Systems, Networks and Digital Signal Processing (CSNDSP) 5th International Symposium 2006; 504-508.

Shuttleworth JK, Todman AG, Naguib RNG, Newman RM and Bennett MK. (2002) Multiresolution Colour Texture Analysis for Classifying Colon Cancer Images. Proc. Int. Conf. IEEE Eng. in Med. and Biol. Soc. 2002; 1118-1119.

Shuttleworth JK, Todman AG, Naguib RNG, Newman RM and Bennett MK. Fine Grading of Colorectal Biopsy Images Using Colour Texture Analysis. Proc. Medical Image Understanding and Analysis Conf. 2003; 169-172.

Shuttleworth JK, Todman AG, Naguib RNG, Newman RM and Bennett MK. Enhancing Feature Extraction from Colon Microscopy Images Using Colourspace Rotation. Proc. Medical Image Understanding and Analysis Conf. 2005; 11-14.