An Automated System for the Detection of Stratified Squamous Epithelial Cancer Cell Using Image Processing Techniques

Ram Krishna Kumar [#1], M.Monica Subashini [*2], Sarat Kumar Sahoo [#3]
[#1] BO-In (operations), Ericsson Global India Service Pvt. Limited
Kolkata, India
[#3]School of Electrical Engineering, VIT University
Vellore, India
[1] ram.krishna.kumar@ericsson.com
[3] Sarata1@rediffmail.com
[*2]School of Electrical Engineering, VIT University
Vellore, India
[2] monicasubashini.m@vit.ac.in

***Abstract*---Early detection of cancer disease is a difficult problem and if it is not detected in starting phase the cancer can be fatal. Current medical procedures which are used to diagnose the cancer in body parts are time taking and more laboratory work is required for them. This work is an endeavor to possible recognition of cancer cells in the body part. The process consists of image taken of the affected area and digital image processing of the images to get a morphological pattern which differentiate normal cell to cancer cell. The technique is different than visual inspection and biopsy process. Image processing enables the visualization of cellular structure with substantial resolution. The aim of the work is to exploit differences in cellular organization between cancerous and normal tissue using image processing technique, thus allowing for automated, fast and accurate diagnosis.**

Keyword- Cancer and Non-Cancerous Sample Images, Pre-processing, Morphological Pattern, Diagnostics, MATLAB

I. INTRODUCTION

Cancers that originate in the stratified squamous epithelia caused more than 300,000 cancer deaths in 2006 and accounted for more than 50% of all diagnosed cancers [1]. Each year, tens of millions of people are diagnosed with cancer and half of them die eventually [2]. Early detection and treatment of these cancers is important to minimize death risks. Tissues with stratified squamous epithelia include the cervix, colon and oral cavity. Currently, the diagnosis of most stratified squamous epithelial cancers is carried out through biopsy and detection by histological evaluation.

In the traditional method of medical microscopic image processing, the doctors have to extract the features manually through the microscope, and draw conclusion according to the stylebook. This is a time consuming and a tedious job. In order to overcome these problems, an automation of medical image analysis that previously requires manual operations is performed on the basis of developments in image processing algorithms. The objective of this work is to develop an automatic image processing model. The segmentation of cancer cells in image processing gives more accurate results, comparable to that from a specialist.

The results from this work could be incorporated into diagnostic tools that provide early diagnosis of epithelial pre-cancers, potentially increasing the diagnostic accuracy and versatility of these techniques.

II. METHODOLOGY

The algorithm followed in this work has been high-lighted as a flowchart in Fig.1

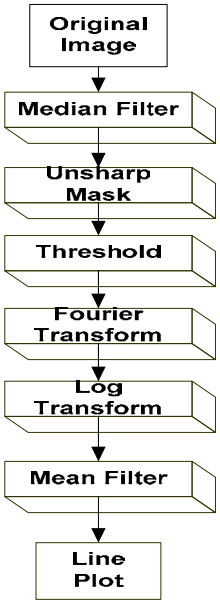

Fig 1. Flow Chart on the Algorithm Implemented

A. Median Filter

Median filtering helps in the removal of salt and pepper noise. A 2-dimensional median filter was applied using the 'medfilt2' function in Matlab. Each output pixel contains the median value in the 5-by-5 neighbourhood around the corresponding pixel in the input image. 'Medfilt2' pads the image with zeros on the edges, so the median values for the points within 3 pixels of the edges may appear distorted. It is more effective than convolution when the goal is to reduce noise and also preserve edges.

B. Unsharp Mask

The idea of unsharp masking is to subtract a scaled unsharp version of the image from the original. The schema for unsharp masking is shown in Fig 2. The result of unsharp masking appears to be a better image than the original; the edges are crisper and more clearly defined.

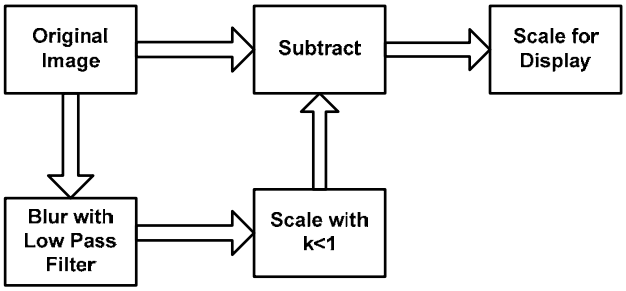

Fig 2. Process of Unsharp Masking

Contrast between the cytoplasm, and nuclei and extracellular components were enhanced using an unsharp filter [4]. The filter has been applied to the image by subtracting the gaussian filtered input image, multiplied by a scaling factor, from the input image. The gaussian filter was created using the built-in Matlab functions 'fspecial' and 'gaussian'. A rotationally symmetric Gaussian lowpass filter with a standard deviation of 10 pixels was used, with a total filter size of 15-by-15 pixels. The scaling factor was 0.9.

C. Threshold (T)

A greyscale image was turned into a binary (black and white) image by first choosing a grey level in the original image, and then turning every pixel black or white according to whether its grey value is greater than or less than T :

A pixel becomes: White if its grey level is > T
Black if its grey level is < T

This function is used to convert image into binary image. The value of 'level' lies between [0, 1]. The graythresh function uses Otsu's method which chooses the threshold to minimise the intraclass variance of the

black and white pixels [5]. The function ignores the imaginary non zero part of the image. As shown in Fig.3 and Fig. 3a, after applying the threshold function to the unsharp masked image, it was found that the cytoplasm was white (or 1) and the nucleus and extracellular components were black (or 0).

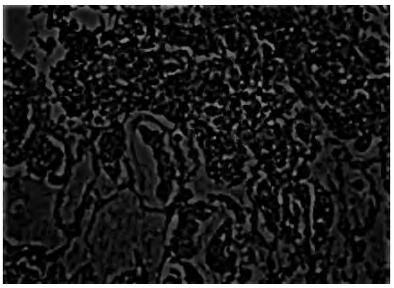

Fig 3. Unsharp masked image

Fig 3(a). Image after thresholding

D. Fourier transform

The binary images were converted into the spatial frequency domain using the two-dimensional discrete Fourier transform. The Fourier Transform allows to isolate, process particular image frequencies and also to perform low-pass and high-pass filtering with a great degree of precision [3]. The built-in Matlab function 'fft2' was used to convert the binary images into the spatial frequency domain using the two-dimensional discrete Fourier transform. The image was shifted before the Fourier transform so that the zero frequency component was at the center of the frequency space. The image becomes easy to visualize when the shift is done in the centre.

E. Log transform

The log transform compresses the values of the light pixels of the image and expands the values of the dark pixels of the image. This reduced DC values relative to the rest of the pixel values, allows the details of the transform to become visible. The general formula of log transformation is

$$s = c\log(1+r) \tag{1}$$

where c is a constant and it is assumed $r \geq 0$.

This transformation maps a narrow range of low gray level values in the input image into a wider range of output levels.

F. Mean Filter

Mean filters are linear filters. The value of each pixel is replaced by the average value of the pixels of the neighbourhood as defined by the mask. It filters the data 'a' with 2-d filter matrix in 'h'. The images obtained after log are fairly noisy which may make automatic detection schemes challenging. To reduce the noise, a 5 by 5 pixel mean filter was implemented. This filtered averaged 25 points thus reduced the noise by 5. Because a single pass of this filter did not seem to provide sufficient noise reduction, the image was passed through the filter a second time. The only disadvantage is that it blurs image.

G. Line plot

The centre row of pixels are extracted and their values are plotted against their positions to get two dimensional images such that they could easily and quantitatively be analyzed by one dimensional signal processing techniques. The line plot is done so that the features can easily be studied which is little difficult in a 2-d image.

III. MATLAB IMPLEMENTATION

MATLAB is a high-performance language for technical computing and interactive environment for algorithm development [6]. In this paper, the algorithm developed for differentiating normal and cancer cells which has been shown as a flowchart in Fig.1 has been successfully implemented using image processing techniques [7].

Steps Followed in MATLAB to differentiate cancer cells from normal cells:

- RGB images are converted to Gray scale images
- Images are enhanced using histogram equalization.
- Salt and pepper noise present in the images are removed by median filter.
- Edges are clearly defined with the usage of gaussian Filter
- Essentials in the images are highlighted by thresholding the images.
- Log transformation is applied to differentiate the cancer images from normal images.
- Line plot was applied in order to generate the local minima.

IV. RESULTS AND DISCUSSION

The varying features start to become apparent after 'log transformation' of the input images which might be used to automatically separate the normal from the cancer. In the cancer samples, the low frequency bright spot is fairly uniform. Looking closely at the normal samples, a dark ring is visible. The results are shown below in the sub section 4.1

A. Log Transformed Cervix Images

The images of a cervix normal cell and cervix cancer cell were log transformed and the results are shown in Fig. 4 and Fig. 4(a).

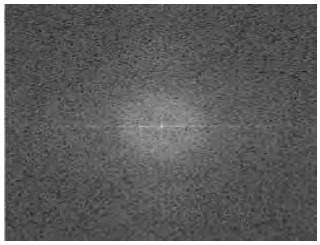

Fig 4. Cervix Normal Cell

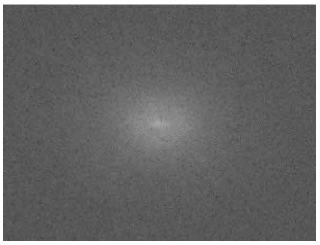

Fig. 4(a). Cervix Cancer Cell

B. Line Plot on Cervix Cell Images

The line plot has been analysed and automated diagnosis of the cancer is done. Once the plotting is done, the difference between cancerous and non-cancerous cells becomes apparent. In a normal cell, the plot gives a local minima which is absent in case of a cancerous cell and the graph is shown in Fig..5. The graph of cancerous cell appears smoother and is shown in Fig.5 (a). The presence of local minima suggests repeating nature of normal cell while its absence suggests changes in the cell structure.

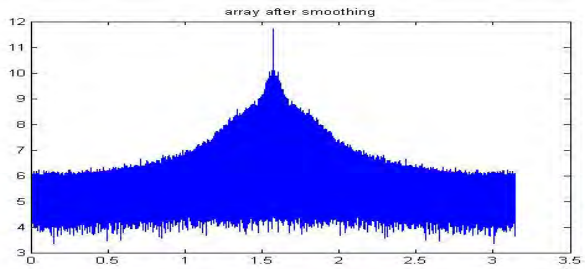

Fig. 5 Cervix Normal Cell with local minima

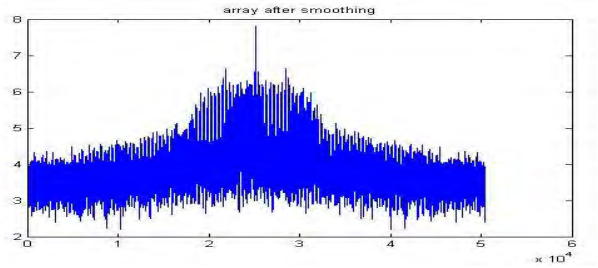

Fig. 5(a): Cervix Cancer Cell without local minima

There are a number of benefits that result from an automated analysis. These include an acceleration of time and reduction in cost as well as decrement in a false inspection due to fatigue. The results highlighted in this paper are based on normal cell (cervix) and cancer cell (cervix). Similarly the images on normal and cancer cell of pancreas, lung, stomach and breast were tested with the algorithm and successfully cancer cells were differentiated.

V. Conclusion and Future Scope

Multiple image enhancement steps are needed to exaggerate the differences between the frequency-domain images of normal and cancerous tissues (median filter, unsharp mask, threshold). Additional enhancements are needed to improve contrast between the power spectrum of normal and cancerous tissues (averaging). After these enhancements, clear differences could be seen between the normal and cancerous power spectrum.

With more extensive testing, it is believed that the local maximum could be used to quantify the organization of the tissue structure. This could be investigated in the future and potentially lead to automated diagnostic techniques, reducing the cost and increasing the accuracy of epithelial biopsy procedures.

Hence the graph can be analysed and automated diagnosis of the cancer can be done. The same technique further can be used to identify the stage of cancer the patient is in. This can be done by creating a database and then comparing the results with it. Further by quantising the peaks of minima formed, the location of cancer can be found which in turn will help in biopsy of the patient more accurately.

Hence, the algorithm resulted in differentiating cancer cell from normal cell, which can be implemented in the real world for better, fast, and accurate cancer diagnostics.

VI. Acknowledgements

The sample images were collected from Kasturba Medical College, Manipal and Mammalian cell culture lab, VIT University. This work has been supported by School of Electrical Engineering, VIT University.

References

[1] Melissa C. Skala 1 , Jayne M. Squirrell3, Kristin M.Vrotsos1, Jens C. Eickhoff2 , Annette Gendron-Fitzpatrick4, Kevin W.Eliceiri3 and Nirmala Ramanujam, “Multiphoton Microscopy of Endogenous Fluorescence Differentiates Normal, Precancerous, and Cancerous Squamous Epithelial Tissues”, doi: 10.1158/0008-5472.CAN-04-3031 Cancer Res February 15, 2005 65; 1180 .

[2] Xiaomei Ma and Herbert Yu*, “Global Burden of Cancer”, Yale Journal of Biology and Medicine 79 (2006), pp.85-94.

[3] F. W. Fitzke, B. R. Masters, R. J. Buckley and L. Speedwell, “Fourier Transform Analysis of Human Corneal Endothelial Specular Photomicrographs”. Exp. Eye Res. 65, 205±214 (1997)

[4] Ling Li, Guitao Cao, Jun Shi, Heng Wu, Xianxia Zhang, “Detecting Immature Precursor Cells in Pathological Images of Bone Marrow Based on Morphology”, Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010) 978-1-4244-5934-6/10/$26.00 ©2010 IEEE (2010).

[5] E. Bengtsson , C. Walhby and J. Lindbald, “Robust cell image segmentation methods”, (1994)

[6] Pankaj Agrawal, S.K.Shriwastava, and S.S.Limaye, “ MATLAB Implementation of Image Segmentation Algorithms”, (2010) IEEE.

[7] Rafael C. Gonzalez, Richard E.Woods, Steven L. Eddins. Digital Image Processing Using MATLAB. Publishing House of Electronics Industry, Beijing 2005, 9: 252-326