Computer Aided Molecular Modeling approach of H & E (Haemotoxylin & Eosin) images of Colon Cancer

Venkata Subbaiah Kotakadi
DST PURSE Centre, S.V. University, TIRUPATI-517502.

Gaddam Susmila Aparna
Department of Virology, S.V. University, TIRUPATI-517502.

Prashanth B.U.V
DST PURSE Centre
S V University, Tirupati-517502.

D.V.R. Sai Gopal
DST PURSE Centre, S.V. University, TIRUPATI-517502.

ABSTRACT
Cancer detection and classification of histopathological images is the standard clinical practice for the diagnosis and prognosis of any cancer. In this paper we present the colon cancer detection and classification of benign and malignant tumor (Nuclei) based on H & E stained histopathology and color segmentation based staining method to distinguish the different types of tissues in biomedical application. Nucleus detection in H&E is a challenging problem considering the variability, heterogeneity, low contrast, K means clustering, and differing typologies of nuclei to distinguish different types of tissues. There are strong indications that morphological analysis in H&E can serve as a biomarker. The segmentation approach is completely colour based and uses k-means clustering technique. This technique uses a series of algorithm steps which is an image processing approach in distinguishing the different tissue types. These algorithm steps are modelled in image processing tool box of MATLAB v7.0. Modelling steps involved are from reading the image to segmentation of the nuclei into a separate image. Further there are also the intermediate steps that are involved between reading the image and segment the nuclei into a separate image in MATLAB real-time simulation environment.

Keywords
Histopathology and Colon Cancer, K means clustering,Color Segmentation, Image Processing

1. INTRODUCTION
Inflammatory bowel disease (IBD) is a chronic inflammatory condition that is frequently associated with increased colon cancer risk and Ulcerative colitis is a chronic inflammatory condition of the colon associated with an increased colon cancer risk.. The colon cancer risk increases with length and severity of the disease (1). Computer-aided histological analysis for Colon cancer diagnosis, Colon cancer is the most frequently diagnosed non-skin cancer and the leading cause of cancer death among men in the United States and other countries. Conventionally, pathologists review colon tissue visually under a microscope and a diagnosis of colon cancer is rendered based on subjective perceptions of a number of histological features. In the present paper is that computer histological image-processing techniques have is developed to detect colon cancer automatically at a level of accuracy comparable to an expert pathologist. The long-term goal of this research is to develop a computer-aided diagnosis (CAD) system that can be used to help pathologists diagnose colon cancer with improved accuracy, reproducibility, and efficiency. In present research work, we developed computer image-processing techniques for classifying colon cancer and non-malignant tissue in digital histological images acquired from tissue sections with immunohistochemistry (IHC) and hematoxylin and eosin (H&E) stains. And the Digital color histological images were acquired using light microscopes and charge-coupled device (CCD) cameras from regions of interest (ROIs) marked by pathologists on colon tissue sections.

2. MATERIAL AND METHODS

DESIGN OF ALGORITHM
The different steps are involved in the design of algorithm, are explained as follows: (Ref: 2, 3,4,5,6,7,8)

1. Reading the image
In this step it will read the image and this read image is taken as the input image and is displayed as follows (Fig.1). The haematoxylin and eosin (H&E) stained colon tissue image was taken for the analysis.
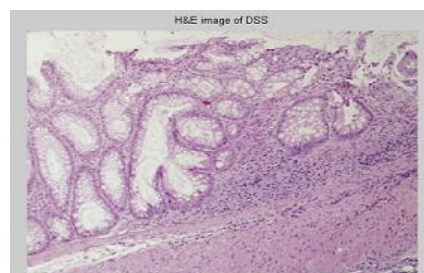

Figure 1.

2. Convert Image from RGB Colour Space to L*a*b* Colour Space
In this step, the input image was analysis based on the different colors that we see in the image, the variation in the image brightness is ignored. In the input image there are three colors white, blue and pink by converting RGB colour space into L*a*b color space we can visually distinguish these colours from one another. The L*a*b* color space is derived from the CIE XYZ tristimulus values. The L*a*b* space consists of a luminosity layer 'L*', chromaticity-layer 'a*' indicating where color falls along the red-green axis, and chromaticity-layer 'b*' indicating where the color falls along the blue-yellow axis. All of the color information is in the 'a*'

and 'b*' layers. we can measure the difference between two colors using the Euclidean distance metric.

3. Classify the Colors in 'a*b*' Space Using K-Means Clustering.

In this step we classify the colors by using Clustering. Clustering is a way to separate groups of objects.

k-means clustering treats each object as having a location in space. It also finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. K-means clustering requires that you specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other. Since the color information exists in the 'a*b*' space, our objects are pixels with 'a*' and 'b*' values. Use kmeans to cluster the objects into three clusters using the Euclidean distance metric.

4. Label Every Pixel in the Image Using the Results from K- MEANS

In this step, the every object in the input are taken and k-mean is calculated, the k-means returns an index corresponding to a cluster. The cluster_center output from k-means will be used for determining the three cluster images. Labeling has to be done for every pixel in the image with its cluster index.
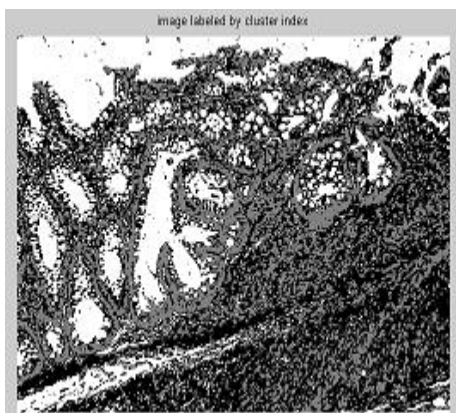

Figure2.

5. Create Images that Segment the H&E Image by Colour

In this step by using Using pixel_labels, we can separate objects in the input image by color, which will result in three different images as seen in the figures. 3, 4 and 5 below.
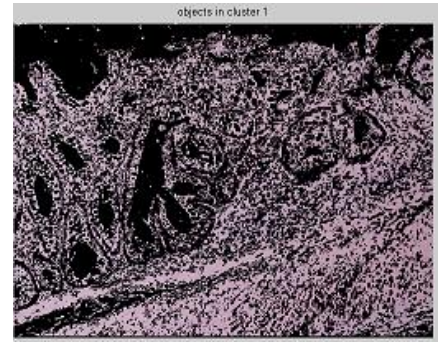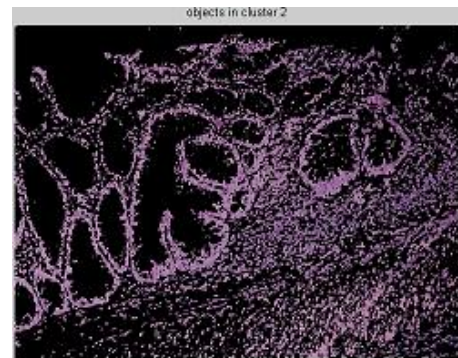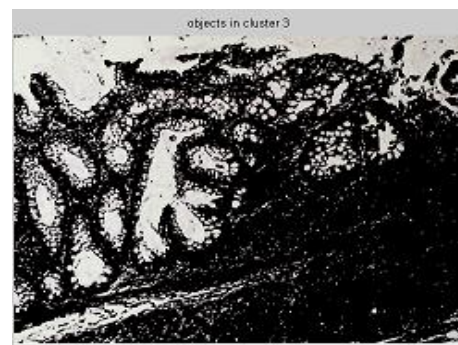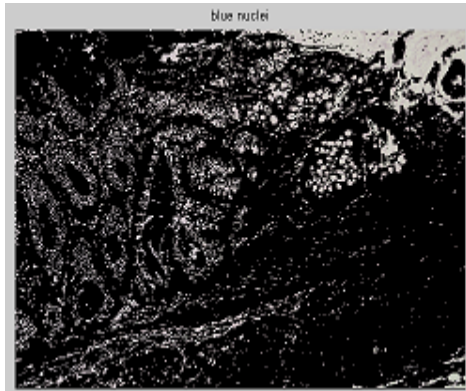

Figure 3.

Figure 4.

Figure 5.

6. Segment the Nuclei into a Separate Image.

In the final step the segmented nuclei or the blue nuclei is separated, in the process there are dark and light blue objects in one of the clusters. We can separate dark blue from light blue using the 'L*' layer in the L*a*b* color space. The cell nuclei are dark blue. Also the 'L*' layer contains the brightness values of each color. The cluster that is found contains the blue objects. After finding the cluster containing blue objects we extract the brightness values of the pixels in this cluster and threshold them using im2bw.The next step is to programmatically determine the index of the cluster containing the blue objects because kmeans will not return the same cluster_idx value every time. We can do this using the cluster_center value, which contains the mean 'a*' and 'b*' value for each cluster. The blue cluster has the smallest cluster_center value (determined experimentally).

Figure 6.

In the next section we are describing the MAT LAB programme and the flow diagram of the work, which gives the basic idea of the image processing software developed in MATLAB.

MATLAB Programme

```
he = imread('colon cancer_H_and_E.png');
imshow(he), title('H&E image of colon cancer');

text(size(he,2),size(he,1)+15,'Image courtesy of Venkata Kotakadi Sri Venkateshwara University-Tirupati','FontSize',7,'HorizontalAlignment','right');
figure;
 cform = makecform('srgb2lab');
lab_he = applycform(he,cform);

ab = double(lab_he(:,:,2:3));
nrows = size(ab,1);
ncols = size(ab,2);
ab = reshape(ab,nrows*ncols,2);

nColors = 3;
% repeat the clustering 3 times to avoid local minima
[cluster_idx cluster_center] =
kmeans(ab,nColors,'distance','sqEuclidean','Replicates',3);

pixel_labels = reshape(cluster_idx,nrows,ncols);
imshow(pixel_labels,[]), title('image labeled by cluster index');
text(size(he,2),size(he,1)+15,'Image courtesy of Venkata kotakadi, Sri Venkateshwara University-Tirupati','FontSize',7,'HorizontalAlignment','right');
figure;
  segmented_images = cell(1,3);
rgb_label = repmat(pixel_labels,[1 1 3]);

for k = 1:nColors
    color = he;
    color(rgb_label ~= k) = 0;
    segmented_images{k} = color;
end

imshow(segmented_images{1}), title('objects in cluster 1');
```

```
text(size(he,2),size(he,1)+15,'Image courtesy of Venkata Kotakadi, Sri Venkateshwara University-Tirupati','FontSize',7,'HorizontalAlignment','right');
figure;
imshow(segmented_images{2}), title('objects in cluster 2');
text(size(he,2),size(he,1)+15,'Image courtesy of Venkata Kotakadi,Sri Venkateshwara University-Tirupati','FontSize',7,'HorizontalAlignment','right');
figure;
imshow(segmented_images{3}), title('objects in cluster 3');
text(size(he,2),size(he,1)+15,'Image courtesy of Venkata Kotakadi, Sri Venkateshwara University-Tirupati','FontSize',7,'HorizontalAlignment','right');
figure;
mean_cluster_value = mean(cluster_center,2);
[tmp, idx] = sort(mean_cluster_value);
blue_cluster_num = idx(1);

L = lab_he(:,:,1);
blue_idx = find(pixel_labels == blue_cluster_num);
L_blue = L(blue_idx);
is_light_blue = im2bw(L_blue,graythresh(L_blue));
nuclei_labels = repmat(uint8(0),[nrows ncols]);
nuclei_labels(blue_idx(is_light_blue==false)) = 1;
nuclei_labels = repmat(nuclei_labels,[1 1 3]);
blue_nuclei = he;
blue_nuclei(nuclei_labels ~= 1) = 0;

imshow(blue_nuclei), title('blue nuclei');
   text(size(he,2),size(he,1)+15,'Image courtesy of Venkata Kotakadi, Sri Venkateshwara University-Tirupati','FontSize',7,'HorizontalAlignment','right');figure;
```
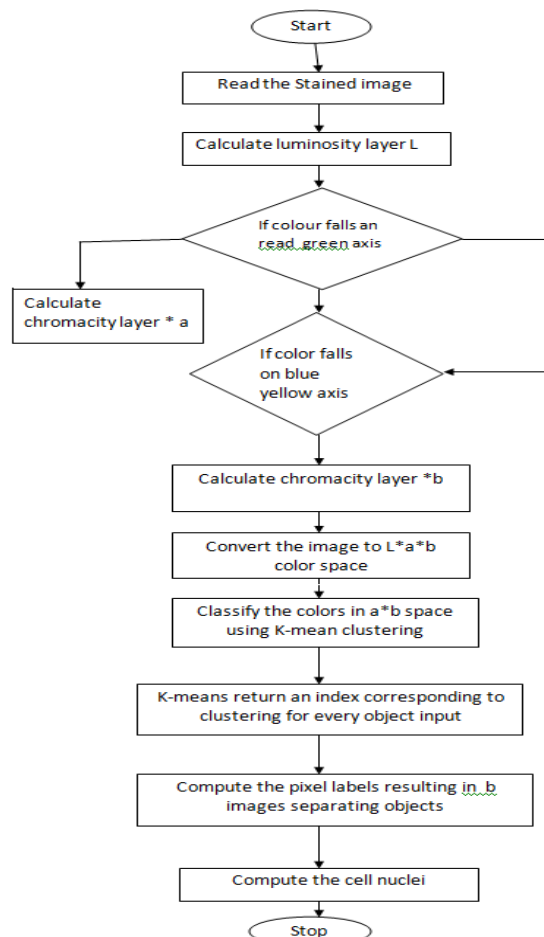
3. FLOWCHART IMPLEMENTATION

The flow chart is derived from the algorithm based on the algorithm above steps.

7

4. RESULTS OBTAINED

The following output simulation result is obtained in MATLAB7.0 real time environment, which takes the original image as the input. According to the algorithm designed the following output result s obtained.

The source code is implemented in the MATLAB 7.0 and the output simulation results are obtained in the real time simulation environment.
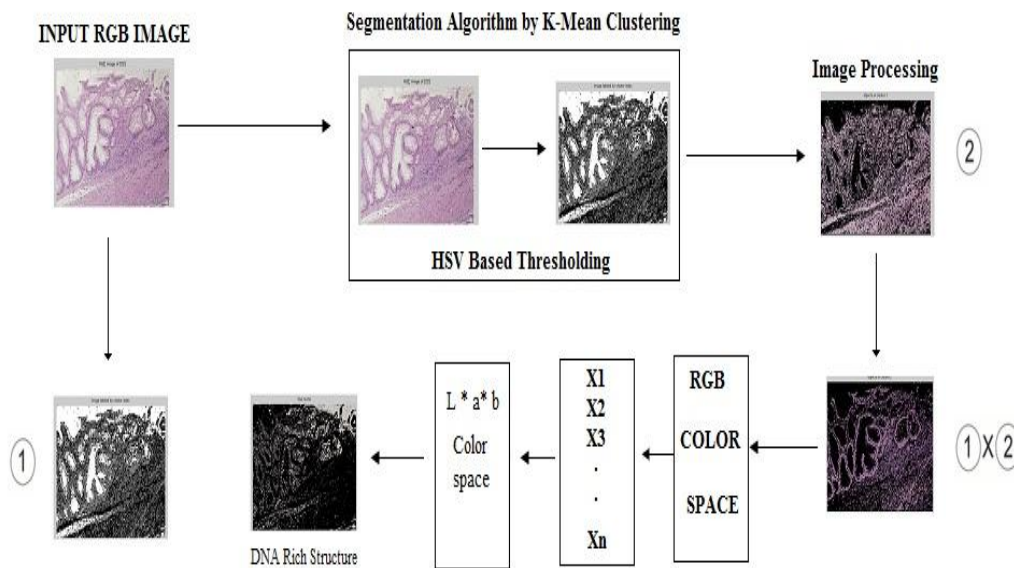
5. CONCLUSION

We have proposed a novel algorithm for color segmentation. Our algorithm improves the color segmentation accuracy significantly. One distinctive advantage of our approach is that our algorithm is also provides an explicit feature for regions which can be used for classifications and recognitions.

6. AKNOWLEDGEMENTS

7. REFERENCES

[1] Itzkowitz,S.H. et al. (2004) Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. Am. J. Physiol. Gastrointest. Liver Physiol., 287, G7–G17.

[2] Otsu Nobuyuki, "A Threshold Selection Method from gray level Histograms", vol. SMC-9, no.1, January 1979

[3] I. Pitas, A.N. Venetsanopoulos, (1990) "Nonlinear Digital Filters

[4] Gonzalez R.C., Woods R.E., (1993) "Digital Image Processing",

[5] Pratt W.K., (1991) "Digital Image Processing",

[6] G.Deng, L. W. Cahill and G.R. Tobin, "The Study of Logarithmic Image Processing Model and Its Application to Image Enhancement"

[7] Joongho Chang, Gunhee Han, Hose M. Valverde, Norman C. Grisworld, J. Francisco Duque-Carillo, Edgar anchez-Sinencio,(1997) "Cork Quality Classification System using a Unified Image Processing and Fuzzy-Neural Network Methodology", vol. 8, no. 4, July

[8] Crabb D.P., Edgar D.F., Fitzke F.W., Mcnaught A.I. & Wynn H.P. (1995). New Approach to Estimating Variability in Visual-Field Data Using An Image-Processing Technique. British Journal of Ophthalmology 79:213-217.