Histological Image Segmentation using Fuzzy C-Means

Rupesh Mandal
Sikkim Manipal University
Sikkim Manipal Institute of Technology
Dept. of CSE

Nupur Choudhury
Assam Don Bosco University
Dept. of CSE & IT

Baishali Goswami
Sikkim Manipal University
Sikkim Manipal Institute of Technology
Dept. of CSE

ABSTRACT
This paper deals with the automatic segmentation of Haematoxylin and Eosin(H&E)stained Histological slide image with the help of advanced soft clustering mechanism. The clustering mechanism used in this proposed framework is Fuzzy C-Means (FCM) algorithm and it is implemented on the human skin dataset. The dataset is obtained by digitally scanning the H&E stained histological slide of human skin tissue with the help of WSI (Whole Slide Image) scanner. The FCM clustering mechanism is implemented on the image obtained after converting to L*a*b* colour space. This paper presents a detailed discussion regarding the L*a*b* colour space followed by the soft FCM algorithm. The experiment is carried out on human skin tissue and the results obtained after segmentation is shown which will prove to be helping hands for the medical practitioners for identifying and extracting the Region of Interest (ROI) for the purpose of diagnosis.

Keywords
Histological slide, Fuzzy C-Means, L*a*b* colour space, colour image segmentation.

1. INTRODUCTION
The process of image segmentation is one of the major and cumbersome process in the image processing domain. It is the process of assignment of various labels to the respective pixels of an image which would result in the sharing of common visual characteristics by the pixels having similar labels. Several existing applications in this domain involves accurate image segmentation like identification of objects, identification of object position and feature extraction. Various mechanisms of segmentation of medical images are already been proposed and implemented which includes methods based on edges, regions or combination of both the methods. The primary motive of the segmentation of medical images is to facilitate an image with clear meaning and easy understandability and analysis.

Reference [1] demonstrates, the various approaches which forms a part of the image segmentation process which includes categories like thresholding, clustering, edge detection andextraction of region. The easiest and the simplest process of segmentation is thresholding in which the pixels are categorised based on the intensity values. On the other hand, various edge based information is used in order to identify the object boundaries of an image which would enable the formation of closed regions. Similarly, the initial seed point neighbouring pixels are examined by the region based segmentation methods which helps in determination of the neighbouring pixels that needs to be added to the region. The most commonly used algorithms in image segmentation include clustering algorithms. Jain et. al[2], in their recent review, classifies clustering methods into two classes namely partition and hierarchy based which differentiates from each other by the process of production of a nested partition series by the later method.(Generally, the schema of the partitions are generated with the help of a dendogram, which is responsible for representing the pattern groups and the levels of similarity at which a change is observed). On the contrary, the partition based methods can generate just single partition.

In this paper, for image segmentation process, a clustering based method would be discussed which essentially the Fuzzy C-Means is.It is the process of categorising or classifying the various object and patterns in a way that objects with similar characteristics belong to one single group than the objects that have varied characteristics. The clustering methods can further be categorised into hard clustering and fuzzy clustering schemes specified by its own characteristics. The method in which each of the points of the data set or the pixels are restricted to a single cluster is known as hard clustering and forms the conventional clustering method. However the Fuzzy C means utilises the concept of fuzzy set theory [3] which proposes the idea of generating partial membership of involvement specified by a membership function. This can also be referred to as soft segmentation method that involves a wider domain and has been successfully applied in image segmentation techniques. Fuzzy C-Means (FCM) algorithm [4] isthe most commonly used clustering method among all the soft clustering methods since ambiguity is prevented due to its robustness and can store larger amount of information as compared to the hard segmentation methods[5].

2. BACKGROUND STUDY

2.1.L*a*b* Color Space
This is also known as CIE L*a*b* Horizontal axes are represented using a* and b* and the vertical axis is represented using L* axis which represent lightness with a range between 0-100. These axes cross each other at the centre which is generally the neutral (grey, black or white) and are placed at right angles to each other. The primary principle is based on the fact that a colour can never be red and green or blue and yellow at the same time. The centre of each of the axes is of the value 0 which is basically used to represent neutral or near neutral colour. The negative part (-a) of the a* axis represents the green colour whereas the positive part stands for red. Similarly, negative b* axis represents blue colour and positive b* axis represent yellow. Theoretically there are no maximum values for a* and b* but practically the range of the maximum values of a* and b* are numbered from -128 to +127(256 levels). This model constitutes of colours which are outside the range of human vision.

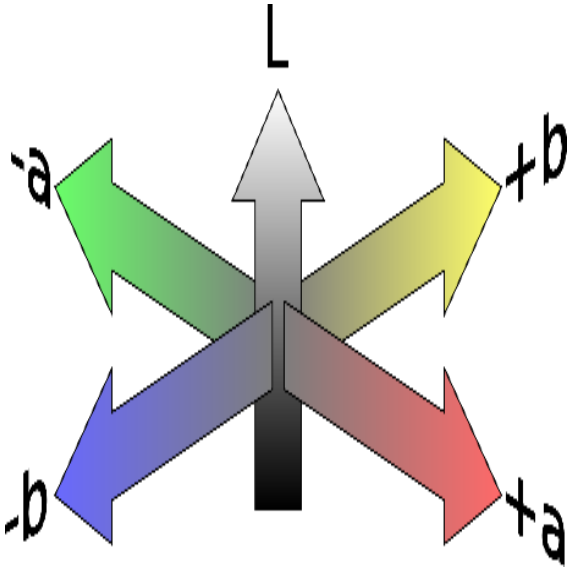

Figure 1: L*a*b* color space

2.2.Fuzzy C-Means (FCM)

This process involves partitioning the set of n objects $x = \{x_1, x_2, \dots, x_n\}$ in R^d dimensional space into c fuzzy clusters $(1 < c < n)$ having y centres or centroids with $y = \{y_1, y_2, \dots, y_n\}$ [6]. A fuzzy matrix μ having n rows representing the number of data objects and c columns representing clusters is used to represent the clustering of the objects in FCM. The matrix element u_{ij} in the ith row and jth column in the fuzzy matrix μ demonstrates the association degree or the membership function of the ith element with the jth cluster. The primary objective of this method is to minimize the below given equation:

$$J_m = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ij}^{m} d_{ij} \qquad (1)$$

$$d_{ij} = \|x_i - y_i\|^2 \qquad (2)$$

Where $1 \le m < \infty$ is the fuzzifier and y_i is thei^{th} centroid of the corresponding cluster which is represented as

$$y_i = \frac{1}{n_i}\sum_{j=1}^{n}(u_{ij})^m x_j \; ; \qquad (3)$$

$$where\ n_i = \sum_{j=1}^{n}(u_{ij})^m ; \qquad (4)$$

$$u_{ij} = \left(\sum_{k=1}^{n}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}} ;\right)^{-1} \qquad (5)$$

The Fuzzy C-Means algorithm is an iterative approach which is stated in following steps [3]:

1. Assign the initial mean y_i and select fuzzifier m such that $(1 \le m < \infty)$; initialization of the membership function values$u_{ij}, i = 1,2,.,n; j = 1,2,\dots,c$.
2. Update the mean (centroid)$y_i, j = 1,2,\dots,c$.according to equation (3).
3. Compute Euclidian distance $d_{ij}, i = 1,2,\dots,n; j = 1,2,\dots,c$.
4. Repeat steps 2 to 4, by incrementing t, until $|u_{ij}(t) - u_{ij}(t-1)| > \in$.

3. METHODOLOGY

In this present research, an input image considered is Haematoxylin and Eosin (H&E) stained histological slide of human skin tissue. This image is obtained by digitally scanning the histological slide using Whole Slide Image (WSI) scanner. The main aim of this research paper is to automatically segment the histological slide image using Fuzzy C-Means algorithm. This kind of work is done manually by medical practitioners at the present time. A framework is developed after discussing with the medical practitioners where input will be an H&E slide image and output will be a segmented image which will show the desired Region of Interest. The entire backbone of this framework depends completely on the segmentation method using Fuzzy C-Means clustering algorithm for segmentation. The workflow of this frame is as follows:

Step 1: The input image is obtained from WSI scanner which is a high resolution RGB image.

Step 2: Convert the RGB image into L*a*b* color space.

Step 3: Apply Fuzzy C-Means clustering technique which produces output in a* and b* pixel values.

Step 4: Label every pixel in image with cluster index obtained by Fuzzy C-Means clustering method.

Step 5: The objects in image are separated using pixel labels and finally segmented image is created.

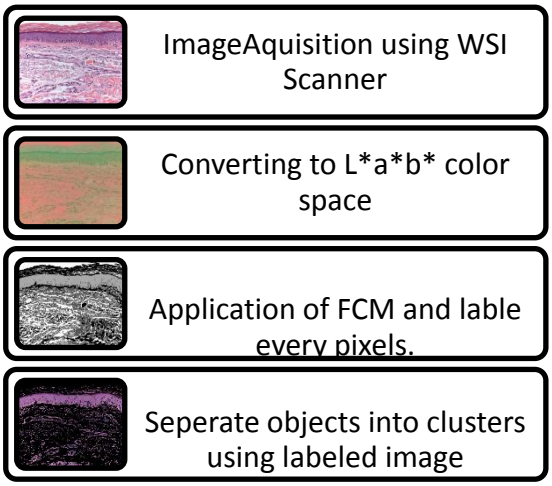

Figure 2: Workflow of the entire framework

4. RESULT AND DISCUSSION

The input image obtained from the WSI scanner is an RGB image which is then converted to L*a*b* color space.

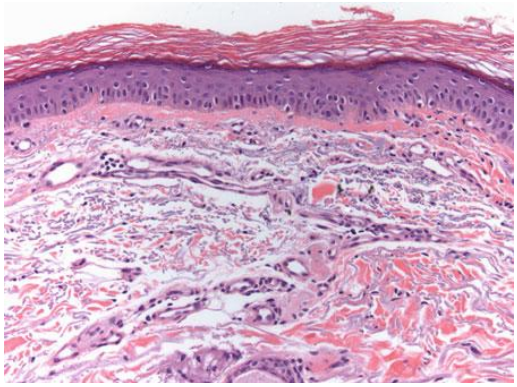

Figure 3: Original input image

This basic approach is followed because the H&E slide images conventionally contain four major colours viz. white, black, pink and blue. The segmentation is done on the basis of these colour differences. Converting to L*a*b* colour space makes it able to quantify theircolour differences. The image obtained after conversion to L*a*b* colour space is as below.

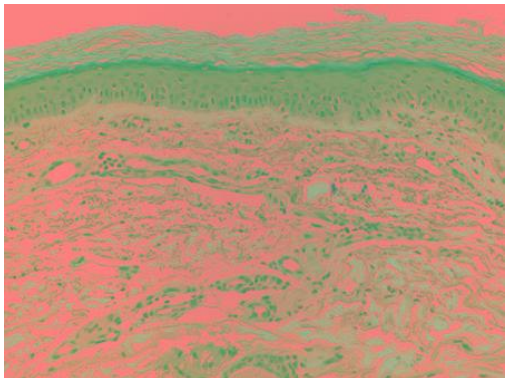

Figure 4: Image in L*a*b* color space

Once this conversion is done the clustering technique is applied. The Fuzzy C-Means algorithm is used for this purpose. The FCM algorithm incorporates membership function which separates the algorithm from the conventional hard K-means clustering method. Each and every pixel is compared based on the distance metric used.Euclidean's Distance is used as the distance metric. FCM returns an index corresponding to a cluster. Every pixel of the image in L*a*b* is thenlabelled using the cluster index obtained from FCM. The image obtained by doing this shown below:

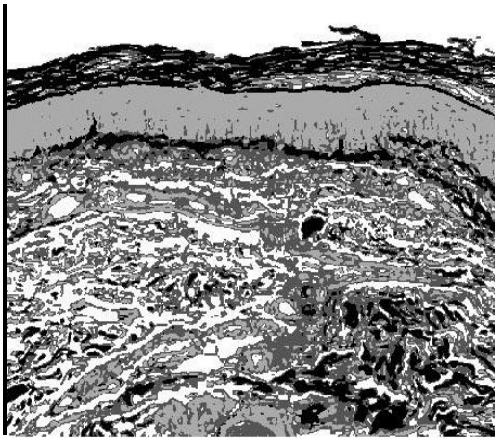

Figure 5: Image labeled with cluster index of FCM

The above labeled image is then used to segment and separate objects belonging to different clusters. By performing this step the desired area of interest is then segmented out. The final segmented image is shown below.

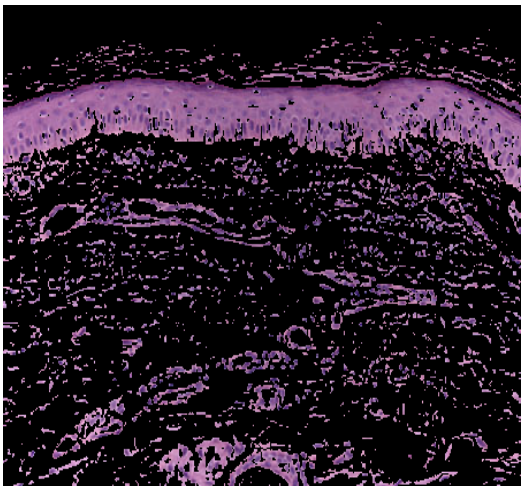

Figure 6: Segmented image using FCM

5. CONCLUSION

In this paper a novel approach for a framework has been proposed that leads to the automatic segmentation of the H&E histological slide which will boost the process of finding the area of interest by the medical practitioners. Fast and efficient soft clustering technique that is Fuzzy C-Means algorithm is implemented in this framework which gives accurate result that will help to find abnormalities in the human skin tissue to the optimum level. This research will provide a great helping hand to the medical practitioners in order to identify and extract the Region of Interest (ROI) from high resolution H&E stained histological slide images, which the medical practitioners have to do manually at present time.

Moreover, a common platform for viewing and giving expert commentsviasmartphone and desktop application can be implemented as a future scope where the medical practitioners can view and comment on the histological images at any point of time and from anywhere just using the device network.

6. REFERENCES

[1] Fu, S.K.—Mui, J.K.: A Survey on Image Segmentation. Pattern Recognition, Vol. 13, 1981, pp. 3–16.

[2] A. Jain, M. Murty, and P. Flynn, "Data clustering: A review," ACM: Computing Surveys, vol. 31, no. 3, 1999, pp. 264–323.

[3] Zadeh, L.A.: Fuzzy Sets. Inform. and Control., Vol. 8, 1965, pp. 338–353.

[4] Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press, 1981.

[5] Bezdek, J.C.—Hall, L.O.—Clarke, L.P.: Review of MR Image Segmentation Techniques Using Pattern Recognition. Med. Phys., Vol. 20, 1993, pp. 1033–1048.

[6] Lin, W., E. Tsao and C. Chen,Constraint satisfaction neural networks for image segmentation, In: T.Kohonen, K. Mkisara, 0.Simula and J. Kangas (eds.), Artificial Neural Networks (Elsevier Science Publishers), 1991,pp: 1087-1090.

[7] Elnomery Allam Zanaty, An Adaptive Fuzzy C-Means Algorithm for Improving MRI Segmentation,Open

Journal of Medical Imaging, Volume 3, 2013, pp: 125-135.

[8] Pradipta Maji and Sankar K. Pal, RFCM: A Hybrid Clustering Algorithm Using Rough and Fuzzy Sets, Fundamenta Informaticae, volume 80,no. 4, November 2007, pp: 475-496.

[9] Pradipta Maji and Sankar K. Pal, Maximum Class Separability for Rough-Fuzzy C-Means Based Brain MR Image Segmentation, LNCS Transactions on Rough Sets, IX(LNCS 5390), December 2008, pp: 114-134.

[10] Pradipta Maji and Shaswati Roy, Rough-Fuzzy Clustering and Multiresolution Image Analysis for Text-Graphics Segmentation, Applied Soft Computing, Volume 30, May 2015, pp. 705-721.