Segmentation of Blood Smear Images using Normalized Cuts for Detection of Malarial Parasites

Subhamoy Mandal[1], *Graduate Student Member, IEEE*, Amit Kumar[1], *Graduate Student Member, IEEE*, J Chatterjee[1], *Member, IEEE,* Manjunatha M[1] , *Member, IEEE,* and Ajoy K Ray[2,3], *Member, IEEE*

[1]School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, India
[2]Dept. of Electronics & Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India
[3]Bengal Engineering and Science University, Shibpur, India

***Abstract*— This paper presents an optimized normalized cut method for segmentation of RBCs infected with malarial parasites using peripheral blood smears. The algorithm is applied over various color spaces to find its optimal performance for microscopic blood smear images. We tested the efficacy of results in RGB, YCbCr, HSV and NTSC using the Rand's Index. The work is useful in telepathology applications and can automate the screening of malaria in rural areas where healthcare manpower is limited.**

***Keywords*— Image Processing, Segmentation, Color Image Processing, Normalized Cuts, Medical Imaging**

I. INTRODUCTION

Malaria is one of the predominant tropical diseases in the world causing wide spread sufferings and deaths in the developing countries. The WHO reports 300 to 500 million clinical cases of malaria each year resulting in 1.5 to 2.7 million deaths [1]. About 40% of the world's populations - about two billion people - are at risk in about 90 countries and territories. The present paper aims at automating the process of blood smear screening for malaria parasite. The method uses the normalized cut (NCut) algorithm [2-3] and tests it over various color spaces. The system is optimized by using HSV color space and recursively repartitioning the segments to find the appropriate order of the smallest eigenvector. The approach can be used to increase the total number of samples screened and can also be used for telemedicine application to enable pervasive healthcare at the base of the pyramid.

II. METHODOLOGY

A. Microscopic Image Grabbing & Preprocessing

The basic aim of our system is to scan the peripheral blood smear and detect any parasitaemia by using image analytics on the acquired images. The acquisition of the image is an important step in the entire process. We use a compound bright field microscope to scan the slides in-situ. This is presently done manually under the supervision of a trained pathologist. Once the image of the entire slide is obtained, the pathologist selects the specific regions of interest (ROI). The ROI selected is grabbed and securely transmitted over the network for analysis [4-5]. Several techniques using morphology [8], Shading Correction [9], and wavelet based techniques [12] have given encouraging results for detection of malarial parasites from blood smear images. It is seen that the quality of the acquired and transmitted images are low and are noisy. Hence, we use de-noising and image enhancement strategies to increase the suitability of images to be segmented. For de-noising we use the interscale orthogonal wavelet based threshold which is a new Stain's Unbiased Risk Estimator (SURE) approach to image de-noising justified by Unser et al [6]. The histogram of the image shows a unique peak; thus we obtain good results using the Brightness Preserving Dynamic Histogram Equalization (BPDHE) algorithm [7]. We however, will concentrate on the strategy of segmentation and abstain from going into details of the de-noising and enhancement technique in this paper.

B. Normalized Cut algorithm

Normalized Cut is a method in which we cut a graph into two components so as to express the cost of the cut as a small fraction of the total affinity within a group [10].

$$NCut(\mathrm{A},\mathrm{B}) = \frac{Cut(A,B)}{assoc(A,V)} + \frac{Cut(A,B)}{assoc(B,V)} \qquad (1.1)$$

The score of the cut expressed by (1.1) where V is a weighted graph and decomposed into two components A and B. Cut (A,B) is sum of weights of all edges in V that have one end in A and other in B. The assoc (A, V) and assoc (B, V) are sum of weights of weights of all edges with one end at A and B respectively.

The NCut essentially searches for the minimum (min) value of the criterion in NCut (A, B). The min value signifies cutting the graph between two components or regions having lesser weighted edges between them and higher internal edges.
We assume y be a vector of elements, each element is located at graph nodes. The graph V is partitioned into two sets A and B. Let us assume x to be a dimensional indicator vector given N = |V|, x_i = 1 if the node i lies in A, else x_i = -1. Now, we denote the affinity matrix by A and degree matrix by D. Each diagonal element of the matrix is the sum of weights entering a given node. It is expressed as,

$$D_{ij} = \sum_j A_{ij} \quad (1.2)$$

each off diagonal elements of D is zero. Consequently, we get

$$\min_x \mathrm{NCut}(x) = \min_y \frac{y^T(D-A)y}{y^T D y} \quad (1.3)$$

The expansion is termed as Rayleigh Quotient [11] and we need to minimize the expression by finding a suitable value of 'y'. To reduce the computational complexity we allow 'y' to take only real values. Thus (1.3) can be minimized by solving the generalized eigenvector system.

$$(D - W)x = \lambda Dx \quad (1.4)$$

The smallest eigenvalue being zero the eigenvector corresponding to 2nd eigenvalue is sufficiently used. Thus we re-write the expression as,

$$D^{-\frac{1}{2}}(D-A)D^{-\frac{1}{2}}z = \lambda z \quad (1.5)$$

Using the transformation, $z = D^{-\frac{1}{2}} y$

It is significant to mention here, solution is equivalent to the solution to №, where № is termed as the normalized affinity matrix.

$$№ = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} z = \mu z \quad (1.6)$$

C. Color Space Conversion

Color space conversion is the first step in image analysis and applied before running Normalized Cut (NCut) algorithm for cell segmentation. We are converting the microscopic images into various color spaces (RGB to HSV, RGB to YCbCr and RGB to NTSC). Images are usually stored and displayed in the RGB space. However, to ensure the isotropy of the feature space, a uniform color space with the perceived color differences measured by Euclidean distances should be used. Here we are comparing the performance of NCut algorithm in four color space (RGB, HSV, YCbCr and NTSC).

D. Segmentation using Normalized Cut algorithm over the converted color space

NCut is an unsupervised segmentation technique that does not require initialization and approaches the segmentation problem as a graph-partitioning problem. NCut is based on a global criterion; and it maximizes both the total dissimilarity between the different groups and the total similarity within the groups.

RGB:

It is a hardware-oriented model and is well known for its color-monitor display purpose. The NCuts on RGB shows encouraging results but it is also corrupted with lots of spurious segments which make the system not suitable for automated methods.

HSV:

Hue (H) is a color attribute that describes a pure color, while saturation (S) defines the relative purity or the amount of white light mixed with a hue; value (V) refers to the brightness of the image. This model is found to be most suitable. In general, HSV is widely used for image processing applications but the basic implementation of NCut is based on RGB color space. The results here clearly illustrate an optimized performance of NCut in HSV for microscopic images.

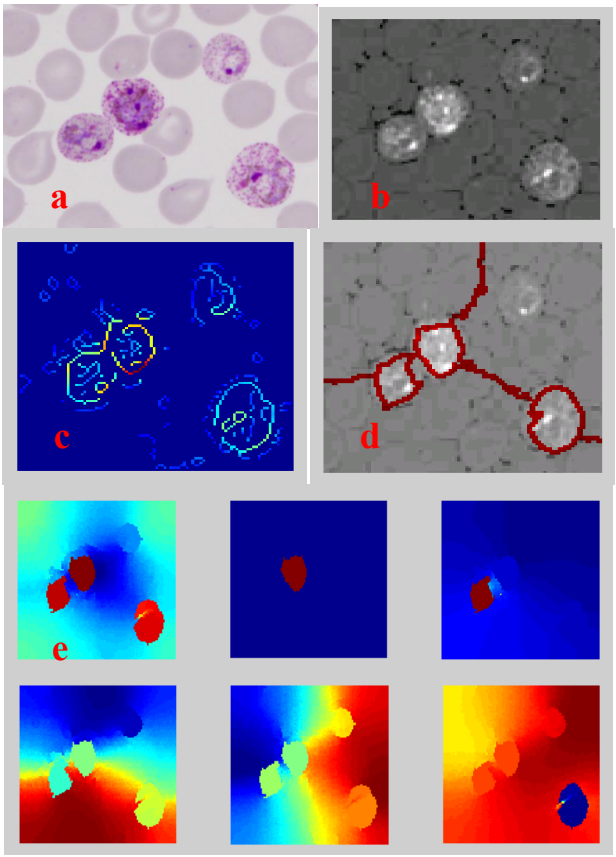

Fig 1.The images* show segmentation in HSV color space using nbsegment=6 (a)Image grabbed from Microscope, (b) Brightness Image, (c) Displaying Edges,(d) Displaying the Segmentation,(e) Displaying the Eigenvectors

YCbCr:

This is yet another hardware-oriented model. However, unlike the RGB space, here the luminance is separated from the chrominance data. The Y value represents the luminance (or brightness) component, while the Cr and Cb values, also

known as the color difference signals, represent the chrominance component of the image. The YCbCr color space is tested here to evaluate performance of the system on a Digital Signal Processor (DSP) based platform for integration with the tele-pathology unit developed by our group.

NTSC:

The components of the NTSC color space are Y (the luminance component), I (the cyan-orange component), and Q (the green-purple component). It performs sub-optimally in multiple cell scenarios, hence not investigated further.

E. Adjusting Segments for determining smallest Eigenvectors

The algorithm uses the given features to solve (1.4) for eigenvectors with the smallest eigenvalues. Then it uses the eigenvector with the 2nd smallest eigenvalue to bipartition the graph and finds the splitting point that minimizes NCut. To decide whether to divide the current partition it will check stability and NCut threshold. If necessary it recursively repartition the segmented parts.

F. Evaluation of Segmentation Performances

For the purpose of evaluation, the images were hand segmented with feedback from pathologist and medical students. The hand segmented images were used as standard labeled references and the segmented images are compared against it. In the present experiment we considered the Rand Index (RI) [13]. RI measures the fraction of the measures where the clustering agrees for both segmented and reference image. The index is bound between 0-1 with the upper bound denoting perfect match. The method however, works for binary images only. The segmented images (S) were obtained by converting them from respective color space to binary. The original images were converted to binary and the reference image (R) was extracted. Let i and j range over all pairs of pixels where $i \neq j$, then each pair falls into one of four categories: (p) $Ri = Rj$ and $Si = Sj$, (q) $Ri \neq Rj$ and $Si = Sj$, (r) $Ri = Rj$ and $Si \neq Sj$, (s) $Ri \neq Rj$ and $Si \neq Sj$. If we let p, q, r, s refer to the number of pairs in its corresponding category, then the Rand index is defined as:

$$\mathrm{RI}\,(\mathrm{R},\mathrm{S}) = \frac{p+s}{p+q+r+s} \qquad (1.7)$$

The index was calculated for all color spaces for the different developmental stages of the parasite. For better visualization we have scaled up the values between 0-100 while plotting (figure 3) by multiplying by a factor of 100.

III. Results and Discussion

The DPDx image library has been used for testing and analysis of algorithms. This database is made available publicly by Center for Disease Control and Prevention. The Plasmodium Vivax dataset (37 images), contains images from both thick and thin blood smears for all stages of development of the organism (Trophozoit, Gamatocytes, Ookinete and Schizonts). The Gamatocyte and the Ookinete stages being outside the human body are excluded from our study. In the present experiment we have segmented the trophozoits (ring form and developmental stages) and Schizonts. The developmental

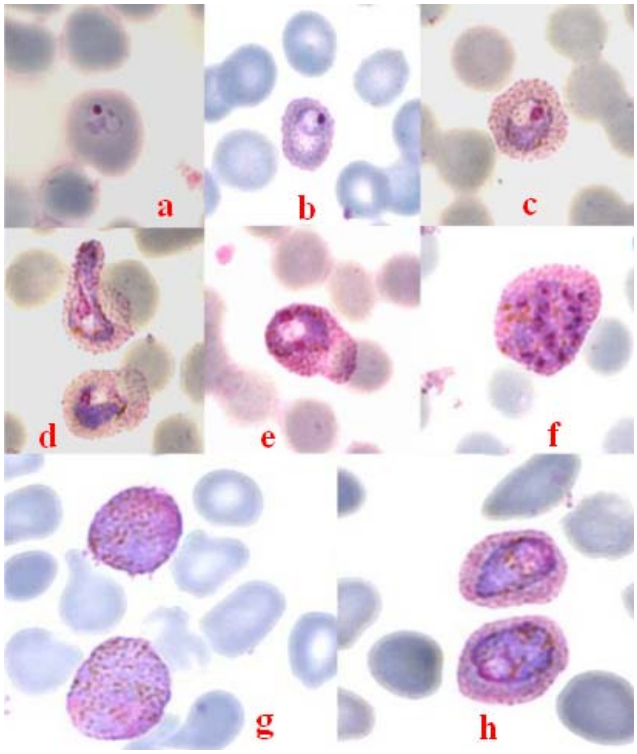

Fig 2. Images of various stages of development of malaria parasite within human blood grabbed using bright field microscope**

stages are illustrated in figure 2 (a-h), the segmentation results of the same are given in figure 4. The observations indicate that the results obtained using the HSV color space gives the most accurate results. The input images containing 1- 4 positive protozoa infected cells are taken into consideration. We also tested the algorithm on images having multiple overlapping RBCs, the same yields similar results under those conditions as well. The RGB performance seems to go down with complex multi-cell scenes but HSV is observed to be more stable and robust even with alteration of cells in background and in absence of malarial cells (figure 3).

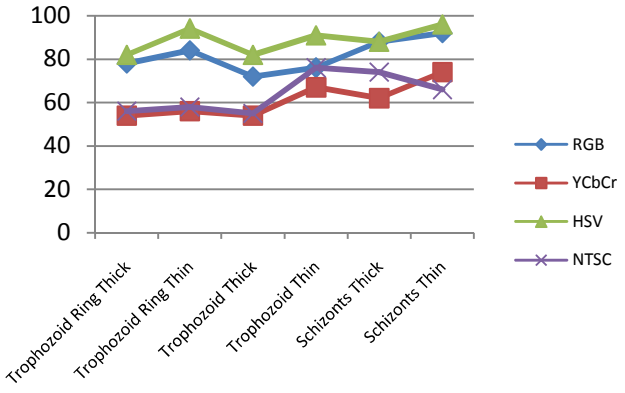

Fig 3 The comparative performances of the segmentation using different color spaces at various stages parasitic proliferation using scaled RIs

The system was also tested for finding the most optimal number of segments for automated detection. On a statistical study it is seen that the 3 repartitioning is the optimal showing

maximum true positive detections. However, cells with three or more than three number of infected cells are better delineated in 6 repartitioning.

The system when applied over various stages of parasitic development i.e. schizont from sporozoits, it is able to detect traces of parasite into stage 'c' (2-3.5 days of incubation) and generate perfected segmentation and detection by the end of stage d (5-6.5 days of incubation) (see fig. 4). Thus this method can be used sufficiently in sync with the present methods of pathological detection techniques. It is, however, to be noted that the study is preliminary and detailed clinical trials or pilot studies have not been concluded. The algorithm performs better than other algorithms commonly used in segmentation of blood smears, viz. morphological image processing and region growing algorithms. In NCut framework, finding the best partition is equivalent to computing eigenvectors, so the algorithm is efficient. The algorithm often gives a more precise and faster convergence rate than the algorithms mentioned. The drawback, however, is that it is a segmentation strategy based on global criterion so unintended noises (e.g. Patches cause by accumulation of strains) can seriously render the segmentation results inaccurate. The final output is seen to improve remarkably in these cases if we introduce some pre-processing routines, but

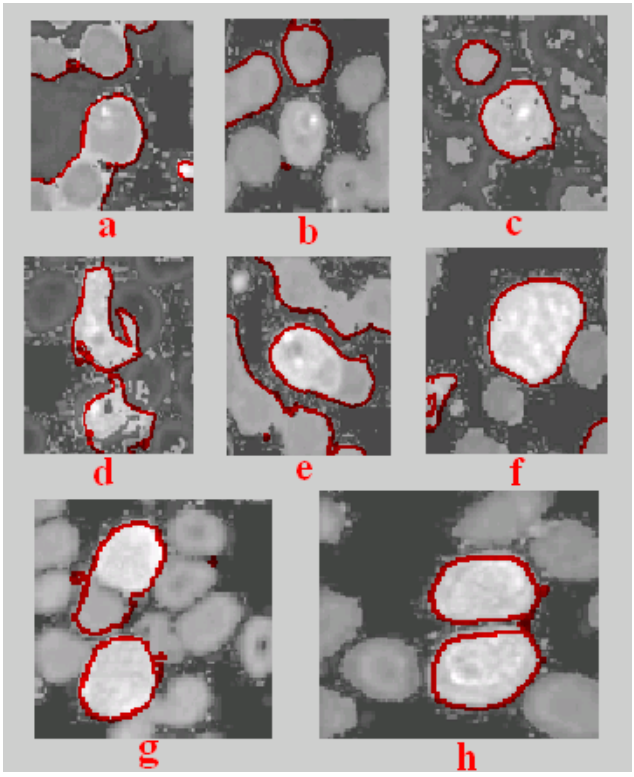

Fig 4. The segmentation of corresponding stages of parasitic proliferation with reference to various stages of development of malaria parasite

again, the same tends to increase the screening time per slide. The system performance can easily be optimized on DSP Platforms thus helpful in development of embedded devices that can be used in rural outreach centers.

IV. Conclusion

This paper shows a computationally efficient method for segmentation and characterization of malarial parasites from peripheral blood smear images. The results indicate that the performance of the NCut algorithm is best in HSV color space. The work is tested primarily based on feedbacks from pathologists and post-graduate medical students, efforts have been taken to quantify the same using Rand's Index. The applicability of the method can be noted in telepathology in areas lacking trained pathologists. However, scope of further improvement in the results can be achieved using mean shifts segmentation as a pre-processing step for NCut algorithm [14]. Use of other indices and non-linear measures for evaluating segmentation performance might be able to provide more insight into pathogenesis and detection.

References

[1] The Malaria Control Programme, World Malaria Report, World Health Organization, 2008.

[2] Jianbo Shi and Jitendra Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. of Pattern Recog. & Machine Intl*, August 2000.

[3] Jianbo Shi, Serge Belongie, Thomas Leung, Jitendra Malik, "Image And Video Segmentation: The Normalized Cut Framework", *IEEE Int'l Conf on Image Processing*, 1998

[4] Silvia Halim, Timo R. Bretschneider, Yikun Li, Peter R. Preiser and Claudia Kuss, "Estimating Malaria Parasitaemia from Blood Smear Images," Nanyang Technological University, Singapore.

[5] A. Fitzgibbon, M. Pilu and R.B. Fisher, "Direct Least Square Fitting of Ellipses," *Pattern Analysis and Machine Intelligence*, vol. 21 no. 5, 1999, 476-480.

[6] Florian Luisier, Thierry Blu and Michael Unser, "A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding," *IEEE Trans., Image Processing*, vol. 16 no. 3, 2007, 593-606.

[7] N. S. P Kong and H Ibrahim, "Color Image Using Brightness Preserving Dynamic Histogram Equalization," *IEEE Trans., Consumer Electronics*, vol. 54 no. 4, 1963-1968, Nov.2008

[8] Cecilia Di Ruberto, Andrew Dempster, Shahid Khan, and Bill Jarra, "Morphological Image Processing for Evaluating Malaria Disease," IWVF4, LNCS 2059, pp. 739{748, 2001

[9] Tomazevic D., Likar B. and Pernus F., "A comparison of retrospective shading correction techniques," Pattern Recognition, 2000. Proceedings of 15th International Conference on, Volume: 3, 3-7 Sept. 2000 Pages 564 – 567.

[10] David A. Forsyth and Jean Ponce, "Computer Vision A Modern Approach," Pearson Education.

[11] Halim, S.; Bretschneider, T.R.; Yikun Li; Preiser, P.R.; Kuss, C.; , "Estimating Malaria Parasitaemia from Blood Smear Images," *Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on* , vol., no., pp.1-6, 5-8 Dec. 2006

[12] Makkapati, V.V.; Rao, R.M.; , "Segmentation of malaria parasites in peripheral blood smear images," *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* , vol., no., pp.1361-1364, 19-24 April 2009

[13] L.P. Coelho, A. Shariff and R. F. Murphy," Nuclear segmentation in microsope cell images: A hand-segmented dataset and comparison of algorithms," *Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging* (ISBI 2009), p. 518-521

[14] Wenbing Tao; Hai Jin; Yimin Zhang; , "Color Image Segmentation Based on Mean Shift and Normalized Cuts," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* , vol.37, no.5, pp.1382-1389, Oct. 2007

*Image courtesy of Drs. JoAnn Sullivan and William Collins, Division of Parasitic Diseases, Centers for Disease Control and Prevention.