DEVELOPMENT AND VALIDATION OF AUTO-DETECTION TOOL FOR MALARIA PARASITES

[1]Dr. Narendra Mustare, [2]Mrs.Kaveri
[1]Professor, [2]Assistant Professor
[1]Department of Electronics and Instrumentation Engineering
[1]CVR College of Engineering, Vastunagar, Mangalpalli, Hyderabad, India
[2]Research Scholar (VTU), Department of Bio-Medical Engineering, KBN College of Engineering, Gulbarga, India

Abstract: A main aim of this research work is to develop and validate an automatic tool for parasite detection in thick blood films. Malaria is still a major health problem but the diagnosis has not improved and it is at the level of technicians due to the lack of Medical resources. An automatic tool developed here would be a useful filter for the specialist to test technician diagnosis. This paper presents the development and validation of an algorithm to detect parasites in thick blood films. The algorithm is trained by selecting a number of correctly classified pixels from each tissue, which are then used as input to a statistical classifier based on the k-Nearest Neighbor (k-NN) decision rule in the RGB color space. A few samples of the classes to segment, i.e. red cells (RC), parasites (PA) and background (BG) are picked to obtain a representation of the pattern space. The k-NN rule classifies a given pixel within the category the most heavily represented among its k-Nearest Neighbors. Validation was performed on a group of six images. Parasite was detected in the entire set of test images and the number of misclassified pixels was observed to amount to 11.41 %.

Index Terms **- Malaria, k-NN classifier, segmentation**

I. INTRODUCTION

Malaria is a one of the leading cause of morbidity and mortality in all countries, with an estimated infections of 100 to 300 million per year worldwide and 1 to 2 million deaths [1]. It is caused by any of the four different species of Plasmodium parasite, vivax, ovale, malariae and falciparum. Disease is transmitted via the bite of an infected female of the Anopheles mosquito. Diagnosis is mainly based on a complete clinical history, however parasite confirmation goes through laboratory tests. The parasite that causes malaria can be seen on a blood smear under a microscope, yet exhaustive microscope evaluation is highly time consuming, not precise enough and depend on the skill of the observer [5]. This paper presents the development and validation of an original automatic method to detect parasites in thick blood images. Firstly, a few samples of correctly classified pixels are manually selected to obtain a good representation of the different tissues in the image to classify. Then, these pixels are used to train a statistical classifier that uses the RGB color components as the space of patterns. The classifier used here is the k-Nearest Neighbors (k-NN) classification rule [2, 3].

II. METHODOLOGY

Image Acquisition

Thick films were obtained by placing a drop of blood in the middle of a clean microscope slide, then smeared with the corner of a second slide and colored with Giemsa stain. Giemsa stains red cells in a deep red color, while parasites are trained in a pink diffuse color (see Figure 1). Images were acquired with a Nikon Coolpix 950 (640 × 480 to 1600 × 1200 pixels) high resolution digital camera coupled to an Leica DM LS microscope Use of intermediate lens and a × 20 power objective yielded a total magnification of ×900. Optical image size was 256 ×192 μm^2 for a 1 600 × 1 200 pixel image, resulting in a total resolution of 0.0256 μm^2/pixel.

Image Analysis

The k-NN decision rule

Given the knowledge of N prototype patterns (vectors of dimension Σ) and their correct classification into M classes, the k-NN rule assigns an unclassified pattern to the class that is most heavily represented among its k neighbors in the pat-tern space (under some appropriate metric). For the k-NN rule, the risk is bounded by $(1 + \frac{1}{k})R^*$ [2], so that theoretically R^* can be approached as close as desired by increasing k. Practically, though, this is limited by the requirement that $k/N \rightarrow 0$, that is to say, the more k is chosen larger the more number of training samples are needed.

k-NN classification in RGB color space

Despite these remarkable statistical properties, k-NN classification is often discarded because of its high computational cost. Implementing the k-NN rule with a brute-force method to classify F pat-terns with N prototypes requires F ×N distance computations and o(F × N × log(N)) comparisons. Different approaches have been devised to improve algorithm performances. The approach that is

better suited to our problem, was first pro-posed by Warfield [6] who applied it to double echo spin echo MR image classification. In such an application, the number of possible patterns is smaller than the number of patterns to classify, so that it becomes efficient to pre-compute a lookup table (LUT) for every possible pattern, then to classify patterns by accessing this lookup table. The computation of this LUT is essentially a k-distance transformation problem. Distance transformations are algorithms that compute for every pixel of an image the distance to the nearest pixel of a given object. Once the prototype pixels are selected, information is propagated by using a k-Euclidean distance transformation. The pattern space is scanned by using ordered propagation, starting from the prototype patterns, by order of increasing Euclidean distance [4]. The strict respect of the increasing distance order is achieved by bucket sorting the propagation front of pixels, which insures that no unnecessary computations are performed.

Choice of parameters

The parameters of this method are the number F of training samples required, the number k of nearest neighbors used by the k-NN rule and the size $2^{D.B}$ of the pattern space where D is the number of dimensions, i.e 3 for the RGB space, and B the number of bits used to quantify each color. Optimal k, F and B were established by using a 10-fold cross-validation in a sample of 1000 pixels per class.

Segmentation of parasites

Automatic algorithm was implemented in user interface (GUI) written in TCL/TK. Validation of this method was performed on eight images randomly selected from three different malaria cases, used for an unrelated study. Evaluation was performed two folds: on the one hand an expert in the domain determined visually if the parasite has been detected in any of the eight group of images and on the other hand, an error classification was estimated for the parasite class by comparing the number of misclassified pixels to the total number of pixels in the image and summed over the whole set of images.

III. RESULT

A subsection of a typical blood thick film is illustrated in bottom and lower panels (76.9 × 108.2 μm^2) of Figure 1. This coloration stained para-sites (PA) in violet, the red cells (RC) in a pale red and the background (BG) was observed to be close to a pink. Segmentation, identification and classification of the three classes were achieved by using 205 (PA), 212 (RC) and 265 (BA) correctly classified pixels. The resulted segmented image is displayed, with a k value of 9, in lower panel. Pixels are colored in white, black and a gray level for the PA, BA and RC classes, respectively.

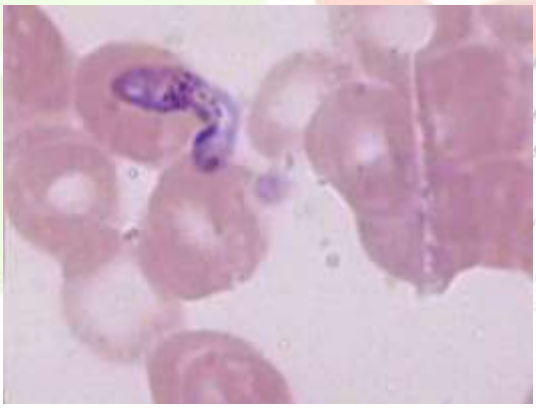

Figure.1: A Typical thick blood film

Parameter selection

The main parameters that will influence the classification results are the number of training samples, the number k of neighbors considered by the k-NN rule, and the quantification of the LUT in the RGB color space, i.e. the number of bits B used to describe each color The importance of the number of neighbors used to train the classifier is illustrated in Figure 2, where error rate R is displayed. Samples included pixels randomly picked from a particular tissue. Note that error steadily maintains down after a training with k = 9.

Table: I Error response

Number of samples needed to train the classifier	Error Rate in (%)
100	7.1
200	6.5
300	1.8
400	0

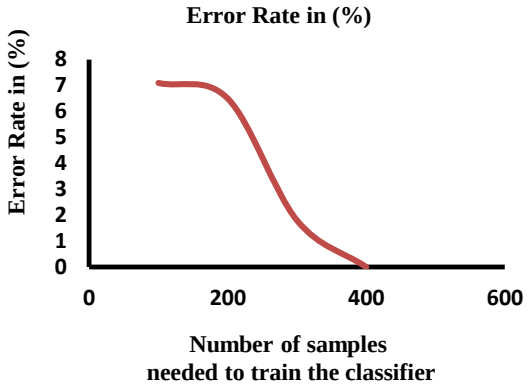

Figure 2: significance of no. of neighbors (k) used decision rule

The number of samples needed to obtain a good representation of the RGB space is shown in Figure 3. Observe that error goes down after 300 samples.

Table: II: Error response for decision rule

No. of neighbors used for decision rule	Error rate (%)
2	1.3
4	1
6	0.8
8	0.63
10	0.43
12	0.43
13	0.43
14	0.4
16	0.39
18	0.38

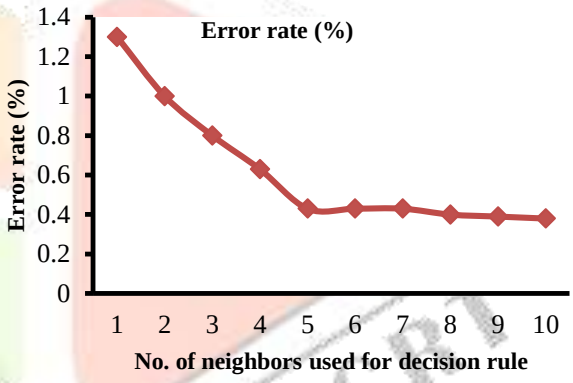

Figure.3: Error response for decision rule

Table: III: Comparison of error rate w.r.t different bit color bits

No. of neighbors used for decision rule	Error rate (%)				
	3 bits	4 bits	5 bits	6 bits	7 bits
1	55	40	30	27	9
2	55	40	29	26	8
4	54	39	28	25	7
6	54	38	27	24	6
8	54	37	26	23	5
10	54	36	26	22	4
12	54	35	24	21	3
14	52	35	23	19	3

16	52	35	22	15	2
18	51	34	21	14	2
20	51	34	20	12	2

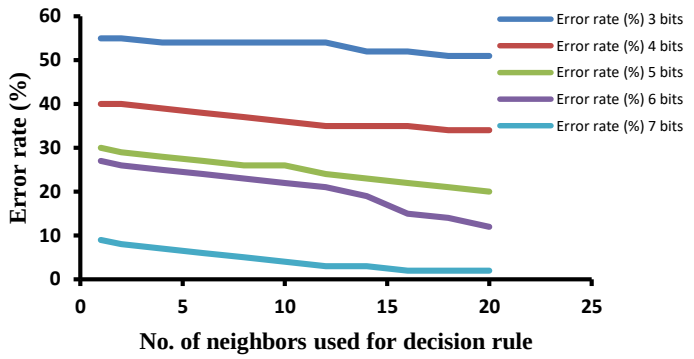

Figure.4: Comparison of error rate w.r.t different bit color

Finally, Figure 4 shows the error rate obtained by using bit quantification, i.e. classifying from two to the entire set of eight bits. Note error changes little with six or seven bits. For k = 5 and B = 6, the size of the k-DT space is $5 \times 8^6 = 1.31 \times 10^6$. CPU time increases linearly with k and is multiplied by 8 for each ex-tra bit used to quantify the colours. Practically, we routinely use k = 9 and B = 6 samples. Best results k were obtained when $k \geq 9$ and $B \geq 6$. An expert in the domain evaluated the performance of the method. Parasite was always detected on the entire set of eight images, irrespective of which image was used for training. Also, when counting on the whole set, the percentage of misclassified pixels was 11.41 %. The time for processing was also considered, while a LUT creation took about 3 seconds, processing an individual image requires less than a second on a Pentium IV, 800 MHz computer to process a 1 600 × 1 200 image.

IV. Conclusions

This research work explains an automatic method to segment and detect parasites in thick blood films. The algorithm allows consistent and reproducible measures of the deferent classes present in thick blood films. A minimal user help is needed to select the sample pixels required to train the statistical classifier. This automatic approach works at the level of the color space by simple distance criteria defined on the k-Nearest neighbors (k-NN) of any particular pattern. The k-NN rule is essentially a technique for local classification in the pattern space that is used here to determine whether or not a particular pixel belongs to one or another tissue. Thus, there is no a defined model of the tissue color properties i.e. no prior knowledge of the probability density function associated with any particular population of pixels. Moreover, classification is based on the final state of the acquired image and, in outcome no special emphasis is placed on any step of the capturing procedure. This rule requires that $_N{}^k \rightarrow 0$ to insure that there exists enough prototypes to achieve good local approximations of the diverse probability density functions associated with the deferent tissues. Once the different parameters selected by the user give satisfactory results then, the method can be considered as fully automatic and the chosen settings can be applied to a larger set of images in the same batch.

V. Acknowledgment

The authors are very much thankful to the management of CVRCE, Hyderabad and KBNCE, Gulbarga for providing research facility and encouragement to carry out the work.

References

[1] The Malaria control program, “World Health Organization,” Third World Net-work Features, Health Canada, The Center for Disease Control and Prevention.
[2] Cover, T.M. and P.E. Hart, “Nearest neigh-bor pattern classification, ” IEEE Trans. on Inf. Theo., IT-13, 1967, pp. 21–27 .
[3] Fix, E. and J.L. Hodges, “Discriminatory analysis, non-parametric discrimination, ” , USAF School of Aviation Medicine, Ran-dolf Field, Texas, Project 21-49-004, Rept. 4, 1951, Contract AF41(128)-131.
[4] Cuisenaire, O. and B. Macq, “Fast k-NN classification with an optimal k-distance transformation algorithm, ” Proc. of the 10th European Signal Processing Confer-ence, September, 2000, pp. 1365–1368
[5] Romano, L.A. and M.D. Ferder and I.Y. Stella and F. Inserra and L.F. Ferder, “High correlation in renal tissue between computed image-analysis and classical morpho-metrical analysis, ” J. Histo-technol. , 19, 1996, pp. 121–123
[6] Warfield, S., “Fast k-NN classification for multichannel image data, ” Pattern Recog. Lett., 17, 1996, pp. 713–721