International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue (4) July, Technovision-2014, ISSN 2249–071X

Technovision-2014: 1st International Conference at SITS, Narhe, Pune on April 5-6, 2014

A Fast Segmentation Method for the Recognition of Acute Lymphoblastic Leukemia using Thresholding Algorithm

Mrs. Trupti A. Kulkarni	Prof. D. S. Bhosale	Dr. D. M. Yadav
E & TC Department, JSPM's Bhivarabai Sawant Inst. of T & R (W), Pune, India	E & TC Department, JSPM's Bhivarabai Sawant Inst. of T & R (W), Pune, India	E & TC Department, JSPM's Bhivarabai Sawant Inst. of T & R (W), Pune, India
truptijoshi21@yahoo.in	bhosale.dilip@rediffmail.com	Email: principalbsiotr@gmail.com

***Abstract* – In medical diagnosis system, vast number of diseases are diagnosed on the basis of counting and classification of blood cells. Acute lymphoblastic leukemia (ALL) is the most common type of blood cancer of white blood cells (WBCs) in children below 7-8 years. It can be fatal if diagnosed late or untreated. ALL cells are abnormal lymphocytes called blast cells or lymphoblast. Careful microscopic examination of stained blood smear or bone marrow aspirate is the best way to diagnose leukemia. There are various techniques such as cytochemistry, FISH-Fluorescence In Situ Hybridization, cytogenetic analysis etc. are available for specific leukemia detection. These tests are done manually so time consuming as well as costly. Therefore low cost and efficient solution is automatic detection and analysis of microscopic blood images. This paper presents complete and fully automatic method for WBCs identification and classification of blasts from microscopic images. The proposed method is to segment normal and ALL lymphocytes into two parts: nucleus and cytoplasm. This is done by using Otsu's thresholding algorithm. The MATLAB is used to develop the whole work.**

***Keywords* – Acute Lymphoblastic Leukemia, Public Image Dataset, WBC Segmentation, FAB (French-American-British) Classification.**

I. Introduction

Leukemia is a disease of unknown cause where the bone marrow produces large numbers abnormal cells [1]. The diagnosis is based on the blood tests and bone marrow tests. Both of the tests can help doctors decide on the best choice of treatment. It is a bone marrow disorder that arises when abnormal white blood cells replicate itself. Acute Lymphoblastic Leukemia is basically associated with the lymph nodes. According to FAB (French – American – British) classification, leukemia is classified as ALL- Acute Lymphoblastic Leukemia, AML- Acute Myeloid Leukemia, CLL- Chronic Lymphoblastic Leukemia, CML- Chronic Myeloid Leukemia [2]. This classification is based on the cell type involved and the clinical course.

This work focuses on ALL, a serious illness caused by abnormal growth and development of lymphocytes which is one of the types of white blood cells. The early and fast detection is very much important in providing the appropriate treatment as there are many types of leukemic cancer. If some symptoms are found on clinical examination, doctors suggest for complete blood count test (CBC) [3]. If the white blood cell count is found enormously increased, the patient is suggested to perform a bone marrow biopsy.

Thus there is a need of study of morphological bone marrow and peripheral blood slide analysis. For the classification of normal, abnormal and leukemic cells, hematologists observe some cells under a microscope. The aim is to find out if the abnormalities are present in the nucleus or cytoplasm of the lymphatic cells. The presence of large of number of blast cells in peripheral blood is a significant symptom of leukemia [4]. In the present work ALL is considered. The main objective is to classify a lymphocyte as a normal or a blast cell.

II. Literature Review

Over the years, many works have been conducted in the area of general segmentation methods of blood smear images for the detection of leukemia. An automated system for WBC segmentation by applying active contours or color-based segmentation is proposed in [5] [6]. Teager Energy-Based Segmentation is used in [7]. Watershed algorithm and optimal thresholding is used for WBC segmentation in chronic lymphocytic leukemia (CLL) detection in [8]. A two step segmentation method by applying the HSV color model is used in [9]. Segmentation of WBC cells using active contours model is proposed in [10]. A complete classification system to detect the acute leukemia from blood image based on morphological features and using gray level images is used in [11]. The cell type classification by using an artificial neural networks and morphological operators is proposed in [12]. ALL-Acute Lymphoblastic Leukemia recognition based on different approaches such as the analysis of gene expression is presented in [13]. A high throughput algorithm for leukemia cell population statistics on a hemocytometer is presented in [14]. Fuzzy based blood image segmentation for automated Leukemia detection is presented in [15]. In the present paper we propose a segmentation method for automatic analysis of

peripheral blood smear examination. The segregation of nucleus from other blood components is performed here.

III. Materials and Methods

Dataset Description

The image dataset used for the processing is used which are digitized using digital microscope captured under 100X magnification. The slides which are prepared are blood smear images and stained using Leishman stain for better clarity of blood components [1]. The dataset is available for download at http:// en.wikipedia.org/ wiki/Blood [16].

System Description

The system follows the procedure using following components to segment and classify the microscopic image: (1) Preprocessing, (2) Segmentation, (3) Feature Extraction, (4) Classification. The blood smear image consists of red blood cells (RBCs), platelets along with various types of WBCs. Our main goal is to segregate lymphocytes from all these blood components for further processing. These lymphocytes consist of round, blue nucleus with scanty cytoplasm. In particular, lymphocytes present a regular shape. It consists of a compact nucleus with regular, continuous edges, whereas lymphoblast is of irregular shape and sometimes shows small cavity in the cytoplasm and spherical particles in the nucleus. The whole process is schematized as given in figure1.

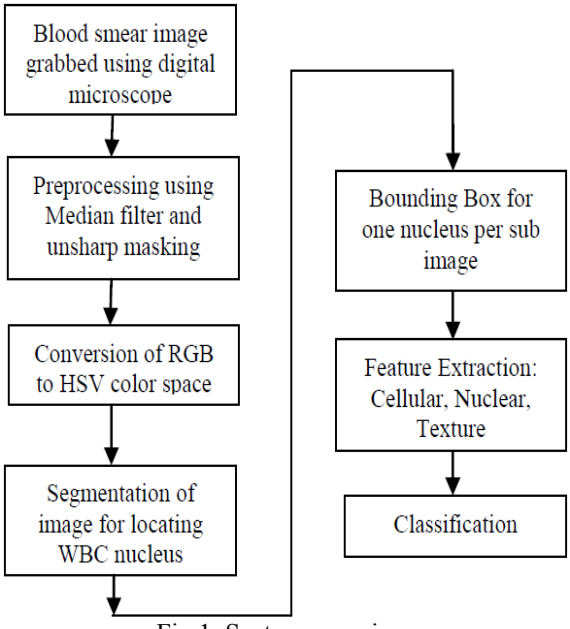

Fig.1. System overview

1. Pre- processing

The image acquisition may contain noise due to staining or electronic components present in digital microscope. All the images are subjected to Median filtering to remove the noise and unsharp masking is used for better enhancement of image.

2. Color conversion

The images which are generated using a digital microscope are color images i.e. in RGB color space which is quite difficult to segment. To make easy and fast segmentation, RGB color space is converted to HSV color space. This conversion reduces the color dimension from three to two. The HSV color space consists of Hue, Saturation, and Value components. For better distinction of cells in the foreground, we have selected Saturation plane as it shows better contrast by observing the results. A redistribution of image gray levels is essential for subsequent segmentation process. Therefore, a histogram equalization or a contrast stretching is used at this stage.

3. Image Segmentation

Image segmentation is performed using Otsu's thresholding method [17]. In this method, the threshold value is selected automatically to minimize the interclass variance of the black and white pixels and the Saturation threshold plane is converted to binary image. It is necessary to remove background components and artifacts from the image. Background removal can be performed with some arithmetic operations. To clean up the background, we have used the operation called area opening. It allows to delete the unwanted cells and objects with a size smaller than the structuring element. Once the image with only leukocytes is obtained, it is now easier to segment further for overlapping cells and their separation. There are various methods as shown in the literature. In the present method, we have used labeling technique which labels all of the connected components in the binary image and returns the number of objects it finds in the image in the output value. In short, it labels all isolated regions with separate numbers. The labeled cells are now nuclei of WBCs are now individual cells and ready to find out the features and hence leukemia detection.

4. Sub Imaging

The peripheral blood smear image consists of cluster of blue nuclei. To extract specific features for accurate detection of leukemia, it is important to extract the individual nucleus for classifying it as a blast cell. Therefore, sub imaging consisting of single nucleus per image is essential. We have used Bounding Box technique [18].

5. Image Segmentation algorithm

The present work is implemented using the following algorithm.

1. Input color blood smear Leishman stained image to the system.
2. Pre-process the image for removing noise using Median filter and for better enhancement of image use unsharp filter.
3. Convert color image into H, S, V color space.
4. Select H, S, V planes and plot histograms separately.
5. Select S plane as it gives better distinction of contrast between foreground and background for postprocessing.
6. Obtain threshold value for this plane.

7. Apply Otsu's optimal thresholding algorithm on this S plane to get threshold S plane and convert it into a binary image.
8. Morphological opening is used to remove small pixel groups.
9. By applying labels to the cells, we can make analysis of overlap and non-overlap cells.
10. Overlapped cells are removed using regionprops.
11. Isolated and individual cells are further used to find out the features. Hence are useful to identify whether cells are leukemic or normal.

It is well observed that this segmentation method yields better and faster results than other methods.

6. Feature Extraction

An image feature is a distinguishing primitive characteristic or attribute of an image. Image features are of major importance in the isolation of regions of common property within an image (image segmentation) and subsequent identification or labeling of such regions (image classification) [19]. The transformation of the input data into the set of features is called feature extraction [20]. In the present paper, nuclear features, shape features are extracted from the segmented nucleus image.

i) Shape Features:

According to doctors the shape of the nucleus is an important feature for describing the cell as a blast. The various features which are mainly used by doctors are evaluated here. These are regionally based and boundary based features obtained from a binary image which is equivalent to the input color image. The quantitative evaluation of each cell is done independently using these features.

- Area: The area is the total number of nonzero pixels which is the object in the image.
- Perimeter: The number of pixels in the boundary of the shape.
- Eccentricity: It is the roundness of the object, with the value 0 up to 1, a circle is perfectly round and has an eccentricity 0, while a line segment has eccentricity 1.

$$\text{Eccentricity} = \frac{\sqrt{a^2 - b^2}}{a}$$

- Form factor: The form factor is a function of area and perimeter of the object under consideration.

$$\text{Form factor} = \frac{4 * \text{pi} * \text{area}}{\text{Perimeter} * \text{perimater}}$$

- Solidity: Solidity is defines as the ratio of actual area and convex hull area. This is an important feature for blast cells classification.

$$\text{Solidity} = \frac{\text{Area}}{\text{Convex Area}}$$

ii) Bounding Box and Contour Signature:

It is the measurement of the rough boundary of a nucleus. It is an important feature to label the nucleus as blast cell. In the present work, we concentrate on finding the centroid and by using a Euclidean theorem the distance between the boundary x, y coordinates and the centroid is calculated.

The Euclidean distance is given by

$$Distance = \sqrt{(x-j)^2 + (y-k)^2}$$

Where (x, y) is the coordinates of the origin and J is the number of columns and K is the number of rows. Bounding Box plotting and Euclidean distance measurement is explained in following algorithm.

1. The single cell is cropped out of labeled image. Bounding box and the centroid is plotted.
2. The bounding box is a rectangle enclosing all points of particular cell. x1, y1 and x2, y2 are the points of boundary box.

Where x2 = x1 + x_width and y2 = y_width

3. Distance of a peripheral point separated by each 10° from centroid is calculated using above Euclidean formula.
4. The distances are stored in 36 length matrix.
5. These 36 distances of one cell describe contour signature of that particular cell.
6. Standard deviation describes the roundness of that particular cell.
7. If the cell is of regular shape, standard deviation is less, otherwise it is larger.
8. The plot is shown below in the results.

iii) Texture Analysis:

Image texture is defined as a function of spatial variations in pixel intensities (gray values). The function Gray_Level Co occurrence Matrices (GLCM) is used in the present work. For the texture measurements, gray images of nucleus obtained after sub-imaging are used. Using this function, GLCM standard deviation is calculated.

IV. Experimental Results and Discussions

The present technique is applied to 10 images of blood smear. A microscopic image of size 347 X 395 is used and processed for evaluation. The various results obtained are shown below.

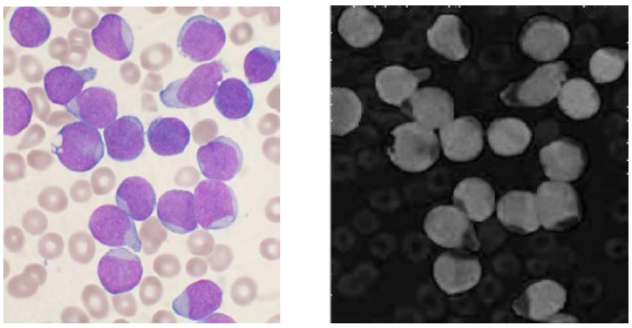

Fig.2. Original RGB image and S component image

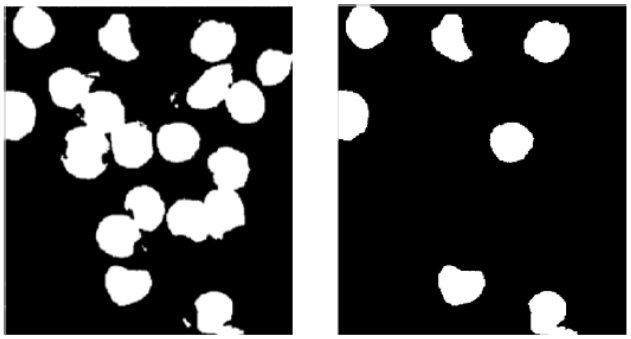

Fig.3. Saturation thresholded output and final segmented result

Fig.4. Separated Nucleus sub images result using Bounding box

Fig.5. Separated Nucleus sub images using Bounding box, Centroid located and Edge image using Canny edge detector

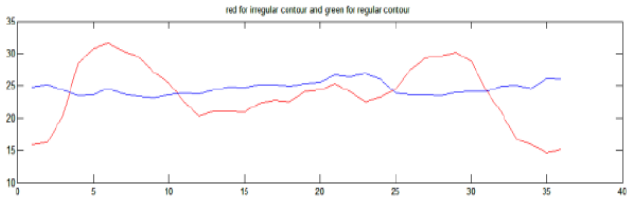

Fig.6. Sub images and contour signature plotted for the respective nucleus

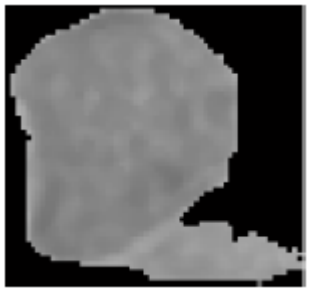

Fig.7. Texture analysis output image for a single cell

Table I: Results of Various Shape Measurements

Cells	A	P	S	E	F-F
1	1817	173.3970	0.9753	0.7745	0.7594
2	1965	165.4386	0.9810	0.4254	0.9022
3	1807	174.6102	0.9135	0.7212	0.7448
4	2038	180.2670	0.9501	0.5982	0.7881
5	1897	168.6102	0.9703	0.3813	0.8385
6	2113	176.6102	0.9684	0.4875	0.8513

Table II: Standard Deviation and GLCM-Standard Deviation

Cells	Std_Deviation	GLCM-Std_Deviation
1	4.74229408774867	193.2994
2	1.19391697759786	181.5236
3	4.62813637898772	203.4461
4	3.16180704580719	241.1813
5	1.07544239421758	172.7403
6	1.44770232101902	259.6359

In the present technique the results which are obtained are shown above. In the result, input image and saturation component image which is selected for processing is shown in figure 2. By using Otsu's thresholding algorithm the saturation threshold which is binary image and its final segmented output is shown in figure 3. Sub imaging of various nucleus is done using Bounding box technique for the analysis of individual nucleus is shown in figure 4. Cropped nucleus with centroid and edge detected using Canny edge detector is shown in figure 5. Figure 6 shows Contour signature of individual nucleus. Figure 7 shows texture analysis image resulted after processing the gray nucleus. This is obtained after computing GLCM-standard deviation. In this analysis, more standard deviation shows rough texture and less shows relatively smooth texture. Table I shows the various shape features obtained for segmented cells in the image output. In table I, A is area, P is perimeter, S is solidity, E is eccentricity and F-F is form-factor. These are calculated for individual cells. Table II shows the measurements of standard deviation and GLCM-standard deviation i.e it shows how we have calculated the shape of a lymphocyte or lymphoblast and for the texture analysis, we have used GLCM function to calculate GLCM-standard deviation. In the shape analysis, if the standard deviation is less, the cell is of regular shape otherwise irregular and in texture analysis, if it is less, the

texture is relatively smooth otherwise it is found to be rough.

V. Conclusion

The basic aim of the present work is the ALL slide image segmentation followed by feature extractions. We mainly considered shape features like area, perimeter, eccentricity, form factor etc. for the detection of lymphoblast i.e. leukemia. The feature contour signature is also considered for calculating the roughness of the boundary. This is important feature which can be further used for classification of cells as normal lymphocytes or lymphoblast. Results obtained encourage the future work to develop a robust segmentation system independent of stains used in blood smear images.

Acknowledgment

We are really thankful to Dr. Ramanan, Dr. G. Kulkarni, Dr. Nita Munshi, Ruby Hall Cancer Research Clinic, Pune for guiding us in the work.

References

[1] B. J. Bain, "A Beginners Guide to Blood Cells," Blackwell Publishing, 2nd edition, 2004.
[2] H. Theml, H. Diem, T. Haferlach, "Color Atlas of Hematology, practical Microscopic and Clinical Diagnosis," 2nd revised edition, Thieme publications, 2004.
[3] H. T. Madhloom, S. A. Kareem, Department of Artificial Intelligence, "A Robust Feature Extraction and Selection Method for the recognition of Lymphoblastic Leukemia," International Conference on Advanced Computer Applications and Technologies, 2012.
[4] C. Haworth, A. Hepplestone, P. Morris Jones, R. Campbell, D. Evans, M. Palmer, " Routine Bone Marrow Examination in the Management of Acute Lymphoblastic Leukemia of childhood", Journal of Clinical Pathology, 34: 483 – 485, 1981.
[5] Ongun G, Halisi U, Leblebicioglu K, Atalay V, Beksac M, Beksac S, "An automated differential blood count system," Engineering in Medicine and Biology society, Proceedings of the 23rd Annual International Conference IEEE, 2001.
[6] Ramesh N, Dangott B, Salama ME, Tasdizen T, "Isolation and two-step classification of normal white blood cells in peripheral lood smears," J Pathol Inform, 2012.
[7] Kumar BR, Joseph DK, Sreenivas, TV, "Teager energy based blood cell segmentation," IEEE 14th International Conference, 2002.
[8] Emad A Mohammed, MostafaM A. Mohamed, Christopher Naugler, BehrouzH. Far, "Chronic Lymphoblastic Leukemia Cell Segmentation from Microscopic Blood Images Using Watershed Algorithm and Optimal Thresholding," 26th IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2013.
[9] N. Sinha and A. G. Ramakrishnan, "Automation of Differential Blood Count," In Proceedings Conference on Convergent Technologies for Asia Pacific Region, 2:547 – 551, 2003.
[10] Atin Mathur, Ardhendu Tripathi, Manohar Kuse, "Scalable system for classification of white blood cells from Leishman stained blood stain images," J Pathol Inform, DOI:10.4103/2153-3539.109883, 2013.
[11] V. Puri and F. Scotti, "Morphological classification of blood leukocytes by microscopic images," IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA), July 2004.
[12] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA), July 2005.
[13] F. Scotti, "Robust segmentation and measurements techniques of white blood cells in blood microscope images," IEEE Instrumentation and Measurement Technology Conference (IMTC), April 2006.
[14] B. Prasad and W. Badawy, "High Throughput algorithm for leukemia cell population statistics on a hemocytometer," IEEE Biomedical Circuits and Systems Conference (BIOCAS), November 2007.
[15] S. Mohapatra, S. S. Samanta, Dipti Patra, S. Satpathi, "Fuzzy based Blood Image Segmentation for Automated Leukemia Detection, IEEE International Conference, 2011.
[16] http://en.wikipedia.org/wiki/Blood.
[17] Otsu, N, "A Threshold Selection Method from Gray – Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol.9, No.1, pp. 62 – 66, 1979.
[18] A. K. Jain, "Fundamentals of Digital Image Processing, Pearson Education," 1st Indian edition, 2003.
[19] William K. Pratt, "Digital Image Processing," 4th edition, 2007.
[20] The Wikipedia the Free Encyclopedia Website, http://en.wikipedia.org.