Malignant White Blood Cell Image Segmentation

Ahmed A.Abdul Hameed[1], Loay E. George[2]
Computer Science Department, College of Science, University of Baghdad, Iraq

Abstract— Leukemia is a type of cancer of blood or bone morrow which is characterized by abnormal increase in the number of immature white blood cells called "Blasts". Blood and bone morrow smears are usually inspected using microscope to properly identify the presence of leukemia and type of leukemia if it is located. To properly identify the disease, white blood cell should be isolated from the rest of the image and the components of white blood cell which are the nucleus and cytoplasm areas must be differentiated from each other. Our system first enhances the colors of the image sample then image segmentation is done using the proposed color segmentation method. The proposed segmentation method gives good results, the number of test samples used were 60 samples for both blood and bone morrow samples.

Index Terms— Computer aided diagnosis, image processing, image segmentation, morphological operators.

1 Introduction

Pattern recognition (PR) is the scientific discipline whose goal is the classification of objects into a number of categories or classes. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified **[1]**.

The advent of computers increased the demand for practical applications of PR, which in turn set new demands for further theoretical developments. As our society evolves from the industrial to its postindustrial phase; automation in industrial production and the need for information handling and retrieval are becoming increasingly important. This trend has pushed pattern recognition to the high edge of today's engineering applications and research **[1]**.

Within medical science, pattern recognition is the basis for computer-aided diagnosis (CAD) systems. CAD describes a machine dependent procedure that supports the doctor's interpretations and findings **[2]**. The major problem in medical field is to diagnose disease. Human being always make mistake and because of their limitation, diagnosis would give the major issue of human expertise. One of the most important problems of medical diagnosis, in general, is the subjectivity of the specialist. It can be noted, particularly in pattern recognition activities, that the professional experience should closely related to the final diagnosis. This is due to the fact that the result does not depend on a systemized solution but on the interpretation of the patient's signal **[3]**.

The primary goal of CAD is to increase the detection of disease by reducing the false negative rate due to observational oversights. The use of computer rather than a second human observer has the advantage of no increasing the demands on the pathologist (or trainer observer) pool. An important aspect of either approach is to increase disease detection without an undo impact on the recall and work up rates. In some applications of CAD, with its associated automated software tools, has the potential to provide workflow efficiencies. CAD algorithms are developed to search for the same features that a pathologist looks for during case review **[4]**.It has been proven that the capabilities of PR are so encouraging in many domains such as medical applications. Pattern recognition has the ability to learn by example which makes it very flexible and powerful in medical diagnosis. Nowadays, physicians combined the opportunity that is given by medical pattern recognition and their expertise to detect early stages of patient's disease **[5]**.

2 Related Works

Many researches have been conducted to study the problems of utilizing pattern recognition methodology in medical images processing. Part of the researches aimed to develop the methodology that capable of detecting white blood cells or identifying the existence of blood cancer using microscopic images.

Cseke **[6]** introduced a system consists of three components: First a novel simple algorithm for localization of white blood cells. The algorithm is based on a priori information about blood smear images. Second is a segmentation method applied for separating the different cell components using automatic thresholding. The thresholds are selected according to simple recursive method derived from maximizing maximizing the interclass variance between dark, gray and bright regions. Third, and finally, the segmented regions are smoothed by a set of morphological operations. Theerapattanakul **[7]** proposed a segmentation scheme that utilizes a benefit of active contour. Specifically, the binary image is obtained by thresholding of the input blood smear image. The initial shape of snake is then placed roughly inside the white blood cell and allowed to grow to fit the shape of individual white blood cell. The white blood cell is then separated using the extracted contour. Experimental results show that the proposed method can handle well with the problem of touching red blood cell. Qiongshui Wu **[8]** introduced multispectral imaging techniques. After a high quality image was acquired, the RGB values of each pixel were directly fed into a trained support vector machine (SVM) for classification, and then morphological binary operations were performed to correct the small error-classified regions. The experiments showed that the segmentation results are highly satisfactory and inspiring. It shows that the introduction of multi-spectral imaging analysis techniques into white blood cells detection is fruitful. Cheewatanon **[9]** introduced a new algorithm that consists of two models. Firstly, the mean shift filter is used to

remove noise. Secondly, a simple but effective region growing algorithm is suggested to segment the image. The experimental results show the excellent performance in both color spaces.

3 Proposed System Description

Our proposed method consists of several steps to properly segment the white blood cell from the digital image and then further segment the white blood cell into two regions: nucleus and cytoplasm area. With these two regions, the identification of leukemia can be achieved and leukemia type can be identified because each region contains information needed for properly stating the type of leukemia.

Below are the steps for our proposed segmenting method:

3.1 Image Loading

As a first task, the digital blood cell images are loaded. Typically, blood sample is first taken from the patient then the lab operator prepare the sample by adding dye solution, with regard to the percentage of the added dye solution to the sample and the concentration of the dye compound. Then the lab operator places the prepared sample in a microscope and uses an imaging device like camera mounted of the microscope to save a digital image of the sample. Both blood and bone morrow smears can be used in our segmentation method. Figure (1) shows samples on medical images to be used in our segmentation method.

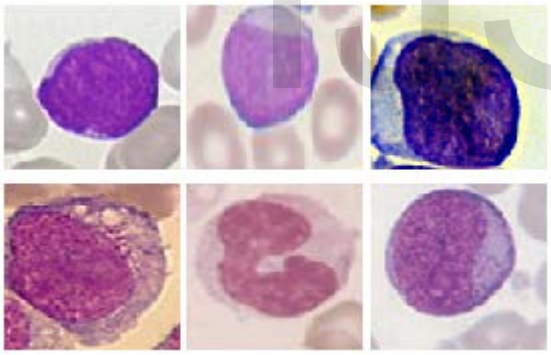

Figure 1: Medical image samples

3.2 Color Enhancement

As a first task, the digital blood cell images are loaded. Typically, blood sample is first taken from the patient then the lab operator prepare the sample by adding dye solution, with regard to the percentage of the added dye solution to the sample and the concentration of the dye compound. Then the lab operator places the prepared sample in a microscope and uses an imaging device like camera mounted of the microscope to save a digital image of the sample. Both blood and bone morrow smears can be used in our segmentation method.

This step is made to make a contrast stretching for the red, green, and blue color channels, such that each color compound is stretched individually. The applied stretching is of linear type which shifts the lowest existing color value toward 0, and the highest found toward 255. In order to avoid the effects of impulsive noise existence in the image, the lowest and highest color component value are assessed according to some statistical bases, i.e.:

$$V_{lowest} = m - \alpha\sigma \quad (1)$$

$$V_{highest} = m + \alpha\sigma \quad (2)$$

Where Vlowest and Vhighest are the assessed values for the lowest and highest values may exist in the image, m is the mean value for the color components (R, G or B), σ is the corresponding standard deviation value. α is a predefined parameter used to control the strength of the stretching parameter.

Contrast stretching is very important because some of the blood cell images may have faint colors such that the colors of different regions in the image are very close to each other. This is due to the staining process of the blood film, and to the bad optical setting of microscope-camera devices. The poor color contrast have negative effect on the performance of the proposed system (i.e., localization of nucleus and cytoplasm regions) so, color stretching is important to make the image color data more usable in the next stages of the developed system and its shown in figure (2).

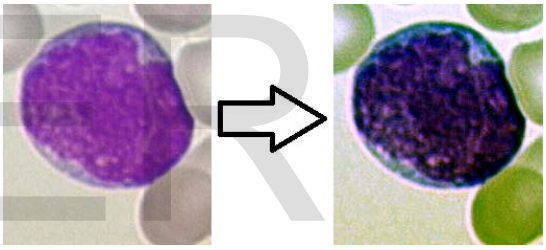

Figure 2: Color enhancement

3.3 Nucleus Segmentation

Segmentation is probably the hardest and the most time consuming step in the proposed system because the algorithms used to collect the whole parts of the desired region of interest are iterative and usually should perform many different tasks at a time. It was noticed that most probably a merge occurred between two or more different regions (e.g., merge white blood cell area with red blood cell area and treat them as one distinct part). The developed nucleus segmentation method, in our proposed system, consists of the following tasks:

A- RGB segmentation: we decided to use RGB color model to do the required segmentation; this is due to its simplicity besides its low computational requirements to accomplish the segmentation. The relevant parameters value (i.e., the color bounding thresholds for each color component) used in the developed algorithm, have been carefully selected after many trials to gain the best segmentation results for all tested image samples. The algorithm implies a binarization process whose result will be stored as a bi-

nary mask refers to the region of interest for the next processing stages. When the color of the tested pixel satisfy the color bounding conditions then the pixel is decided to be a point belongins to ROI (i.e., assigned a value 1) otherwise it is considered as part of the background region (i.e., assigned a value 0).

B- Hole Detection: although many segments shapes may appear in the binary mask image, all resulted segment(s) may have hole region(s) that needs to be filled and assembled with the nearst segment bulk. An 8-directional pixel checker is invoked, which takes a pixel and ensure that this pixel and its neighboring 8 pixels are of the same color and that color is not equal to segment color (i.e., to ensure it is a hole); and this pixel is inside the boundary region of the segmented bulk part (i.e., to ensure it is mostly a part of the segmented region). The radius of search is set to be the longest distance for the checker to search in one direction before it switches to another direction. After the checker verifies that all directions are of the same color within the specified radius, a color filler is applied when a hole is found inside the segmented region. This algorithm is used for filling the small holes (i.e., of radius R or less) which don't need the use of seed filling algorithm, because this later algorithm may consume more time in unnecessary places (i.e., very small holes).

C- Noise Cleaning: After the holes being identified and replaced to be ROI pixels; and additional algorithm is invoked to remove the produced noise pixels and some small regions, which are basically produced as wrong outcomes of the binarization task. This process is very similar to that accomplished by the above mentioned algorithm, but it checks for ROI existence and then checks its neighbors' pixels in neighborhood up to certain radius and if it is found that the neighbors satisfy the detection condition then the tested pixel will be counted as noise (gap) pixel or very small (gap) region.

D- Nucleus Island Removal: after the application of noise removal task, the resulting image will be a binary image which still has little amount of small island regions beside some gaps still exist in the nucleus segment body; these small region(s) are scattered throughout the image. Gaps inside ROI region should be filled by the ROI color to form a uniform segment, while the small island regions (i.e., very small isolated areas are considered as ROI) in the image should be removed because they can significantly degrade the whole diagnosis task accomplished by the system.

The adopted gap filling method is based on flood fill algorithm. The basic principle of the applied method is to select a black pixel (gap) and starts the flood fill process; if the flooded area is 30% or less of the entire image area (as the medical samples acquired for the research are one WBC in each sample so 30% is an acceptable area size to be selected as parameter) then it will be counted as a part of ROI area whose points should be flagged as ROI points, otherwise this area is treated as background and kept as it is.

E- Removal of Small Island Regions: this step is very similar to the previous one; it differs only in the utilization of gap filling results. Instead of collecting the gap points, here the adjacent ROI pixels are collected, using seed filling algorithm. If its size is less than 10% of the image size then the region will be treated as background, {such that all the counted pixels are changed to be background instead of ROI pixels}; otherwise, the collected ROI pixels are left as they are. So this step will remove the small island regions from the medical image.

F- Dilation: This stage is important for smoothing the boundary of the segmented nucleus which may appear ragged. The appearance of ragged edges is not acceptable, because they will give false signatures about the nucleus shape. Dilation can efficiently solve the boundary irregularity. We have applied dilation two times just to be sure that the boundary irregularity is properly handled. The continual application of dilation process, for more than two times, may produce problems, that lead to disfigure the accurate shape of the nucleus which in turn leads to wrong impressions about the nucleus shape attributes.

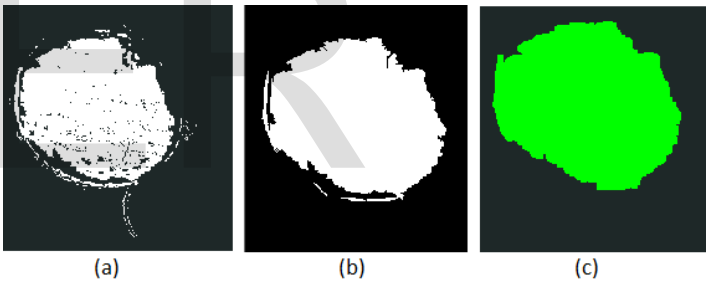

Figure 3: Nucleus segmentation process uses (a) RGB segmentation, (b) noise cleaning and gaps filling, (c) dialation and final nucleus segment

3.4 Cytoplasm Segmentation

All the previous applied stages will lead only to the segmentation of the nucleus of the white blood cell. The cytoplasm of the cell is required to make a proper diagnosis decision for leukemia presence. The segmentation process is nearly similar to nucleus segmentation with some differences. The steps taken are:

A- Boundary Extraction: The boundary edge of extracted nucleus area is required for the next coming steps. The detection of nucleus boundary edge using the produced binary mask image is a simple task. A scan for allocating the color change is applied. The color change occurs at the boundaries separating the nucleus region from background region.

B- Nucleus Neighbor Area Check: A window is created for checking task and it is moved over all flagged edge boundary points of the nucleus area. All points that lay within the window are checked individually so that the window points that are not part of the nucleus region and boundary edge points are collected and are considered as part of the cytoplasm region. The sliding window size used in our established method is of size 5x5 pixels. All collected pixels are considered as parts of the cytoplasm area, and they will be used in the next step of the cytoplasm segmentation process. From the collected cytoplasm points, the mean values of their color components are determined.

C- RGB Color Segmentation: similar to color segmentation scheme used for allocating nucleus, a color based segmentation criteria is applied to segment the cytoplasm region(s) image. First, the standard deviation of the three color channels is calculated using the mean color values resulted from the previous stage, then standard deviation is calculated for each color channel and then lower and higher bound values are calculated using equations (1) and (2). The factor is different from nucleus segmentation process to have a better segmentation result.

D- Dilation of Cytoplasm Area: After cytoplasm segmentation, dilation is invoked to smooth the ragged edges and fill some of the gaps found inside the cytoplasm area.

E- Cytoplasm Location Checking: A search for nodes is conducted to allocate the new boundary points of the cytoplasm area which are adjacent to the already registered cytoplasm area and the collected to be a new overlay of the area of cytoplasm. Taking into consideration, that stage (A) may produce small regions scattered through the image and only the area surrounding the nucleus should be considered as the true cytoplasm.

F- Gaps Filling: as in the case of nucleus, gaps may occur in cytoplasm area; so the gaps filling algorithm is invoked here to construct a well-shaped cytoplasm area that surrounds the nucleus. Small change on the steps of the process in nucleus segmentation is made to work properly on filling the gaps of the cytoplasm region, such as the percentage of gap size to image size is changed from 30% (used in nucleus segmentation) to 5%; where it is better to lower this percentage to get better filling result.

With the end of these steps, we now have a fully segmented white blood cell with two distict parts: nucleus and cytoplasm, ready for operation using diagnostic systems. Figure (4) display (a) the original sample image and (b) final fully segmented white blood cell.

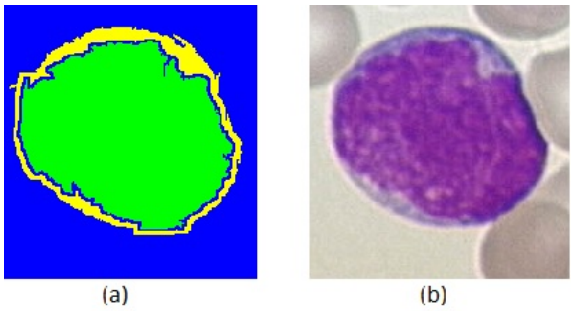

Figure 4: White blood result with (a) final segmentation result (b) original sample image

4 Result and Conclusion

Using different stained medical image from different organs (blood tissue and bone morrow tissue), the job of cognition between the white blood cell and the background and then segmenting the ROI area into nucleus from cytoplasm regions is found more challenging and complex task. Many complicated operations have been utilized for better performance with less execution time.

Color segmentation based on RGB color model is proven to be the most suitable method, in contrary to the case of using the Hue-Saturation-Intensity model, color histogram and gradient color transform which all failed. The use of RGB color space led to developing a simple and fast method for WBC segmentation from the color image.

References

[1] S. Theodoridis and K. Koutrombas, "Pattern Recognition", 4th Edition, Elsevier Inc, 2009.
[2] R. Milewski and V. Govindaraju, "Binarization and cleanup of handwritten text from carbon copy medical form images". Pattern Recognition (31 March 2008)
[3] L. Laura and G. Armando, "Pattern Recognition in Medical Images Using Neural Networks", Universidad Nacional dela Plata, 2001.
[4] R.A. Castellino, "Computer Aided Detection (CAD): An Overview", article published by International Cancer Imaging Society, vol. 5, pp. 17-19, 2005.
[5] L. Laura & G. Armando, "Pattern Recognition in Medical Images Using Neural Networks", Universided Nacional de La Plata, 1999.
[6] I. Cseke, "A fast segmentation scheme for white blood cell images" , Conference C: Image, Speech and Signal Analysis, Proceedings., International Conference on Pattern Recognition ,Vol.III., 11th IAPR 1992.
[7] J. Theerapattanakul, " An efficient method for segmentation step of automated white blood cell classifications" , TENCON IEEE Region 10 Conference, Nov. 2004.
[8] Q. Wu, "Precise Segmentation of White Blood Cells by Using Multispectral Imaging Analysis Techniques", Conference on Intelligent Networks and Intelligent Systems,

2008. ICINIS '08. First International, Nov 2008.

[9] J. Cheewatanon, T. Leauhatong, S. Airpaiboon, and M. Sangwarasilp, "A New White Blood Cell Segmentation Using Mean Shift Filter And Region Growing Algorithm", International Journal of Applied Biomedical Engineering vol.4, No.1 2011.