ISSN (Oline): 2164-1374, ISSN(Print): 2164-1366, Published online February, 2012, (http://www.researchpub.org/journal/ijcs/ijcs.html)

Automated Red Blood Cell Counting

Alaa Hamouda, Ahmed Y. Khedr, and Rabie A. Ramadan

***Abstract*— One of the vital information that help diagnosis many of the patients' sicknesses is the Red Blood Cell (RBC) count. In fact a normal red blood cell count helps the body to perform nearly every function involved with surviving. At the same time counting RBC using traditional methods is very costly and time consuming. In addition, it is often an arduous task to get a precise and accurate result for red blood cells counting using manual cell counting. This paper introduces an efficient method for RBC and the concept of RBC.**

***Index Terms*— health care applications, red blood cell count.**

I. Introduction

One of the important information that doctors usually use to diagnose different diseases is the RBC count. Blood is a connective tissue with two main components—the plasma, a clear extracellular fluid, and the formed elements, which consist of the blood cells and platelets. Blood cells are composed of red blood cells, white blood cells, and platelets. A major function of the red blood cells, erythrocyte, is to carry oxygen to all the tissues from the lungs. The red blood cells than transport the carbon dioxide from the cells because of breaking down the nutrients. An erythrocyte is a disc-shaped cell with a thick rim and a thin sunken center where the nucleus used to be [5].

The RBC is a count of the number of red blood cells per cubic millimeter of blood. This measurement is made with a microscope and a specially ruled chamber (hemacytometer). Counting RBC in this way is costly and time consuming. The RBC is recorded as millions of cells per cubic millimeter (Normal Ranges: males, 5.4 ± 0.8; females, 4.8 ± 0.6). In almost all of the medical laboratories, blood cell count reports are taken based on physicians' recommendation in order to assist the diagnosis of the particular ailments of the patients. In a manual RBC count, 10µl of blood is diluted in 1990µl of dilution solution. This results in a dilution of 1:200. This suspension is usually well-mixed and be immediately placed into the counting chamber. After approximately 3 minutes, the RBCs will have settled, and the MLT begins counting the RBCs in 80 small squares. The calculation is achieved by following the formula.

$$\frac{RBCs}{\mu l} = \frac{\text{number of counted RBCs} \times \text{dilution}}{\text{number of counted squares} \times \text{volume above one small square}} \quad (1)$$

Several attempts are conducted to automate the process of red blood cell from an image. In [6], the authors adapted neural network along with Genetic Algorithms (GA) for red blood cell classification for Thalassemia diagnostic tool. Similarly, the authors of [3] studied the usage of different neural network technique for extracting the red blood cell components from microscopic images and classify them using neural networks. Also, there are many commercially products have been developed to automatically count RBCs or WBCs such as CellaVisionAB [1]. Although, these products are widely available, they are costly and unbearable for medical laboratories; especially in developing countries.

In this paper, we present our first step towards developing automatic process of RBC count from an image. The approach used in this paper is based on some image preprocessing such as equalization and segmentation as well as on clustering using k-mean clustering algorithm. Our results showed similar results to the neural network counting presented in [6][3].

The paper is organized as follows; an overview on classification and clustering algorithms including k-mean and ID3 algorithms are presented in section 2; section 3 elaborate on our proposed RBC count approach; in section 4 , the experimental results are depicted; finally the paper is concluded in section 5.

II. Overview

Clustering and classification are widely used in many applications. Some of these applications are real-time application. Clustering in most cases is defined as the process of grouping objects into a set of similar clusters; dissimilar objects are classified into different clusters. On the other hand, classification analyzes and categorizes the data into known classes. In most cases, unsupervised classification, called clustering or exploratory data analysis, no labeled data are available. There are two types of clustering which are hard and hierarchal clustering. For hard clustering, each pattern only belongs to one cluster. However, a pattern may also be allowed to belong to all clusters with a degree of membership as in hieratical clustering. In this paper, we utilize the hard clustering for extracting the RBC from an image. Figure 1 shows the steps for a clustering procedure. As can be seen in the figure, given data to be classified, distinguishing features from a set of candidates, while feature extraction utilizes some transformations to generate useful and novel features from the

Manuscript received October 9, 2001.

Alaa Hamouda is with the Systems and Computers Department , faculty of Engineering , Al-Azhar University, Cairo Egypt. Phone: +(202) 011-229-7719; (e-mail: dr.alaa.hamouda@gmail.com).

Ahmed Y. Khedr is with Systems and Computers Department , faculty of Engineering , Al-Azhar University, Cairo Egypt. Phone: +(202) 010-124-1440; (e-mail: akhedr@azhar.edu.eg).

Rabie A. Ramadan is the Computer Engineering Department , Faculty of Engineering, Cairo University, Cairo , Egypt. Phone: +(202) 016-764-3227; (e-mail: rabie@rabieramadan.org).

ISSN (Oline): 2164-1374, ISSN(Print): 2164-1366, Published online February, 2012, (http://www.researchpub.org/journal/ijcs/ijcs.html)

original ones. Efficient selection to the features greatly simplifies the subsequent design process. Feature extraction and selection is usually followed by clustering design or selection of suitable algorithm. In other words, in this step, we select a proximity measure and criterion function. Patterns are classified based on their similarity to each other according to the selected proximity. Sometimes, cluster validation and interpretation processes follow the clustering design if there is some previous knowledge about the input data.

One of the clustering algorithms is K-mean clustering; K-means clustering algorithm is used to cluster observations into groups of related observations without any prior knowledge of those relationships. The k-means algorithm is one of the simplest clustering techniques and it is commonly used in medical imaging, biometrics and related fields. In the first step, a number of desired clusters, k, is chosen. In the second step, k starting points to be used as initial estimates of the cluster centroids are selected. After all, each point (i.e., pixel) in the workload data set (i.e., image) is examined and assigned a cluster with the nearest centroid to it. When each point is assigned to a cluster, new k centroids are recalculated. These steps are repeated until no point changes its cluster assignment, or until a maximum number of passes through the data set is performed. As can be seen, it is a very simple clustering algorithm that can does not take that much time or large memory to run.

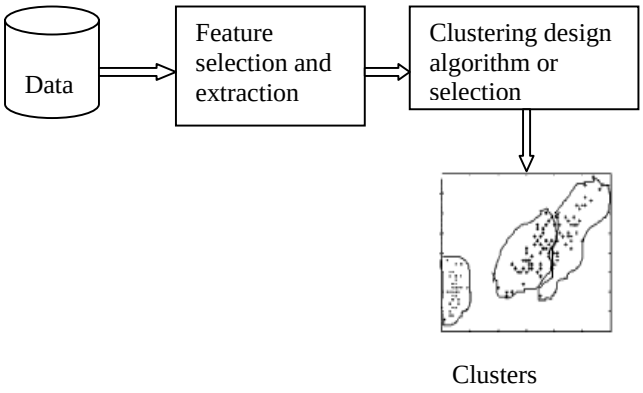

Figure 1: General clustering steps

Classification algorithms are based on the assumption that the image depicts one or more features and that each of these features belong to one of several distinct and exclusive classes. The classes may be specified a priori by an analyst into sets of prototype classes, where the analyst merely specifies the number of desired categories. One of the most famous classification algorithms is the decision tree. It is a supervised classification model used for learning discrete-valued target functions. The tree consists of decision nodes and leaf nodes. The decision trees algorithm creates hierarchical structure of classification rules "If ... Then ..." looking like a tree. One of the decision trees algorithms is ID3. The main idea behind ID3 is:

Each non-leaf node in the tree represents an input attribute and each arc constitutes a possible value of that attribute. While, the leaf-node symbolizes the expected value of the output attribute when the input attributes are described by the path from the root node to that leaf node. For efficient decision trees, each non-leaf node corresponds to the input attribute which is the most informative about the output attribute amongst all the input attributes not yet considered in the path from the root node to that node. In such case, we will be able to predict the output attribute using the smallest possible number of questions, on average. The entropy is used to decide on how informative certain input attribute is related to an output attribute for a subset of the training data.

III. Our Approach

In this section, we show our approach for automatic RBC count. Figure 2 shows an image that might involve some unwanted particles (noise). Therefore, some preprocessing is needed which is called image preparation phase as shown in Figure 3. As can be seen, this phase consists of two steps which are histogram equalization and segmentation. In the first step, histogram equalization, the intensity value of the given image is adjusted using image intensity transformation. In the second step, segmentation, blood cell are detected by differentiate them from the background in terms of contrast. Changes in contrast can be detected by image processing operators that calculate the gradient of an image. Then a threshold can be applied to create a binary mask containing the segmented cell. The edge detection is done by using the Sobel operator.

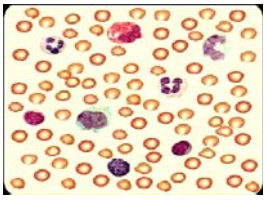

Figure 2: Blood Sample

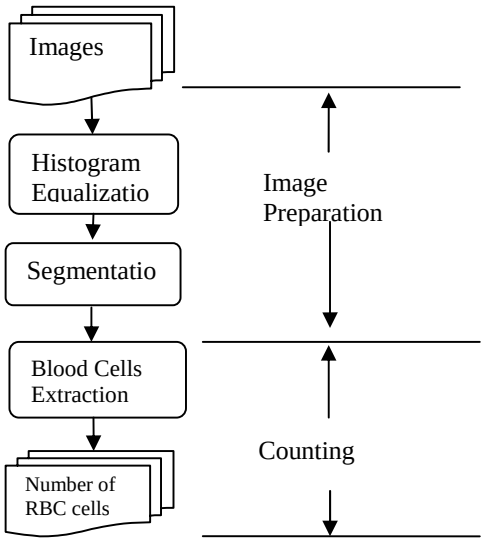

Figure 3: RBC Counting Phases

In the second phase, blood cells extraction, we applied different methods looking after the highest accuracy as well as less complexity. We experimented with four different methods which are K-means clustering, classification, hybrid between clustering and classification. In addition, we investigated the

ISSN (Oline): 2164-1374, ISSN(Print): 2164-1366, Published online February, 2012, (http://www.researchpub.org/journal/ijcs/ijcs.html)

suitability of decision trees on the problem on hand. The following subsections elaborate on the utilized methods in this phase.

A. Clustering using K-means

After applying image enhancement of phase one, we segment colors in an automated fashion using the L*a*b* color space and K-means clustering to differentiate between red and white cells, figure 4. The L*a*b* color space is derived from the CIE XYZ tristimulus values. The L*a*b* space consists of a luminosity layer 'L*', chromaticity-layer 'a*' indicating where color falls along the red-green axis, and chromaticity-layer 'b*' indicating where the color falls along the blue-yellow axis. All of the color information is in the 'a*' and 'b*' layers. The difference between two colors can be measured using the Euclidean distance metric. Finally, simple RBC counting is done.

B. Classification

Here, we apply image enhancement in the first phase, as mentioned. Then, we use image segmentation to locate objects and boundaries in input images. All classification algorithms are based on the assumption that the image in question depicts one or more features and that each of these features belongs to one of several distinct and exclusive classes. The classes may be specified a priori by an analyst (as in supervised classification) into sets of prototype classes, where the analyst merely specifies the number of desired categories. We use size feature to separate between RBCs and WBCs. Considering Gaussian distribution, we use the threshold of *(average size + 3* standard deviation)*. After the classification, we could easily count the RBCs, figure 4.

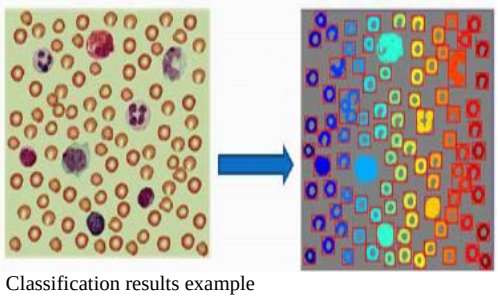

Figure 4: Classification results example

C. Hybrid of Clustering and Classification

In this method, we combined the previous two methods. First, we use K-means clustering algorithm to find the cluster of the red cells as shown in figure 5. Then, we apply the classification by the same threshold for the red cells clustering. Finally, we count RBCs.

D. Learning Using Decision Tree

For learning using decision trees, we have to have some training data sets. As shown in figure 6, after the usual images enhancement, a learning process is done. We used 15 images to train a decision tree to discover the dominant features. Each cell in the learning database was determined to be RBC, WBC, or overlap RBC cells. Then, the extracted features in the decision tree are used to classify the cells and find the RBCs.

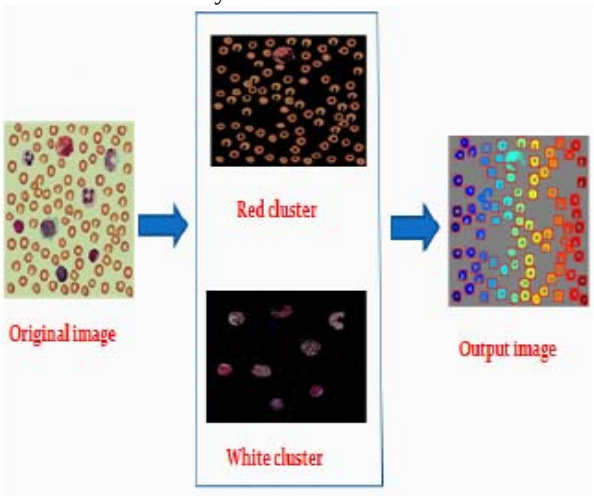

Figure 5: Hybrid of Clustering and Classification

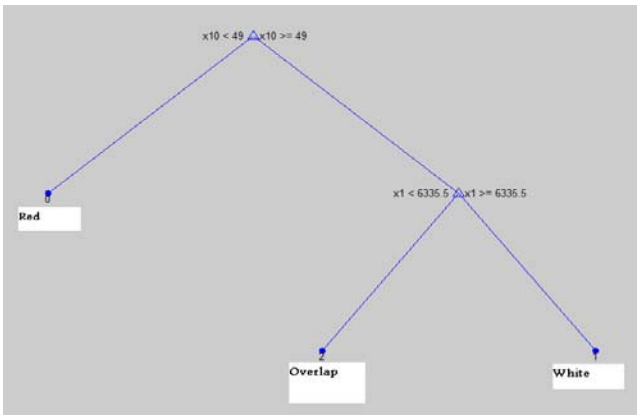

Figure 6: Resultant Decision Tree

IV. Experimental Results

To test our previous approach, we selected 30 sample images for blood testing with different resolutions from different laboratories. Images are manipulated using histogram equalization and segmentation. Then in the second phase, we applied one of the four proposed methods on all given images. Since the original images are bright, the histogram equalization and adjusting of an image affects the contrast of blood cells from the background. As a consequence, this might affect the edges of blood cells.

To investigate the performance of each technique, we use the sensitivity and accuracy parameters according to the following equations:

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Where TP is the True Positive), TN is the True Negative, FP is False Positive, and FN is the False Negative.

Based on our data set, 30 images, the average sensitivity and accuracy for all techniques are presented in table 1. The results seem promising in terms of accuracy and sensitivity. However,

ISSN (Oline): 2164-1374, ISSN(Print): 2164-1366, Published online February, 2012, (http://www.researchpub.org/journal/ijcs/ijcs.html)

seems that clustering by color has better sensitivity and accuracy than classification by size. Merging between both clustering and classification improves the sensitivity with additional 5% and 1% accuracy. Since the decision tree depends on the size as a dominant feature for classification, it yields the best sensitivity and accuracy with, 97%.

Table 1: The experimental results for the different techniques

Technique	Sensitivity	Accuracy
Clustering using K-means	86%	83%
Classification	83%	82%
Hybrid of clustering and classification	88%	83%
Learning using decision tree	97%	97%

V. Conclusion and Future Work

This work shows that automatic red blood cell recognition and counting using image processing and decision tree to classify RBCs yielded a good classification with high sensitivity and accuracy of 97%. From this result, we could apply with confident on labs using desktop or laptop machines or even cell phone. Another way to utilize the results of this paper is to develop a small FPGA based device as a component of E-health system to count RBCs. Also as future work, more blood diseases (e.g. anemia) that can be detected using image processing on the same device (sensor, mobile or FPGA based device) can be investigated and developed.

VI. References

[1]. A. Kratz, H.Bengtsson, J. E. Casey,J. Keefe, G. Beatrice, D. Grzybek, K.. Lewandrowski, and E. Van, "Performance Evaluation of the CellaVision DM96 System WBC Differentials by Automated Digital Image Analysis Supported by an Artificial Neural Network", American Society for Clinical Pathology, Am J Clin Pathol, 2005.

[2]. Emerging Wireless Technologies for Healthcare Applications," Frost & Sullivan, March 30, 2005.

[3]. J. Poomcokrak1 and C. Neatpisarnvanit, "Red Blood Cells Extraction and Counting", The 3rd International Symposium on Biomedical Engineering (ISBME 2008).

[4]. K. Bourzac, "Tiny $10 Microscope, A high-resolution, lens-free microscope fits on a dime-size chip," Technology Review, MIT , 2008. , http://technologyreview.com/biomedicine/21147/

[5]. K. Saladin, Anatomy and Physiology: The unity of form and function, McGraw-Hill, NY, 4th, chap. 18, pp. 680-696, 2007.

[6]. W. Wongseree and N. Chaiyaratana, "Thalassemic Patient Classification Using a Neural Network and Genetic Programming," IEEE, pp. 2926-2931, 2003.

[7]. "Wireless Networks in Industrial Automation," Frost & Sullivan, June 30, 2006.

Alaa Hamouda is an Assistant Professor at the Engineering faculty of Al-Azhar University, Cairo, Egypt. His research interests have focused on Data Mining, Arabic Opinion Mining, Arabic Text Summarization, Multi-agent systems, CMMI, Swarm Intelligence and Sensory Network. As a consultant, he participated in developing software quality models for Capability Maturity Models Integration (CMMI). He also participated in establishing reliable standard work procedures in software houses to apply standard project management processes (PMI).

Ahmed Y. Khedr is associated with Al-Azhar University, Cairo, Egypt. He is currently an assistant predecessor at System and Computer Engineering Department at the faculty of Engineering. His research involves WSNs, Mobile Computing, and Data Mining. More information about his personal and research work could be found at faculty.uoh.edu.sa/a.khedr.

Rabie A. Ramadan is associated with Cairo University, Giza, Egypt. He is currently an assistant predecessor at Computer Engineering Department at the faculty of Engineering. He is a member in IEEE and is currently serving as the chair of IEEE Egypt Computational Intelligent Chapter. He is also got his PhD from Southern Methodist University, Dallas, TX, USA. His research involves WSNs, Mobile Computing, and Computational Intelligence. He is the founder of the Applied Sensing Technology Laboratory (ApplS-LAB). More information about his personal and research work could be found at www.rabieramadan.org.