Volume : 2 | Issue : 1 | January 2013 ISSN - 2250-1991

Research Paper Engineering

Classification of RBC And WBC in Peripheral Blood Smear Using KNN

* Chhaya S. Hinge ** Aarti G. Ambekar
*** Samidha S. Kulkarni

* ME Student, Electronics Engineering Department, K.J. Somaiya College of Engg., Vidyavihar, Mumbai

** ME Student, Electronics Engineering Department, K.J. Somaiya College of Engg., Vidyavihar, Mumbai

*** Associate Professor, Electronics Engineering Department, K J Somaiya College of Engineering, Mumbai

ABSTRACT

The objective of the paper is to present the blood cells classification method to develop an automatic differential blood count system. The counting and analysis of blood cells provides a huge amount of information to the pathologist , for finding the disease and its treatment. The proposed system consists of two automatic steps : segmentation of blood cells and classification. For classification, feature vector is created by extraction of various features from the segmented blood cells. Experimental results have shown the effectiveness of system.

Keywords: Blood cells ,feature extraction, kNN classifier

I. Introduction:

The process of life is maintained by blood, which is a specialized body fluid consisting of plasma and blood cells .The blood cells can be classified as erythrocytes, leukocytes and thrombocytes. To perform proper diagnosis of the disease, identification of the blood cells and their relative quantity in the blood samples must be known. For the image acquisition, the blood smear slides are taken and by connecting a high resolution digital camera to microscope, images are captured by adjusting microscope magnification to get good resolution. For identifying different types of blood cells and for counting their quantity in blood smear, image processing is used on various blood smear images. The image is preprocessed, so as to remove noise, and then conversion to gray level is done. For separating out various cells, binarization, morphological operations followed by watershed segmentation and labeling of segmented cells is performed. Different features of cells are then extracted from the labeled image. From this features, database is created and with the help of this database , set of test images and kNN classifier, classification of various cells in an image is done.

II Blood cell types :

In the peripheral blood, blood cells can be classified as erythrocytes - Red Blood Cells (RBC's), leukocytes - White Blood Cells (WBC's), and thrombocytes - platelets. Different categories of blood cell exhibit different morphologic characteristics that can be used, to identify and classify different cells, [3].

Erythrocytes - (RBC's) :

The erythrocytes are the cells which are highest in numbers. It does not have nucleus. Their cytoplasm consists of hemoglobin that gives a typical red color of the cells. It's primary Function is to transport oxygen from the lungs to the cells of the body & assist with CO2 removal .

Leukocytes - (WBC's):

In terms of the size and shape of the nucleus, the colour of the cytoplasmic staining, and percentage ratio of nucleus to cytoplasm, leukocytes can be classified into5 major types as:

Neutrophils: The nucleus of neutrophil is frequently multilobed and its cytoplasm has very tiny faintly pink stained granules with low visibility.

Eosinophils: The nucleus of eosinophil is frequently bi-lobed and its cytoplasm consists of orange to red stained granules.

Basophils: They are very less in blood. The nucleus of basophil is frequently bi-lobed to multi-lobed. After staining, the cytoplasm is full of large, deep-bluish to purple granules.

Monocytes: Monocytes are large, circulating, white blood cells having the kidney shaped nucleus. The cytoplasm is of light blue colour.

Lymphocytes: Lymphocytes are of 2 main types , B cells and T cells. It has a large, dark-staining nucleus with a small cytoplasm to nucleus ratio.

Thus one can recognize the seven different blood cell types via their cytoplasm, granules, staining properties of the granules, shapes of nucleus, the ratio of nucleus to cytoplasm, and the type of nucleus lobes. Therefore, developing an automatic blood cells recognition system is feasible via image processing and pattern recognition techniques.

III Proposed method architecture :

Chart 1 shows the stages of classification.
"Chart 1 about here"

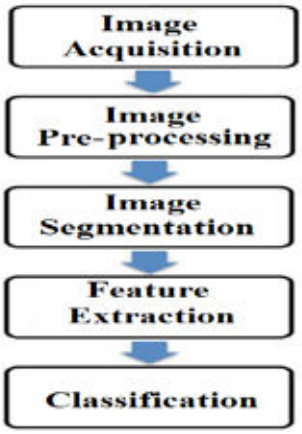


Chart 1: Flowchart of proposed method

1. Image acquisition: The images from blood smear slides are captured by connecting high resolution digital camera to a microscope by adjusting microscope magnification to get good resolution.

2. Image preprocessing: The pre-processing stage includes noise reduction and contrast enhancement of acquired image and is performed to prepare the image for the further stages.

For processing the image it is converted into gray scale image, to avoid being influenced by dye colour. A typical peripheral blood smear image consists of four components, namely background, erythrocytes, leukocytes, and thrombocytes. To segment the desired object from the background, the green channel is used, as it is found that the green component of the RGB input image gives the best contrast between the background and the foreground which consists of blood cells.

3. Image segmentation: In image processing, segmentation is the process of partitioning a digital image into multiple segments. For blood cells classification, segmentation of cells is required. So, for segmentation of blood cells , the segmentation process based on thresholding, morphology, and watershed segmentation is used. It is performed to cover every element in the blood slide in a distinct area.

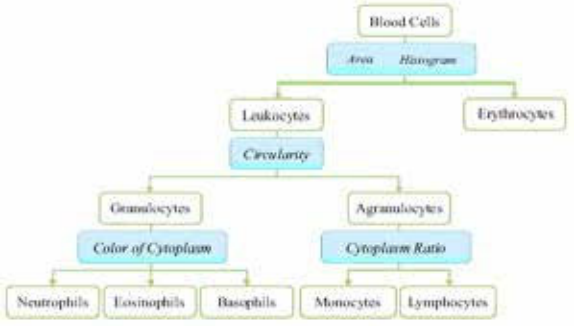
i. Binarization: In order to segment the desired object from the background, i.e. to produce a binary image, that separates foreground and background image pixels, Otsu's adaptive threshold algorithm is applied on the green channel [4].

ii. Mathematical Morphology: Mathematical morphology operations are used to fill the holes in the blood cells and to remove the unwanted points in the red blood cells and background, [4].

iii. Watershed Segmentation: For separation of overlapping cells, watershed transform is used, [2].

4. Feature Extraction: In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. To analyze the individual objects of interest from the background in the segmentation process, border touching cells obtained in binary images are removed and then labelling of the objects within the segmented image is performed .The features of the blood cells such as area, histogram, circularity, cytoplasm ratio , colour of cytoplasm are extracted from a set of images to prepare a database, required for classification of cells. A hierarchical strategy is used here so as to get fast and more accurate classification of cells. Chart 2 shows the hierarchical strategy of classification.

"Chart 2 about here"



proper classification of various blood cells , the important features extracted [1] are:

Area & Histogram: Out of various blood cells WBC's are the largest whereas platelets are the smallest blood cells. So, area can be used as a feature to separate WBC's from RBC's. Also based on histogram, RBC's and WBC's can be separated out, as RBC's are without nucleus and WBC's have nucleus. After the labelling of the segmented cells in an binary image ,WBC's and RBC's can be separated out based on these features.

For further features the segmentation of nucleus of WBC from its cytoplasm is required.

Circularity: The circularity of the nucleus of WBC is calculated using below listed formula, and if the circularity is one, it indicates a perfect circle and if it is 0, it indicates increasingly elongated polygon. So, based on circularity, WBC's can be divided into agranulocytes and granulocytes, as agranulocytes have almost circular nucleus, whereas granulocytes has irregular shape nucleus.

$$Circularity = 4 * \pi * (\mathrm{A}i/P^{2})$$

Where, $\mathrm{A}i$ = Size of the nucleus in the labelled blood cell i, P = Perimeter of the nucleus in the labeled blood cell i.

Color of cytoplasm: The granulocytes are named according to their staining properties. So for each granulocyte, hue component is found and then average and standard deviation is calculated.

Cytoplasm ratio: Lymphocytes have larger nucleus then monocytes, so the ratio of the cytoplasm to the cell can be used to distinguish between monocytes and lymphocytes which are types of agranulocytes.

$$Circularity = 4 * \pi * (\mathrm{A}i/P^{2})$$

Where, $\mathrm{A}i$ = Cytoplasm area of the labeled cell$\mathrm{A}i$ = Cytoplasm area of the labeled cell i.

4.Image classification: There are various techniques to classify images. The algorithm used here is kNN. In pattern recognition, the k-nearest neighbor algorithm is a method for classifying objects based on closest training examples in the feature space. kNN is a type of instance-based learning, where the function is only approximated locally and all computation is deferred until classification. Here an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor. The drawback of this simple approach is the lack of robustness that characterizes the resulting classifiers. The high degree of local sensitivity makes it highly susceptible to noise in the training data.

IV. Experimental Results and Conclusions:
This section explains the experimental results and their analysis.

A. Image Acquisition :
Ten human peripheral blood smear slides were observed under a camera attached microscope and 36 images were taken for testing. There are 659, 15, 9, 7, 7 and 11 objects for all cell classes, erythrocytes, neutrophils, eosinophils, basophils, monocytes, and lymphocytes respectively.

B. Results and conclusions:
Table 1 presents the values corresponding to features computed for different types of segmented cells. From experimental results it is observed that neutrophils, a type of WBC's, are having highest area ,whereas RBC's are of smallest area, if compared between RBC's and WBC's size. It is also observed that , among various types of WBC's, monocytes and lymphocytes are the one which are having nearly circular nucleus. In addition , it is seen that monocytes has larger value of cytoplasm ratio compared to lymphocytes. From results obtained it can also be stated that, mean and standard deviation of hue component of various granulocytes are different, indicating the different colour of cytoplasm of each granulocytes.

"Table 1 about here".

Cell Type	Feature				
	Area	Circularity	Cytoplasm ratio	Cytoplasm colour	
				Mean	Std. dev
1	1.989	NA	NA	NA	NA
2	5.717	0.317	NA	3.366	0.95
3	5.190	0.386	NA	3.930	0.77
4	4.975	0.271	NA	4.619	0.50
5	3.452	0.704	0.3722	NA	NA
6	2.981	0.833	0.2439	NA	NA

Table 1 : Experimental results *1:Erythrocytes, 2:Neutrophils, 3:Eosinophils, 4-Basophils, 5-Monocytes, 6-Lymphocytes, NA – Not applicable

REFERENCES

[1] Wei-Liang Tai, Rouh-Mei Hu, Han C.W. Hsiao & Rong-Ming Chen, and Jeffrey J. P. Tsai1, "Blood cell image classification based on hierarchical SVM ", 2011 IEEE International Symposium on Multimedia,pp.129-136,2011. | [2] S. Beucher, F. Meyer, "The Morphological approach to segmentation: the watershed transform, E.R. Dougherty,Marcel Dekker, 12, pp. 433–481. | [3] L. Sherwood , "Human Physiology: From Cells to Systems", Brooks Cole,2008. | [4] R.Gonzalez , R. Woods & Steven L, 'Digital image processing using MATLAB. |