Indian Journal of Science and Technology, *Vol 10(6), DOI: 10.17485/ijst/2017/v10i6/111137, February 2017*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Segmentation of White Blood Cell using K-Means and Gram-Schmidt Orthogonalization

J. Puttmade Gowda[1] and S. C. Prasanna Kumar[2]

[1]Department of Electronics Engineering, Jain University, #44/4, District Fund Road, Jayanagar, 9th Block, Behind Big Bazaar, Bengaluru - 560069, Karnataka, India; puttamadegowda@rocketmail.com
[2]Department of Electronics and Instrumentation, R V College of Engineering, Mysuru Road, R. V. Vidyanikethan Post, Bengaluru - 560059, Karnataka, India; prasannakumar@rvce.edu.in

Abstract

Testing of blood is a very important examination process, counting of cells is an important laboratory process for identifying blood related diseases. Microscopic evaluation by experts is a slow process and result is depends on skill and experience of technician, also the process is tedious and time consuming. Therefore automatic medical diagnosis system is necessary way to identifying the diseases in short time. For providing information about blood related diseases like leukemia it is necessary to identify and inspect white blood cell in a peripheral blood smear. So Segmentation is an important step in classifying the constituents of blood. This paper represents efficient segmentation of blood image by using K-means clustering method followed by Gram-Schmidt Orthogonal process to detect automatic blood cell nuclei.

Keywords: k-Means Clustering, Orthogonal Set, Orthonormal Set, Segmentation, White Blood Cell

1. Introduction

To check functions of many body organs and for identifying blood related diseases it is necessary to examine the blood. Manual microscopic examination is time consuming and result is depends on experience of the hematologist. The advantage of automation of visual samples made easy for hematologist to detect and identify any abnormality in blood samples.

The white blood cells plays main roll in diagnosing the diseases. So immune system defends against the diseases is totally depends on WBC.

Following are the main process of automation of blood cells. They re image acquisition, image preprocessing, image segmentation using k-means clustering for separate different clusters and analyze the image. In this work we are targeting automatic segmentation of leukocytes nucleus and cytoplasm. For enhance the nucleus and cytoplasm separately we use Gram-Schmidt orthogonalization process. In K-means clustering scheme they use sum of width in variations of cluster with in clustering. It is efficient in computation if clusters are well separated and compact, this method gives satisfactory results. By using Gram-Schmidt orthogonalization process we segment the nuclei of WBC.

2. Related Work

Dorini, Rodrigo Minetto, et. Al.[1], uses two method of segmentation Level set and watershed transform. To segment nuclei they use SMMT operator and identify cytoplasm by granulometric analysis and morphological transformation.

Baidya Nath shah et. al.[2], uses outlier trace by analysis of principal component (PCA), automate active contour (snakes) for object detection. They trace snakes run on an image for detecting white blood cell nuclei.

S. Mohapatra et.al.[3], uses a color segmentation process in two stage based on fuzzy to obtaining WBC from the other blood cells. For classifying nucleus they use Hausddorff dimension and contour signature.

*Author for correspondence

S. H. Rezafotighi et al[4], and M. Mohamed et. al.[9] introduced orthogonalization process of Gram-Schmidt to segmenting nuclei of WBC.

M. Y. Mashor et. al.[5], adopted global contrast strechning process and for improve the quality of image HSI color segmentation is used.

E.A. Mohamed et. al.[6] used watershed algorithm for segmenting the nuclei of lymphocyte. To do this they use optimal threshold system.

Jan duan et. al.[7] select HIS color space followed by threshold segmentation to extracting the nucleus of WBC and region growing for color information for extract the cytoplasm.

E. U. Francis et. al.[8] use partial contrast stretching method for enhancement of image followed by image segmentation. They extract features of WBCs and screening of image by MLP.

M. Y. Mashor et. al.[10] uses two processing methods. In their part one they enhance region of interest by using contrast enhancement technique on leukemia images and in part second they use image segmentation on HIS color space.

Monica Madhukar et. al.[11] presents microscopic images are exposed to correlation of color, contrast enhancement and apply K-means clustering to the resultant images, by using this method they extracting nucleus of the WBC.

Fabio Scotti[12] also uses the same techniques for enhance microscopic image by eliminating background image, then use fuzzy k-means clustering for segmenting the image.

D. C. Haung et. al.[15] works on threshold method of segmentation for extracting WBC nuclei using Otsu's method, co-occurrence matrix was used as a texture measure of segmented images. They clearly distinguish five types of leukocytes by obtaining suitable features using principal component analysis (PCA).

Following are the five types of leukocytes.

Neutrophil: The most populous circulatory leukocyte, short lived, tiny light granules. They have multi lobed nucleus, these lobes are connected each other by a very thin strands. These strands are made by nuclear materials. Accounts for 50-70% of all leukocytes.

Eosinophil: Granulocytes, large granules, red or pink in color. There are two lobes in the nucleus, less than 5% of all leukocytes.

Basophil: There are large number of Basophilic granules appear in this type of cells. When stained looksdeep blue or purple in color, they represents below 1% of total leukocytes.

Lymphocytes: It is an agranular cell and have clear cytoplasm, stained thin blue. The nucleus occupies whole cell surrounded by a lean rim of cytoplasm, accounts for 25-35% of leukocytes.

Monocytes: It has a large nucleus, agranular cell. The shape of the nucleus of monocyteis 'U' or shape of kidney been abundant cytoplasm and cytoplasm looks light blue in color 3-9% of all leukocytes.

3. Methodology

The Architecture of proposed system is shown in Figure 1. In preprocessing step the image is converted in to gray scale followed by filtered and normalization is applied.

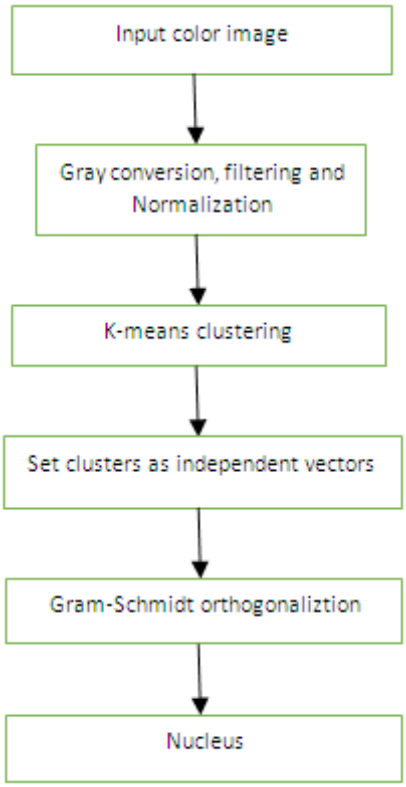


Figure 1. algorithm of segmentation of WBC nucleus.

3.1 Preprocessing

3.1.1 Gray Conversion

The time required by RGB image to process is more than that of gray scale image, so in preprocessing stage, the RGB image is converted to its grayscale image.

3.1.2 Filtering

Median filter replace the value of each and every pixel with median of its neighbor pixels. The process is by selecting values of neighborhood pixels and sort in some numerical analysis and pixel value can replaced by middle most value. The median value is given in equation below.

$$\text{Med (m)} = \begin{cases} X(l+1) = X(m) \ \ for \ N = 2l+1 \\ \dfrac{X(l) + X(l+1)}{2} \ \ for \ N = 2l \end{cases}$$

3.1.3 Normalization

The image function $h_1(x,y)$ transformed to the image function $h_2(x, y)$ and retain information about original image and also satisfies certain conditions. This is related to as normalization. So $h_2(x, y)$ is the normal form of original image $h_1(x,y)$. The normalization process is shown by the relation below.

$$h_2(x_2, y_2) = Hh_1(x_1, y_1) + B$$

$$\begin{pmatrix} x1 \\ y1 \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x2 \\ y2 \end{pmatrix}$$

Here H, B, a, b, c, d. X and Y are real constants. Most of the cases the equation is limited by

$$H > 0, \quad B=0$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \left(\frac{1}{D}\right) \begin{bmatrix} cos\phi & -sin\phi \\ sin\phi & cos\phi \end{bmatrix}$$

$$D > 0$$

H is considered as contrast level and B is bias. These two contrast condition maintain the function of the image is positive.

3.2 K-Means Clustering

Clustering looks the grouping of a multi-dimensional data set by comparing similarity or dissimilarity values. Here we use method of partitional clustering, partition the given samples has K-clusters, the features in a particular sample group are very similar to one another than to those features in different sample groups.

In K-means clustering scheme, sum of with in-cluster variations in a given clustering is used. i.e. within-cluster variation is nothing but the sum of the squared Euclidian distances between the samples and cluster center.

It is computationally very efficient, if the clusters are well separated and compact this method gives satisfactory result.

For partition the intensities observed in the image into similar groups they use image segmentation by using clustering method and segmented image into clusters or regions. This process of image segmentation follows following steps:-

- Initialize number of groups 'm' and centroids C_j.
- Assign every pixel to a particular group of closet centroid.
- Assigning of all pixels are completed, recalculating the centroids.
- Repeat above two steps until there is no change in centroids.

The K-means segmentation method minimizes the sum of the within-cluster variances

$$Q = \sum_{j=1}^{m} \sum_{i=1}^{n} \left\| V_i^j - C_j \right\|_2$$

Where V_i^j indicates i^{th} sample of j^{th} class m_j, C_j the center of the j^{th} region which is mean of $V_{i \text{ belongs to, }} m_j$. For calculating centroids by using the Euclidian distance. The K-means algorithm to segmentation is applied to input blood image and take four clusters.

3.4 Gram-Schmidt Orthogonalization

From linear algebra it is a process for orthogonalizing a set of vectors called Euclidian space (an inner product space) Q. Gram-Schmidt analysis considers linearly independent, finite set Q={m_1, m_2, ----------m_n} and generates an orthonormal set Q'={p_1, p_2, ...p_n} that spans the same sub space as Q.

Projection operator defined as follows

$$\text{Proj}_p \, m = \frac{< p, m >}{< p, p >} \, p = <p, m> \frac{p}{< p, p >}$$

Where <p, m> is an inner product of vectors p and m. By using this operator vector orthogonally projects on to vector m.

This process works as

$P_1 = m_1$ $\qquad e_1 = \dfrac{p1}{abs(p1)}$

$P_2 = {}_m2 - \text{proj}_{p1} \, m_2$ $\qquad e_2 = \dfrac{p2}{abs(p2)}$

$$P_k = m_k - \sum_{j=1}^{k-1} proj_{p_j} \, Mk$$

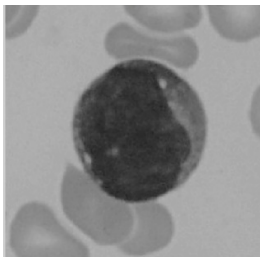
The sequences $p_1, p_2, \ldots\ldots\ldots p_k$ is the set of orthogonal vectors, and normalized vectors $e_1, e_2, \ldots\ldots\ldots\ldots e_k$ we generate a orthonormal set.

Using this method, for linearly independent set $S=\{m_1, m_2, \ldots\ldots\ldots m_n\}$ we choose a particular vector it is maximum orthogonality with one required vector p_k where as minimum with all remaining vectors with in N-dimensional spaces.

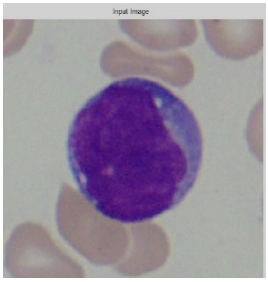
To use this Gram-Schmidt orthogonalization for WBC nucleus segmentation for each pixel, a 3-D feature vector is defined using clusters of image obtained from K-means segmentation method. Then by Gram-Schmidt method, a weighting vector 'e' is obtained by strengthening desired cluster vector, weakening other cluster vectors.

4. Results and Discussion

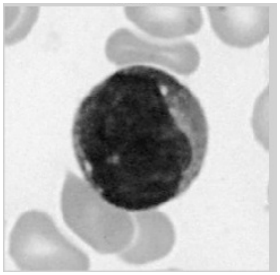
The blood smear images are considered as input images. The WBC segmentation is done by combination of K-means clustering and Gram-Schmidt orthogonalization algorithm. Initially the input image is required filter and normalization.



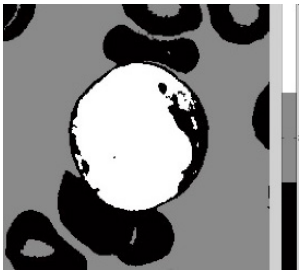
(a)Input image



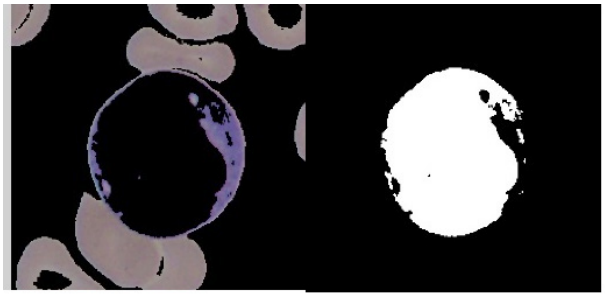
b. Gray scale



c. Image



d. Normalized imaged. K-means segmented Image



e. RBC and cytoplasm f.WBC nucleus Of WBC

Figure 2. Test image with Intermediate Results.

The Figure 2(a) show the original image, its gray scale image is shown in Figure 2(b). the normalized image is shown in Figure 2(c), the resultant image of K-means clustering image is shown in Figure 2(d), the clustered image of RBC and cytoplasm of WBC is shown in Figure 2(e), the nucleus of WBC is shown in Figure 2(f).

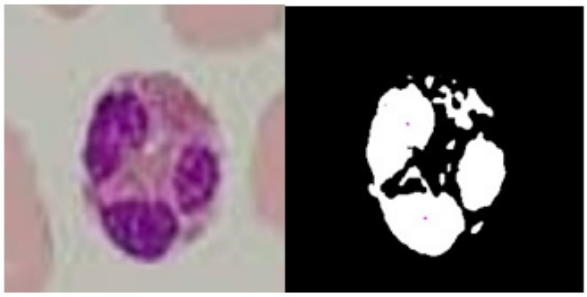


Figure 3. Original image of segmented image Neutrophil.

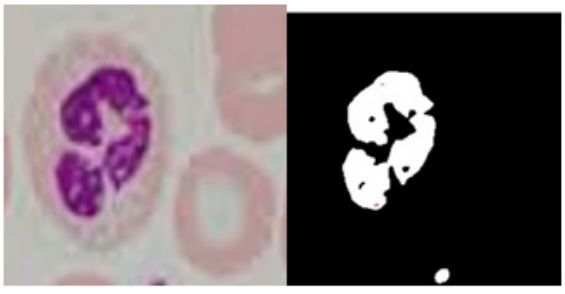


Figure 4. Original image of segmented image Eosinophil.

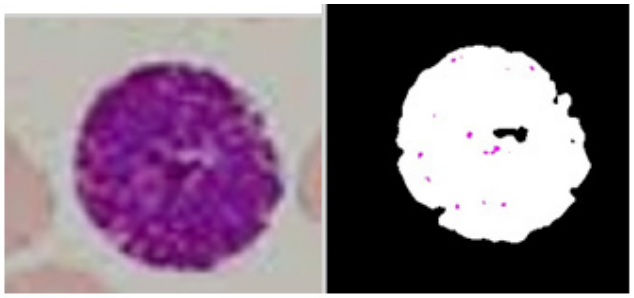


Figure 5. Original image of segmented image Basophil.

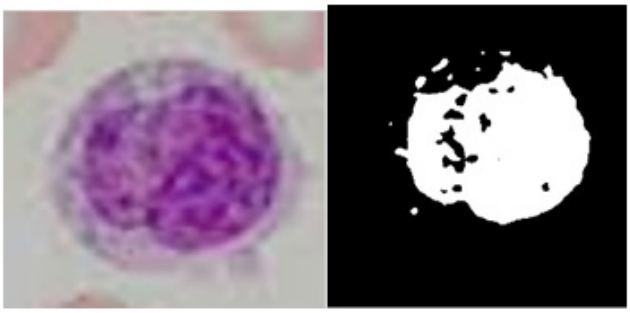


Figure 6. Original image of segmented image Monocytes.

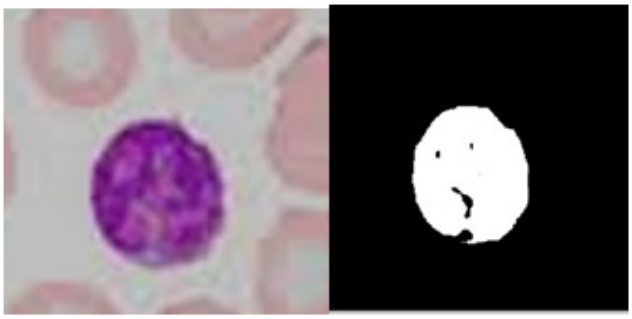


Figure 7. Original image of segmented image Lymphocyte.

Figure 3 shows original and segmented image of Neutrophil, Figure 4 is the original and segmented image of eosinophil, Figure 5 shows the original and segmented image of Basophil whereas Figure 6 is the original and its segmented image of Monocytes and Figure 7 is the original and segmented image of Lymphocytes.

4.1 Method Evaluation

To calculate segmentation performance quantitatively, the measurement defined below is used.

$$A_s = 100 * (A(proposed) \& A(expert))/(\max\{(A(proposed), A(expert)\})$$

Where $A_{proposed}$ is area of segmented nucleus by proposed result and A_{expert} is the area of nucleus determined by hematologist. Table 1 shows the results of segmentation accuracy.

Table 1. Accuracy Rate

Test Image	Existing	proposed	Accuracy (%)
Neutrophil	0.9405	0.9523	95.23
Eosinophil	0.9081	0.9257	92.57
Basophil	0.947	0.9674	96.74
Monocytes	0.967	0.9673	96.73
Lymphocytes	0.8886	0.9021	90.21

5. Conclusion

In this work, we propose a combined effect of K-means and Gram-Schmidt orthogonalization process. K-means is divide the input in to clusters and the desired color vectors were amplified by using Gram-Schmidt orthogonality theory. For final segmentation result we apply an AND operation. This method is very simple to implement and get an accurate results.

6. References

1. Leyza Baldo Dorini, Rodrigo Minetto and Leite NJ. Semiautomatic White blood cell segmentation Based on multiscale analysis. IEEE Journal of Biomedical and Health Informatics. 2013 January; 17(1):250-7.
2. Baidyanath Shah, Nilanjan Ray and Hong Zhang. Snake Validation: A PCA Based outlier detection method. IEEE Signal Processing Letters. 2009 June; 16(6):549-53.
3. Mohapatra S, Samanta SS and Dipti Patra. Fuzzy based Blood image segmentation for automated leukemia detection' IEEE. 2011.
4. Rezotofigi SH, Zadeh HS, Sharifian R, Zoroofi RA. A New Approach to white cell Nucleus segmentation Based on Gram-Schmidt orthogonalization. IEEE, International Conference on Digital Image Processing. 2009; p. 107-11.
5. Abd Halim NH, Mashor MY, Abdul Nasir AS and Rosline H. Nucleus Segmentation Technique for Acute Leukemia. IEEE 7th International colloquium on signal processing and its applications. 2011; p. 192-7.
6. Mohammed EA, Mohamed H, Naugler C, Far BH. Chronic Lymphocytic Leukemia cell segmentation from Microscopic Blood images using Watershed Algorithm and Optimal Thresholding. 26th IEEE Canadian conference of Electrical and Computer engineering. 2013. Available from: https://doi.org/10.1109/ccece.2013.6567770
7. Jun Duan and Le Yu. A WBC Segmentation Based on HIS colorspace. Proceedings of IEEE IC-BNMT. 2011; p. 629-32.
8. Francis EU, Mashor MY, Hassan R. Screening of Bone marrow slide images for Leukemia using multilayer perceptron. IEEE Symposium on indus-

trial electronics and applications. 2011; p. 643-8. Available from: https://doi.org/10.1109/isiea.2011.6108795.
9. Mostafa Mohamed, Far B. A Fast Technique for White blood cells Nuclei Automatic Segmentation Based on Gram-Schmidt Orthogonalization. IEEE 24th International Conference on tools with Artificial intelligence. 2012; p. 947-52.
10. Aimi Salihah, Mashor MY, Harun NH, Rosaline H. Improving colour Image Segmentation on Acute Myelogenous Leukemia using Contrast Enhancement Techniques. Kuala Lumpur: IEEE EMBS conference on Biomedical Engineering and Sciences. 2010; p. 246-51.
11. Monica Madhukar, Agaian S, Antony TC. Deterministic model for Acute Myelogenous Leukemia Classification. Seoul, Korea: IEEE International conference on systems, man and Cybermetics. 2012; p. 433-8.
12. Fabio Scotti. Robust Segmentation and Measurements Techniques of white cells in Blood Microscopic Images. Italy: IEEE, IMTC 2006 Instrumentation and Measurement Technology Conference. 2006; p. 43-8.
13. Mohapatra S, Dipti Patra and Kundan Kumar. Blood microscopic Image segmentation using Rough Sets. IEEE, International Conference On Image Information Processing. 2011.
14. Gonzaleza RC, Woods RE and Eddins SC. Digital image processing using Matlab. New Jersey, USA: Pearson Education, Inc.: Pearson Prentice Hall. 2004.
15. Haung DC, Hang KD, Chan YK. A Computer assisted Method for Leukocyte nucleus segmentation and Recognition in Blood smear Images. Elsevier: The Journal Systems and Software. 2012; 85:2014-8.