ISSN:0975-9646

A Survey on Image Segmentation and Feature Extraction Methods for Acute Myelogenous Leukemia Detection in Blood Microscopic Images

Athira Krishnan, Sreekumar K

Department of Computer Science & Engg, College of Engineering Poonjar, Cochin University of Science and Technology,Kottayam, Kerala, India

Abstract- Acute Myelogenous Leukemia (AML) is a fast growing cancer of the blood and bone marrow. The need for automation of leukemia detection arises since current methods involve manual examination of the blood smear as the first step toward diagnosis. This is time consuming, and also the accuracy of the method depends on the operator's ability. In this paper, various image segmentation and feature extraction methods used for AML detection are discussed.

***Key words*-Web Database, Annotation, Data alignment**

I. INTRODUCTION

White blood cells (WBCs) or leukocytes play a major role in the diagnosis of various diseases. So extracting information about them is valuable for hematologists. Leukemia is the cancer of blood and bone marrow. The bone marrow produces a large number of abnormal white blood cells in the case of leukemia. These cells are immature and they do not function properly. Without treatment, leukemia can be a deadly disease.

Leukemia is broadly classified as: 1) acute leukemia (which progresses quickly); and 2) chronic leukemia (which progresses slowly). The word "acute" in acute myelogenous leukemia denotes the rapid progression of the disease. It is called myelogenous leukemia because it affects a group of white blood cells called the myeloid cells, which in normal develops into different types of mature blood cells like white blood cells, red blood cells, and platelets.

Leukemia can be cured if it is detected and treated at the early stage. The need for automation of leukemia detection arises since current methods involve manual examination of the blood smear as the first step toward diagnosis. This consumes more time and also its accuracy is greatly dependent on operator's ability.

Many attempts have been made in the past to construct systems that aid in acute leukemia segmentation and classification. Image segmentation is a method which can be used to understand images and extract information or objects. It is the first step in image analysis. Feature extraction in image processing is a technique of redefining a large set of redundant data into a set of features (or feature vector) of reduced dimension. This transformation of the input data into the set of features is called feature extraction. The purpose of writing this paper is to provide a literature review in these areas of image segmentation and feature extraction for acute myelogenous leukemia detection.

This paper is structured as follows. Section II focuses in detail on the various image segmentation methods. Section III focuses on feature extraction. Section IV covers the conclusion.

II. IMAGE SEGMENTATION

The research on image segmentation has been a high degree of attention for many years. Different researchers have come up with different segmentation algorithms but till date there is not a single algorithm that is appropriate for all variants of images. As a result, algorithm that is developed for one set of images cannot be applied to a different set of images. So developing a single and unified approach for image segmentation that could be used for all sorts of images remains as a major challenge.

Over years many automatic segmentation and leukemia detection methods for blood smear images have been proposed. Otsu segmentation and automated histogram thresholding were employed to segment WBCs from the blood smear image in [8], [9], and [10].The work in [3] employed contour signature to identify the irregularities in the nucleus boundary. The work in [18] employed selective filtering to segment leukocytes from the other blood components. The work in [5] employed hue, saturation, and value (where hue represents color, saturation indicates the range of gray in the color space, and value is the brightness of the color and varies with color saturation), color space, and expectation–maximization algorithm (which consists of two steps, i.e., expectation and maximization steps) to identify the cytoplasm and nucleus of the WBCs. A watershed segmentation algorithm to segment nucleus from the surrounding cytoplasm of cervical cancer images was proposed by Nallaperumal and Krishnaveni [4].The work in [6] presented unsupervised color segmentation to bring out the WBC from acute leukemia images. A two-step segmentation process using HSV color model is used in [1].

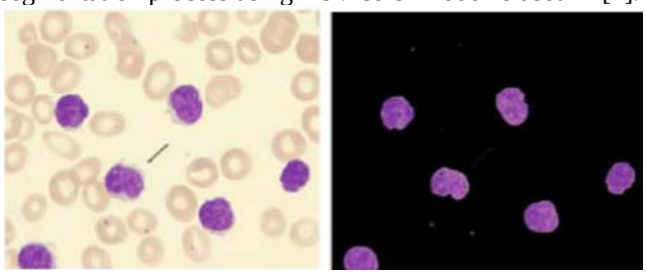

Fig 1: Sample image segmentation

Fig 1 shows sample image segmentation.

III. FEATURE EXTRACTION

Feature extraction in image processing is a method of transforming large redundant data into a reduced data representation. Transforming the input data into the set of features is called feature extraction.

A. *Fractal Dimension*

Fractals have been used in medicine and science earlier for various quantitative measurements [11] [12]. The fractal dimension D is a statistical quantity that gives an indication of how completely a fractal appears to fill space. The most important theoretical fractal dimensions are the Renyi dimension, the HD, and the packing dimension. Practically, the box-counting dimension is widely used. The procedure for Hausdorff Dimension measurement using box counting method [16] is introduced below as an algorithm:

- Each nucleus color (RGB) image is converted to gray and successively to binary image.
- Nucleus edge boundary is extracted using Canny [13] edge detection technique.
- A grid of R squares is superimposed over the edges, while counting the edge occupied squares.
- Step 3 is continued for an increasing number of squares.
- The Hausdorff Dimension HD may then be defined as follows:

$$HD = \frac{\log(R)}{\log(R(s))}$$

where R is the number of squares in the superimposed grid, and R(s) is the number squares or boxes (box count) which are occupied. Higher HD signifies higher degree of roughness.

B. *Local Binary Patterns (LBP)*

LBP is a type of feature used for classification in computer vision. It is a powerful feature for texture classification. The LBP feature vector is created as follows:

- The examined window is divided into cells (e.g. 16x16 pixels for each cell).
- For each pixel in a cell, compare the pixel to each of its 8 neighbours (on its right-top, right-middle, right-bottom, left-top, etc).
- Follow the pixels along a circle, i.e. either clockwise or counter-clockwise.
- If the center pixel's value is greater than the neighbour's value, write "1". Otherwise, write "0". This gives an 8-digit binary number (which is usually converted to decimal for convenience).
- Compute the histogram, over the cell, of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center).
- Optionally normalize the histogram.
- Concatenate histograms of all cells. This gives the window's feature vector.

C. *Texture Features*

Texture is a very useful characterization for a wide range of image. Texture features can be broadly classified into spatial texture feature extraction methods and spectral texture feature extraction methods based on the domain from which they are extracted. In the first approach, texture features are getting extracted by computing the pixel statistics or by finding the local pixel structures in original image domain, whereas the second approach transforms an image into frequency domain and then calculates feature from the transformed image.

TABLE I
COMPARISON OF TEXTURE FEATURES

Texture Method	**Advantages**	**Disadvantages**
Spatial texture	Meaningful, easy to understand, can be extracted from any shape without losing info.	Sensitive to noise and distortions
Spectral texture	Robust, need less computation	No semantic meaning, need square image regions with sufficient size

Table I [7] gives comparison of texture features. It summarizes advantages and disadvantages of spatial and spectral features.

D. *Shape Features*

Shape descriptors are a set of numbers that describe a given shape. Shape feature extraction techniques can be broadly classified into two groups [14], viz., contour based and region based methods. The first technique calculates shape features only from the shape boundary, while the second method extracts features from the entire region. In addition, spatial relationship is also considered in image processing that tells object location within an image or the relationships between objects. It includes following two cases: absolute spatial location of regions [15] and relative locations of regions [16][17].

E. *Color Features*

Color features are defined subject to a particular color space or model. There are different color spaces such as RGB, LUV, HSV and HMMD [19].Color feature can be extracted from images or regions once the color space is specified. Important color features include color histogram, color moments (CM), color coherence vector (CCV) and color correlogram. Among them, CM is one of the simplest yet very effective features.

Table II [7] provides a summary of different color methods with their advantages and disadvantages. DCD, CSD and SCD denote the Dominant Color Descriptor, Color Structure Descriptor and Scalable Color Descriptor respectively.

TABLE II
COMPARISON OF COLOR DESCRIPTORS

Color method	Advantages	Disadvantages
Histogram	Simple to compute, easy to use and understand	Sensitive to noise, high dimension, no spatial info
CM	Compact, robust	Not enough to describe all colors, no spatial info
CCV	Spatial info	High dimension, high computation cost
Correlogram	Spatial info	Very high computation cost, sensitive to noise
SCD	Compact on need, scalability	No spatial info, less accurate if compact
CSD	Spatial info	Sensitive to noise, rotation and scale
DCD	Compact, robust, perceptual meaning	Need post-processing for spatial info

IV. CONCLUSION

This paper explained the various image segmentation and feature extraction methods used for AML detection. The advantages and disadvantages for various techniques are also mentioned. Fractal dimension, local binary patterns, texture features, color features and shape features are discussed.

ACKNOWLEDGEMENT

I sincerely express my gratitude to everyone who supported me to prepare this paper. I am thankful for their aspiring guidance, invaluably constructive criticism and friendy advice during the project work. I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the work.

REFERENCES

[1] 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.
[2] S.Mohapatra, S. Samanta, D. Patra, and S. Satpathi, "Fuzzy based blood image segmentation for automated leukemia detection," in Proc. ICDeCom, 2011, pp. 1–5.
[3] S.Mohapatra, D. Patra, and S. Satpathi, "Image analysis of blood microscopic images for acute leukemia detection," in Proc. IECR, 2010,pp. 215–219.
[4] S.Mohapatra, D. Patra, and S. Satpathi, "Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images," in Proc. ICSMB, 2010, pp. 49–54.
[5] N.Sinha and A. G. Ramakrishnan, "Blood cell segmentation using EM algorithm," in Proc. 3rd Indian Conf. Comput. Vis., Graph., 2002, pp. 445–450.
[6] R.Rangayyan, Biomedical Image Analysis. Series Title: Biomedical Engineering. Boca Raton, FL, USA: CRC Press, Dec. 2004.
[7] Dong ping Tian, "A Review on Image Feature Extraction and Representation Techniques,"in International Journal of Multimedia and Ubiquitous Engineering,Vol. 8, No. 4, July, 2013,pp.385-396
[8] S. Suri, S. Setarehdan, and S. Singh, "Advanced Algorithmic Approaches to Medical Image Segmentatio: State-of-the-Art Application in Cardiology, Neurology, Mammography and Pathology" Berlin, Germany: Springer-Verlag, 2001, pp. 541–558.
[9] F.Scotti, "Robust segmentation and measurement techniques of white cells in blood microscope images," in Proc. IEEE Conf. Instrum. Meas.Technol., 2006, pp. 43–48.
[10] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines,"ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, p. 27, Apr. 2011.
[11] B. B. Mandelbrot," How long is the coast of Britain? Statistical self-similarity and fractional dimension", Science, 156:636 – 638, 1967.
[12] B. T. Milne, "Measuring the fractal geometry of landscapes. Applied Mathematics and Computation", 27:67 – 79, 1988.
[13] J. Canny, "A Computational approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence", 8(6):679 –698, 1986.
[14] D. Zhang and G. Lu, "Review of shape representation and description techniques", Pattern Recognition, vol. 37, no. 1, (2004), pp. 1-19.
[15] C.Yang, M. Dong and F. Fotouhi, "Image content annotation using Bayesian framework and complement components analysis", In Proc. ICIP, (2005).
[16] V.Mezaris, I. Kompatsiaris and M. G. Strintzis, "An ontology approach to object-based image retrieval", In Proc. ICIP, (2003), pp. 511-514.
[17] D. Zhang, M. M. Islam, G. Lu, et al., "Semantic image retrieval using region based inverted file", In Proc. DICTA, (2009), pp. 242-249.
[18] V.Piuri and F. Scotti, "Morphological classification of blood leucocytes by microscope images," in Proc. CIMSA, 2004, pp. 103–108.
[19] P.L.Stanchev, D. Green Jr. and B. Dimitrov. "High level colour similarity retrieval", International Journal of Information Theories and Applications, vol. 10, no. 3, (2003), pp. 363-369.