

Diagnosing Leukaemia Based on Geometric Characteristics of Cells in the Microscopic Blood Images

Jeyagopaal Janarththan, and Somaskandan Suthakar

Abstract — Suspected blood cancer patients need to undergo three different diagnostic exams: physical exam, blood tests, and bone marrow test. These exams are subjective and depend heavily on the experience of the technician. To reduce the complexities in the blood tests, an automatic method for diagnosing leukaemia based on geometric characteristics of cells in the blood sample images is proposed in this paper. The blood sample images are analysed in frequency domain to discriminate the leukemic and non-leukemic blood samples. Experiments are carried out on 162 blood samples, and the results are very promising as the true positive leukaemia detection rate is 100%.

Keywords—Fourier transformation, Image processing, Leukaemia, Blood cancer.

I. INTRODUCTION

BLOOD Cancer is one of the leading health problems worldwide and it is one of the most common cause of cancer-related death. There are many different types and sub-types of blood cancers and related blood disorders. They all affect, to a greater or lesser extent, the normal production of blood cells in the bone marrow and the normal function of blood cells circulating in the blood stream.

Primarily, blood cancer is divided into three basic types. Each of the variety may also include several variations, but in general this cancer is categorized as Leukaemia, Lymphoma and Myeloma [1][2].

- Leukaemia - It affects white blood cells (WBC). Blood tests are likely to present higher counts of white blood corpuscles.
- Lymphoma - Lymphoma is a type of blood cancer that affects lymphatic system. Lymphoma means it makes too many lymphocytes, a type of white blood cell (WBC). Lymphocytes are one of the varieties of white blood corpuscles.
- Myeloma - Myeloma (also called multiple myeloma) is a blood cancer of the plasma cells. Plasma cells are found in mammals bone marrow and produce antibodies which help fight infection. In myeloma, groups of abnormal plasma cells gather in bone marrow and stop it producing normal blood cells.

Jeyagopaal Janarththan and Somaskandan Suthakar are with Department of Computer Science, University of Jaffna, Sri Lanka

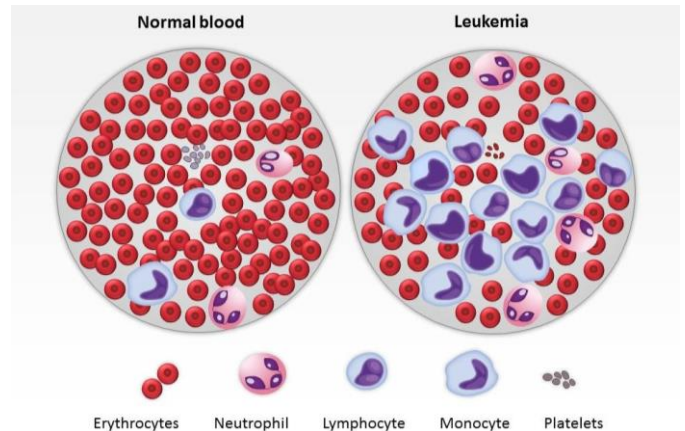


Fig. 1: Normal and Leukemic blood cells [8].

There is no single test that can accurately diagnose cancer. The complete evaluation of a patient usually requires a thorough history and physical examination along with diagnostic testing. Many tests are needed to determine whether a person has cancer, or if another condition (such as an infection) is mimicking the symptoms of cancer. Diagnostic procedures for cancer may include imaging, laboratory tests (including tests for tumour markers), tumour biopsy, endoscopic examination, surgery, or genetic testing.

Leukaemia begins in the bone marrow. It is caused by an excessive production of immature white blood cells that replace normal blood cells. White blood cells are bigger than red blood cells but fewer in number (the ratio between the two is around 1:700) [1]. If a blood sample has abnormal levels of white blood cells or platelets, which may suggest leukaemia (see Figure 1). Hence, by analysing the geometric structures of the blood sample image, the leukaemia can be identified.

In this paper we have proposed an automatic method to detect leukaemia based on the geometric structure of the blood sample. To discriminate the geometric structure of the non-leukemic and leukemic blood images, we have used the Fourier transformation technique.

II. FOURIER TRANSFORM

The Fourier transform is an important image processing tool which is used to decompose an image into its sine and cosine components. The output of the transformation represents the image in the Fourier or frequency domain, while the input image is the spatial domain equivalent. In the Fourier domain image, each point represents a particular frequency contained

in the spatial domain image [3].

The Discrete Fourier Transform (DFT) is the sampled Fourier transform and therefore does not contain all frequencies forming an image, but only a set of samples which is large enough to fully describe the spatial domain image. The number of frequencies corresponds to the number of pixels in the spatial domain image, i.e., the image in the spatial and Fourier domain are of the same size.

For an image $f(x,y)$ of size $M \times N$, the DFT is given by:

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi(ux/M+vy/N)} \quad (1)$$

The expression (1) will be computed for values of $u = 0, 1, \dots, M-1$, and also for $v = 0, 1, \dots, N-1$. The Fourier transform produces a complex number valued output image which can be displayed with two images, either with the real and imaginary part or with magnitude and phase. In image processing, often only the magnitude of the Fourier transform is displayed, as it contains most of the information of the geometric structure of the spatial domain image.

The origin of the Fourier transform corresponds to the average gray level of the image. The frequency components away from the origin of the Fourier transform correspond to faster gray level changes in the image such as the edges of objects and other components of an image characterised by abrupt changes in gray level.

III. PROPOSED METHOD

By looking at the non-leukemic and leukemic blood images (shown in Figure 1), it can be noticed that the relative intensity changing rate in the non-leukemic blood sample is very high than that of leukemic blood sample. Thus the Fourier transform would be a good choice to access this characteristic of a spatial domain image. Because the image in the Fourier domain is decomposed into its sinusoidal components, it is easy to examine or process certain frequencies of the image, thus influencing the geometric structure in the spatial domain.

Hence, it is reasonable to expect more higher-frequency Fourier components for non-leukemic blood images than that of the leukemic blood images, as the non-leukemic blood image has faster gray level changes.

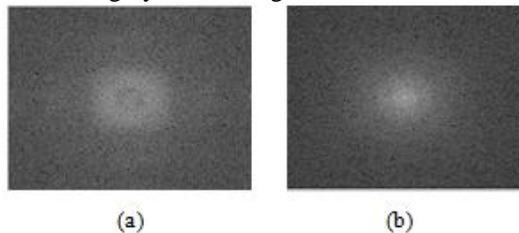


Fig. 2: General patterns of (a) non-leukemic blood samples, and (b) leukemic blood samples respectively.

By analysing more than 160 blood sample images, we found that the non-leukemic and leukemic blood images have specific Fourier patterns as shown in Figure 2.

To discriminate the two Fourier patterns clearly, some image processing techniques are performed before and after the Fourier transform. The whole process is depicted as a flow diagram shown in Figure 3.

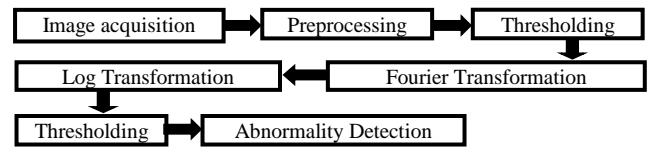


Fig. 3: Stages of the method

A. Image Acquisition

The first step in the process is acquiring the digital images of blood samples. Since we use the Fourier Transform as the main step, the quality of the images severely influence the results.

B. Preprocessing

In the preprocessing step, the image will be prepared for the Fourier transform step.

Colour to Grayscale

The acquired images will be usually colour images. To facilitate the subsequent steps, the colour images will be converted into grayscale images. Grayscale represents the intensity of the image. Grayscale images are obtained by eliminating the hue and saturation information while retaining the luminance [3].

Noise Removal

The noises in the acquired image will affect the Fourier transform severely. Hence, to eliminate the noise from the image, the median filtering method with 5×5 mask is used as it preserves edges while removing the noise [3][5].

Emphasizing Edges

To improve the results of later processing, unsharp masking is used to sharpen the image as it is a flexible and powerful way to increase the sharpness of the images [4][5].

The effect of the unsharp masking on the image is controlled using the following parameter values:

- Radius (standard deviation of the Gaussian low-pass filter) = 1
- Amount (strength of the sharpening effect) = 0.8
- Threshold (minimum contrast required for a pixel to be considered an edge pixel) = 0

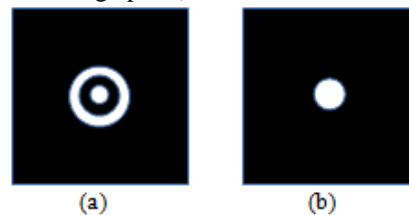


Fig. 4: (a) Expected pattern for non-leukemic blood, (b) Expected pattern for leukemic blood

C. Thresholding

Thresholding is used here for discriminating cells from the blood fluid. By selecting an appropriate threshold value, the gray level image can be converted to binary image. The binary image should contain all of the essential information about the position and shape of the blood cells. We use the Otsu's method [6] to compute a global threshold value for the thresholding function. The Otsu's method chooses the

threshold to minimize the intraclass variance of the black and white pixels. The advantage of obtaining a binary image is that it reduces the complexity of the image.

D. Fourier Transform

The thresholded image is taken for the Fourier transform. If the image size is $M \times N$, the same dimension of Fourier image with the wider range of frequency values will be obtained.

E. Log Transform

We need to scale the resultant values of the Fourier transform using log transform to expand the range of the dark pixels into the bright region so we can better analyse the frequency distribution [5].

F. Thresholding

The output from the log transform of the Fourier transformed image will usually have a range of frequency variations. To clearly discriminate the discontinuities in the Fourier values of the non-leukemic and leukemic blood samples, the thresholding is used. After thresholding the patterns shown in Figure 2, it is expected to have a result like the pattern as shown in Figure 4 (a) for non-leukemic blood images, or a pattern as shown in Figure 4 (b) for leukemic blood images.

Selecting an appropriate fixed threshold value was a challenging task. By analysing a number of log transformed images of the non-leukemic and leukemic blood samples, 0.65 has been fixed as the optimal threshold value.

G. Abnormality Detection

From the resultant binary image of the previous step, the values along the horizontal line passing through the point corresponds to the origin of the Fourier transform will be taken for the abnormality test. From these extracted array (say L) of 0s and 1s, a new array D will be constructed in such a way that

$$D(i) = L(i) - L(i-1)$$

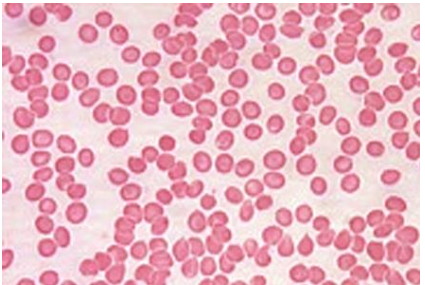
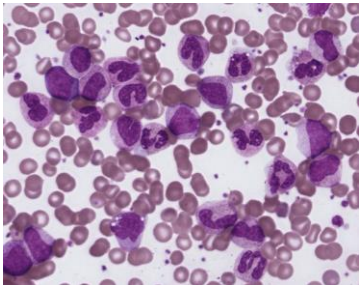
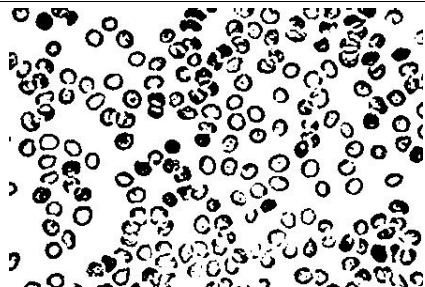
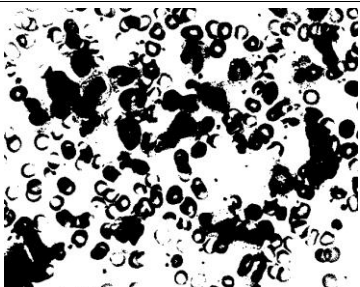
By counting the number of 1s in the array D , it can be decided that whether the blood sample is leukemic or non-leukemic. If there is only one 1, then we can decide it as a leukemic blood.

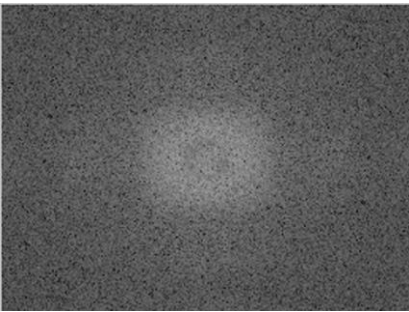
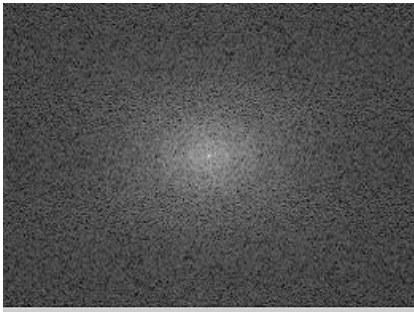

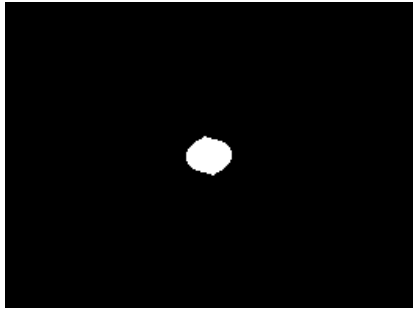
IV. RESULTS AND DISCUSSION

For the experimental analysis, we have obtained 162 true colour blood sample images from the teaching hospital, Jaffna, Sri Lanka. Among them, 93 are clinically identified leukemic blood samples and the rest 69 are normal blood samples. By eliminating the hue and saturation information while retaining the luminance, the grayscale images were obtained. To eliminate the noises, the median filter of size 5×5 is used. Then the unsharp masking is used for sharpening the edges of the cells. For the unsharp masking, we have fixed the following argument values: radius = 1, amount = 0.8, threshold = 0.

Since all the normal and all the leukemic blood samples produce more or less same patterns respectively, the intermediate results of one normal blood sample (Sample-1) and one leukemic blood sample (Sample-2) are shown in Table I.

TABLE I
INTERMEDIATE RESULTS FOR A NORMAL AND A LEUKEMIC BLOOD SAMPLES

Stages of the Detection Process	Sample-1	Sample-2
1) Acquired Image		
2) Stage A: Thresholded result of the preprocessed image		

3) Stage B: Log transformed image of the Fourier transform		
4) Stage C: Thresholded image of the Fourier transform		
5) Stage D: Number of 1s in the constructed array D.	3	1
6) Final decision	Non-leukemic	Leukemic

Two true-colour blood sample images are shown in the first row of Table I. After the preprocessing step, the Otsu's method is used to compute a global threshold value for the thresholding function.

It can be noticed that, the larger cell (i.e., white blood cell) density in sample-1 is very low than that of in sample-2, and hence the sample-1 has higher cell density. Thus, in the thresholded image of sample-1, we have larger black cluster of pixels, while for the sample-2, we only got many small black patches.

As a result, in the stage B, a Saturn planet shaped log transformed image of the Fourier transform is obtained for the non-leukemic blood sample. But for the leukemic sample, the frequency values are highly concentrated around the origin of the Fourier transform.

The stage C is obtained from the log transformed image by the thresholding function with the fixed threshold value of 0.65.

From the thresholded image values, an array L of values along the horizontal line passing through the centre of the image is extracted. From the extracted array L of 0s and 1s values, the array D is constructed such that

$$D(i) = L(i) - L(i-1)$$

For the sample-1, we obtained three 1s in the array D , as the binary log transformed image has three disconnected white regions along the horizontal line passing through the centre. On the other hand, we got only one 1 for the sample-2.

Since there are three 1s for sample-1, it is non-leukemic blood sample. But for the blood sample-2, we found only one 1, hence, it is leukemic blood sample. According to the clinical

records in the Teaching Hospital, Jaffna, the decisions are correct.

The Table II summarises the results for all of the 162 blood samples.

TABLE II
SUMMARY OF RESULTS OBTAINED

Sample type	Number of samples	Correct decisions	Wrong decisions
Non-leukemic	69	62	07
leukemic	93	93	00

To analyse the results, the terms *sensitivity* and *specificity* are very useful [7]. These measures are defined as follows:

- *Sensitivity*: refers to the ability of the test to correctly identify those patients with the disease.
 $Sensitivity = \text{true positives} / (\text{true positive} + \text{false negative})$
- *Specificity*: refers to the ability of the test to correctly identify those patients without the disease.
 $Specificity = \text{true negatives} / (\text{true negative} + \text{false positives})$

Where [9],

- *True positive (TP)*: the patient has the disease and the test is positive.
- *False positive (FP)*: the patient does not have the disease but the test is positive.
- *True negative (TN)*: the patient does not have the disease and the test is negative
- *False negative (FN)*: the patient has the disease but the test is negative.

Table III shows the sensitivity and specificity measures for the results obtained by the proposed method.

TABLE III
SENSITIVITY AND SPECIFICITY OF THE RESULTS

TP	TN	FP	FN	Sensitivity (%)	Specificity (%)
93	62	7	0	100	89.86

The 100% of sensitivity shows that no patients with leukaemia are classified as healthy person. This is very important from the clinical view point.

The specificity is also about 90%, which means that about 90% of the non-leukemic patients are correctly identified. The mistakenly identified 10% of patients will have to go for further examinations.

Therefore, the proposed automatic method can be used as a tool to filter out the non-leukemic patients, which would decrease the time a lot in leukaemia treatments.

V. CONCLUSION

A Fourier transform based automatic method for the detection of leukaemia has been presented. The results of the experiments show that the method is accurate and reliable.

Over 160 clinically analysed blood samples, obtained from the Teaching Hospital, Jaffna, Sri Lanka, have been considered for the experiments. From the experimental results, it has been shown that the proposed method is very accurate with the sensitivity of 100% and the specificity of about 90%.

With this 100% leukaemia detection rate, this method can be used as a tool in the first stage of the cancer treatments.

REFERENCES

- [1] N. A. Talukdar, D. Deb and S. Roy, "Automated Blood Cancer Detection Using Image Processing Based on Fuzzy System," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 8, pp. 467 - 472, 2014.
- [2] T. I. Mughal, J. m. Goldman and S. T. Mughal, *Understanding Leukemias, Lymphomas and Myelomas*, CRC Press, 2009.
<http://dx.doi.org/10.3109/9780203340943>
- [3] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*, Prentice Hall, 2007.
- [4] "Unsharp masking," Wikipedia, the free encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Unsharp_masking. [Accessed 22 October 2015].
- [5] R. C. Gonzalez, R. E. Woods and S. L. Eddins, *Digital Image Processing Using MATLAB*, 2nd ed, Gatesmark Publishing, 2009.
- [6] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62 - 66, 1979.
<http://dx.doi.org/10.1109/TSMC.1979.4310076>
- [7] A. Denniston and M. Philip, *Oxford Handbook of Ophthalmology (Oxford Medical Handbooks) 3rd Edition*, Oxford University Press, 2014.
- [8] "Gallery of What Is A Normal Red Blood Cell Count With Pictures," [Online]. Available: <http://www.vwmin.org/what-is-a-normal-red-blood-cell-count-with-pictures.html>. [Accessed 22 October 2015].
- [9] S. Suthakar and S. Mahesan, "A New Speed Function for Level set Based Deformable Model for Tumour Segmentation in Medical Images," *International Journal of Computational Intelligence and Informatics (IJCI)*, vol. 2, no. 2, pp. 124-131, 2012.