ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 8, August 2014

Connected Component based Medical Image Segmentation

Dr. Mrs. A.J.VYAVAHARE

Professor, E&TC Department, PES's MCOE, Pune, India[1]

Abstract: Medical image segmentation is a difficult problem due to the fact that these images commonly have poor contrast and missing details due to different types of noise and mostly medical images are fuzzy in nature, and segmenting regions based intensity is the most challenging task. Here region growing algorithm is used for segmentation in which selection of seed is important for that connected component is used. Our aim is to study anatomical structure, identify the region of interest, measure abnormality and help doctors in planning for early diagnosis. Connected component labelling works by scanning an image, pixel-by-pixel in order to identify connected pixel regions.\Blood Cell images for segmentation are taken. Leukaemia is a type of blood cancer, and if it is detected late, it will result in death. Leukaemia occurs when a lot of abnormal white blood cells produced by bone marrow. The existence of abnormal blood can be detected when the blood sample is taken and examined by haematologists. Microscopic images will be inspected visually by haematologists and the process is time consuming and tiring. Main objective of analysing through images is to gather information, detection of diseases, diagnosis diseases, control and therapy, monitoring and evaluation. At the end different parameters are calculated from quality metrics to define the accuracy of the algorithm.

Keywords: Leukemia, CLAHE, Connected Component, Region Growing Algorithm

I. INTRODUCTION

Assessment of blood cells smear is a normally experimental test these days and the haematologists are spent most of the time or paying attention on white blood cells (WBCs) merely. The recognition accuracy largely depends on subjective factors like experience and fatigue due to human tiredness. With the need for quality results, there arose a necessity for the automation of the whole process to reduce the burden on haematologists and to accurate results in significantly short period of time therefore new approach of automated detection is shown here. The count and shape, lineage and maturity level of white and red blood cells could aid in the diagnosis of diseases that range from inflammatory to leukaemia. Important information for correct patient diagnoses by Peripheral or marginal blood cell differential counting and therefore the microscopic review is effort exhaustive and requires a extremely trained or qualified expert or professionals. Blood cell images consist of red and white blood cells and also some platelets spread across the whole images. Therefore for the most part research in leukaemia WBC elements recognition and segmentation of nucleus is the important course of action in which the ultimate objective is to take out all the WBCs from a complex haphazard background and then segment the WBCs into their components, such as the nucleus and cytoplasm. In the past when automated system are not used very much Digital image processing techniques have helped to analyse the cells that lead to more accurate, standard, and remote disease diagnosis systems but there are a small amount of technical hitches to extracting the information from WBCs due to the spacious variation of cells in edge, shape, position and size. However there are various problems because of illumination is not fair [2]. The image contrast sandwiched between cell boundaries and the surroundings varies depending on the situation during the image capturing process. If the nucleus size increased then lymphocytes that become large it is called Reactive lymphocytes. The nucleus of a reactive lymphocyte can be round, elliptic, indented, cleft or folded. The cytoplasm is often abundant and can be basophilic. Normal lymphocyte and abnormal lymphocyte are shown in figure1and figure 2.

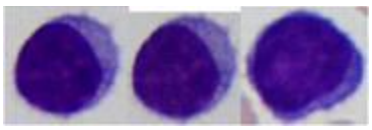


Fig. 1. Normal Lymphocyte

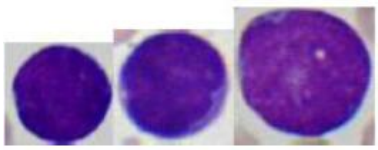


Fig. 2. Abnormal Lymphocyte

This study is focusing on WBC segmentation and detection using microscopic images by the digital microscope. So our main goal is to identify by contour wise WBCs elements and segment and detect each component of WBC and nucleuses and cytoplasm which has developed by means of digital image processing. The intention of the current learning is to build up an automatic tool which can identify, detect, segment and classify the white blood cells lymphocytes in digital microscopic images. Segmenting and classifying WBC was shown to be a difficult task due to various reasons including cell touching, close cell background intensities. In many of the researches presented in literature automatic cell segmentation was avoided to decouple the error due to segmentation with that of classification have done manual

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 8, August 2014

segmentation for all the acquired images to individually get WBC .images are performed in order to define the boundaries of the regions. For colour images RGB image is converted to Y Cb Cr colour space.

Preprocessing of images is performed by means of noise removal and contrast enhancement by contrast limited adaptive histogram equalization (CLAHE). After that connected component is used for seed selection. Connected component labeling of an image is a process of scanning images in order to identify connected pixel regions, which share the same property. Afterwards the seeds are grown to segment the image. After segmentation the images are compared using different parameters e.g. area, centroid, compactness

II. METHODOLOGY

In connected component based region growing algorithm the input is grey scale and colour medical images. The processing on different types of images consist of 3 main stages namely image pre-processing, seed selection using connected component & region growing that have been used in order to find out segmentation of blood cell images.

A. Contrast limited adaptive histogram equalization (CLAHE):

Contrast limited adaptive histogram equalization (CLAHE) is a technique to enhance the visibility of local details of an image by increasing the contrast in local regions. In CLAHE, the enhancement is controlled to avoid excess amplification of the noises in the local regions. CLAHE computes several histograms of intensity values, each corresponding to a distinct region of the image, distributes the histograms to avoid the excess amplification, and remaps the intensity values using the distributed histograms. CLAHE was originally developed for medical imaging and has proven to be successful for enhancement of low-contrast images such as portal films. The CLAHE algorithm partitions the images into contextual regions and applies the histogram equalization to each one. This evens out the distribution of used grey values and thus makes hidden features of the image more visible. The full grey spectrum is used to express the image. Contrast Limited Adaptive Histogram Equalization, CLAHE, is an improved version of AHE, or Adaptive Histogram Equalization. Both overcome the limitations of standard histogram equalization.

Algorithm Steps:

1. Obtain all the inputs: Image, Number of regions in row and column directions, Number of bins for the histograms used in building image transform function (dynamic range), Clip limit for contrast limiting (normalized from 0 to 1)
2. Pre-process the inputs: Determine real clip limit from the normalized value if necessary, pad the image before splitting it into regions
3. Process each contextual region (tile) thus producing gray level mappings: Extract a single image region, make a histogram for this region using the specified number of bins, clip the histogram using clip limit, and create a mapping (transformation function) for this region
4. Interpolate grey level mappings in order to assemble final CLAHE image: Extract cluster of four neighbouring mapping functions, process image region partly overlapping each of the mapping tiles, extract a single pixel, apply four mappings to that pixel, and interpolate between the results to obtain the output pixel; repeat over the entire image. The images processed with CLAHE, lesions appear obvious to the background and the image detail is very good. This algorithm is useful for radiologists to see subtle edge information, such as speculation.

B. Connected Components:

A set of pixels in an image which are all connected to each other is called a connected component. Finding all connected components in an image and marking each of them with a distinctive label is called connected component labeling. It identifies the connected components in an image and assigning each one, a unique label, creating a Label Matrix Connected components labelling scans an image and groups its pixels into components based on pixel connectivity, *i.e.* all pixels in a connected component share similar pixel intensity values and are in some way connected with each other. Since connected component labelling is a fundamental module in medical image processing, it improves the turn-around time of many medical diagnoses and procedures.

Algorithm:

Conditions to check:

1. Does the pixel to the left (West) have the same value as the current pixel?
    1. Yes – We are in the same region. Assign the same label to the current pixel
    2. No – Check next condition

2. Do both pixels to the North and West of the current pixel have the same value as the current pixel but not the same label?
    1. Yes –We know that the North and West pixels belong to the same region and must be merged. Assign the current pixel the minimum of the North and West labels, and record their equivalence relationship
    2. No – Check next condition

3. Does the pixel to the left (West) have a different value and the one to the North the same value as the current pixel?
    1. Yes – Assign the label of the North pixel to the current pixel
    2. No – Check next condition

4. Do the pixel's North and West neighbours have different pixel values than current pixel?
    1. Yes – Create a new label id and assign it to the current pixel

The algorithm continues this way, and creates new region labels whenever necessary.

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 8, August 2014

On the first pass:

1. Iterate through each element of the data by column, then by row
2. If the element is not the background
3. Get the neighbouring elements of the current element
4. If there are no neighbours, uniquely label the current element and continue
5. Otherwise, find the neighbour with the smallest label and assign it to the current element
6. Store the equivalence between neighbouring labels

On the second pass:

1. Iterate through each element of the data by column, then by row
2. If the element is not the background
3. Re-label the element with the lowest equivalent label [3].

Brief explanation of the algorithm is explained below. It will be easy to understand how actually the algorithm works to find the connected components.

Step1. The array from which connected regions are to be extracted is given below (8-connectivity based). We first assign different binary values to elements in the graph. Attention should be paid that the "0~1" values write on the centre of the elements in the following graph are elements' values. While, the "1, 2... 7" values in the next two graphs are the elements' labels.

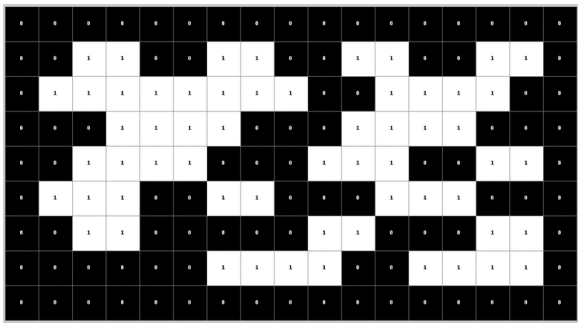


Fig.3 Binary Array

Step2. After the first pass, the following labels are generated. A total of 7 labels are generated in accordance with the conditions highlighted above.

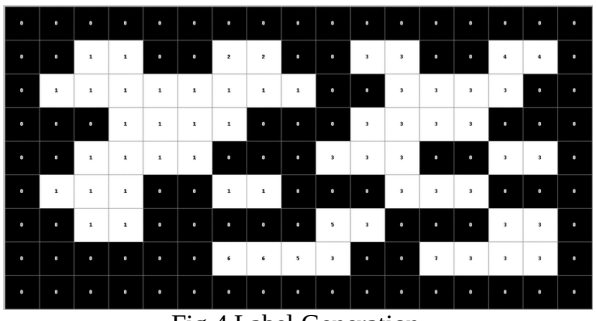


Fig.4 Label Generation

Step3. Array generated after the merging of labels is carried out. Here, the label value that was the smallest for a given region "floods" throughout the connected region and gives two distinct labels, and hence two distinct labels.

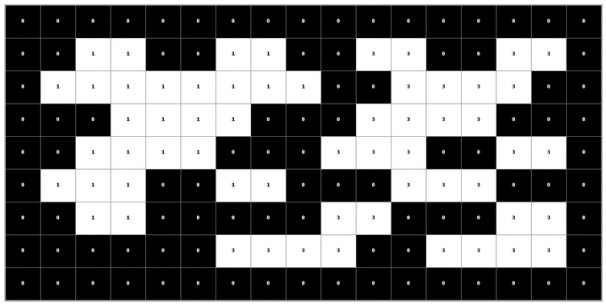


Fig. 5 Array generated after the merging of labels

In this way we get number of connected components in the image.

Step4. Delete the boundary-connected Regions:
Each connected component represents a possible lesion region. Besides the real lesion region, there are some regions connected with the boundary and such kind of boundary-connected regions always have a big area. We cannot simply delete all the regions connected with boundary of the image because sometimes the lesion region is also connected with the boundary. Therefore, we use the centre window to evaluate every boundary region. If a region has no intersection with the centre window and it is connected with any of the 4 image boundaries, we delete this region from the lesion candidate list. Each connected set of these pixels represents an initial region for region of interest. By choosing directly a set as seed rather than a single pixel, the process is speeded up and made robuster as more information about the region is available.

Step5. Rank the regions:
Now the left regions are either not connected with the boundary or having intersection with the image centre window. We use the following score formula to rank each left region. The one with the highest score is considered as the lesion region.

$$S_n = \frac{\sqrt{Area}}{dis\ (C_n, C_O).var\ (C_n)}, n = 1, \ldots., k \ldots\ldots\ldots\ldots \quad (1)$$

Where, k is the number of regions,
Area is the number of pixels in the region,
C_n Is the centre of the region.
C_O is the centre of the image, and
$var(C_n)$ is the variance of a small circular region centred at C_n.

In the implementation, we slightly moved the image centre C_O to the upper part of the image (around row/4) based on our observation that a lesion frequently appears in the upper part of an image and shadow frequently appears in the lower part of an image.

Step6. Determine the seed point:
Suppose the minimum rectangle contains the winning region [x_{min} x_{max} ; y_{min} , y_{max}]. For most cases, the centre of the winning region $((x_{min} + x_{max})/2,\ (y_{min} + y_{max})/2)$ could be considered as a seed point. However, there are cases that the lesion shape is irregular and thus the center point might be outside the lesion. For these special cases, we choose a seed point by the following rule:

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 8, August 2014

$x_{seed} = (x_{min} + x_{max}) / 2,$
$y_{seed} = \{(Vy| (x_{seed}, y) \in$ lesion region$\}$.

C. Region Growing Algorithm:

1. This method takes a set of seeds as input along with the image. (The seeds spot each of the objects to be segmented).
2. The regions are iteratively grown by comparing all unallocated neighbouring pixels to the regions.
3. The difference between a pixel's intensity value and the region's mean, ä, is used as a measure of similarity.
4. The pixel with the smallest difference measured this way is allocated to the respective region.
5. This process stops when the intensity difference between region mean and new pixel become larger than a certain threshold (t)
6. This process continues until all pixels are allocated to a region [35].

(a) $\bigcup_{i=1}^{n} R_{i=R}$.
(b) R_i is a connected region, i = 1,2,…, n
(c) $R_i \cap R_j = \phi$ for all i = 1,2,…, n
(d) P (R_i) = TRUE for i = 1,2,.., n
(e) P ($R_i \cup R_j$)_= FALSE for any adjacent region R_i and R_j.

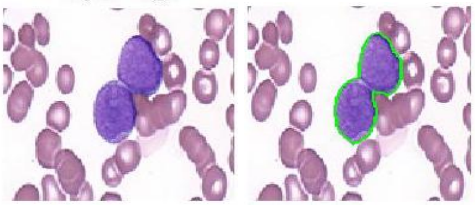
P (R_i) Is a logical predicate defined over the points in set R_i and ϕ is the null set.

a. Means that the segmentation must be complete; that is, every pixel must be in a region.
b. Requires that points in a region must be connected in some predefined sense.
c. Indicates that the regions must be disjoint.
d. Deals with the properties that must be satisfied by the pixels in a segmented region.
e. Indicates that region Ri and Rj are different in the sense of predicate P.

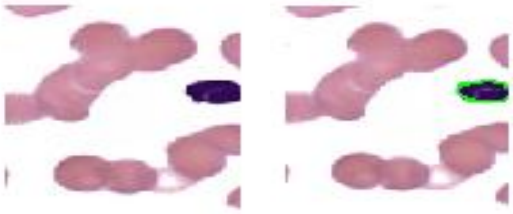
Region growing algorithm starts with a seed pixel, examines other pixels that surrounds it, determines the most similar one, and, if it meets certain criteria, it is included in the region. This process is followed until no additional pixels can be added. The region is iteratively grown by examining all unallocated neighbouring pixels to the region. The difference between a pixels intensity value and the regions mean is used as a similarity measure. The pixel with the smallest difference measured this way is allocated to the iteratively grown region.

This process is finished when the intensity difference between region mean and new pixel is larger than a certain threshold. Essentially, a seed pixel must be selected from which the region growing may commence. The approach selected here finds the minimum pixel from the image on the boundary. This is a dark pixel in most cases. Once the dark pixel coordinates are selected (background air), region growing commences in the form of marking all the pixels which lie in the 4-connected neighbourhood of the seed pixel. The process is repeated for each pixel until all the black pixels are marked in the image. This will deletes the dark regions that are connected to the border of the whole image.

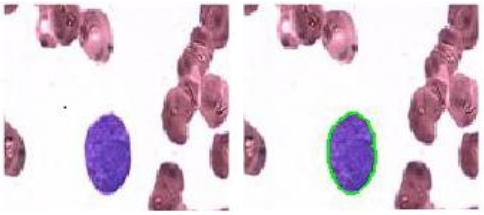
III. EXPERIMENTAL RESULTS



(a) Original image (b) segmented image



(c) Original image (d) segmented image



(e) Original image (f) segmented image

Fig. 6: (a) (c) (e) are the original images
(b) (d) (f) are the segmented images

All Segmented images above extract nucleus from its each element. Normal lymphocyte ratio of cytoplasm and nucleus is almost same but abnormality is due to the growth of nucleus and reduces cytoplasm namely reactive lymphocyte, thus segmentation of nucleus of a lymphocyte is shown in figure 6. In this approach shape features are extracted from the nucleus of WBC and a ratio between the area of nucleus and cytoplasm is also taken into account. Shape features are area, perimeter, compactness, eccentricity, solidity, form factor.

Table No. 1 shows the comparison between original & raw images Blood Cell with parameter

Type of image	Parameters	Blood cell-1	Blood cell-2	Blood cell-3
Original image of Blood Cell	Index	1	1	1
	Area	1492	317	11011
	Centroid	69.299 55.050	114.738 55.246	59.459 93.441
	Perimeter	190.267	72.627	125.720
	Solidity	0.87	0.966	0.974
	Compactness	0.930	0.899	0.850
	Orientation	70.075	-1.131	-89.829
	Elapsed Time	2.532	1.662	1.540
Raw Image of Blood Cell	Index	1	1	1
	Area	1492	316	1009
	Centroid	69.299 55.050	114.778 55.329	59.432 93.484
	Perimeter	190.267	74.627	125.740
	Solidity	0.857	0.924	0.973
	Compactness	0.930	0.897	0.853
	Orientation	70.075	0.003	-89.722
	Elapsed Time	0.360	0.346	0.344

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 8, August 2014

Table shows blood cell colour images to evaluate the performance of algorithm. Different parameters are calculated for original & ground truth image. Ground truth images are constructed from original images manually. Error is calculated by taking difference between the parameters of colour images & ground truth image of blood cell.

Table No.2 Error checking parameters for Blood Cell

Parameters	Blood cell-1	Blood cell-2	Blood cell-3
Area	0	-1	-2
Centroid	0 0	0.04 0.083	-0.067 0.043
Perimeter	0	2	0
Solidity	0	-0.042	-0.123
Compactness	0	-0.002	0.003
Orientation	0	1.134	0.107

Table No.2 is prepared for calculating error from Quality metrics of connected component based region growing algorithm of blood cell images. Here, error is calculated by taking difference between the parameters of colour images & ground truth image of brain. For proper segmentation calculated error should be zero or near to zero.

IV. CONCLUSION

1. Performance of algorithm is evaluated quantitatively by comparing the resulting of extractions with hand-drawn ground-truth images.
2. The successful segmentation and separation of an infected cell into its two sub regions containing cytoplasm and nuclei regions is vital, since these two regions contain features that correlate to the different forms of leukaemia. Algorithm successfully segments WBC images into nucleus, cytoplasm.
3. 3The results obtained show that the proposed method is able to identify in a robust way the WBCs present in the image, being able to properly classify all leukocytes suffering from disease and offering a good level of overall accuracy.
4. Therefore we can say that our systematic method of segmentation gives meaningful, reasonable and effective segmentation for blood cell images.

REFERENCES

[1] B. Senthilkumar1 "A Novel Region Growing Segmentation Algorithm for the Detection of Breast Cancer" 978-1-4244-5967-4/10/$26.00 ©2010 IEEE

[2] Prof.Samir K.Bandyopadhyay and Sudipta Roy "Detection of Sharp Contour of the element of the WBC and Segmentation of two leading elements like Nucleus and Cytoplasm", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol.2,Issue 1,Jan-Feb 2012,pp.545-5.

[3] Sharda. A. Chhabria & Ranjana S. Shende "Connected Component Algorithm for Gestures Recognition" G.H. Raisoni College of Engineering, Nagpur, Maharashtra-440016 (India). International Journal of Computer & Communication Technology

[4] Zhen gang Jiang1 "An Improved Image Segmentation Method Using Three-dimensional Region Growing Algorithm" No.7089, Weixing Road, Changchun, China 2013. The authors - Published by Atlantis Press

[5] Min Li, Xiaolin Zheng1 "Segmentation of brain tissue based on connected component labeling and mathematic morphology" 978-1-4244-9352-4/11/$26.00 ©2011 IEEE.

[6] Fauziah Kasmin, "Detection of Leukemia in Human Blood Sample Based On Microscopic Images: A Study "Journal of Theoretical and Applied Information TechnologyVol. 46 No.2 31st December 2012

[7] Adnan Khashman, "Image Segmentation of Blood Cells in Leukemia Patients" ISSN: 1790-5117 104 ISBN: 978-960-474-151-9

[8] Laurent Busin, Nicolas Vandenbroucke and Ludovic Macaire "Color spaces and image segmentation" Version: July 27, 2007

[9] Kentaro Kokufuta "Real-time processing of contrast limited adaptive histogram equalization on FPGA" i 305-8573 JAPAN, 978-0-7695-4179-2/10 $26.00 © 2010 IEEE DOI 10.1109/FPL.2010.37

[10] Martin Paralic "Fast Connected Component Labeling in Binary Images" 978-1-4673-1118-2/12/$31.00 ©2012 IEEE

[11] Shilpa Kamdi "Image Segmentation and Region Growing Algorithm", International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 1

[12] Sharda. A. Chhabria & Ranjana S. Shende "Connected Component Algorithm for Gestures Recognition" G.H. Raisoni College of Engineering, Nagpur, Maharashtra-440016 (India). International Journal of Computer & Communication Technology

[13] Mangipudi Partha Sarathi "Automated Brain Tumor Segmentation using Novel Feature Point Detector and Seeded Region growing", 978-1-4799-0404-4/13/$31.00 ©2013 IEEE

[14] Shilpa Kamdi "Image Segmentation and Region Growing Algorithm", International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 1.