Segmentation of Lungs in Chest Radiographs using Optimized Clustering Technique

Mary Jaya V J, S Krishnakumar

Abstract: ***Lung Cancer is considered the most common type of disease,*** **when** ***compared to that of the other types. It is one of the leading types of cancer that causes majority of the patient death worldwide. It is often detected only at the later stage as they are usually diagnosed at their advanced stages. Survival of patients with lung cancer is almost impossible. They often die within one year after the onset of clinical symptoms. Screening and early detection play an important role in saving the life of a patient. Chest radiography and computerized tomography scans are the techniques mostly used to diagnose and detect tumor in lungs. They require less radiation dose, and is available in most of the diagnostic centers. Their cost is also less when compared to the other techniques used for diagnosis. Nodule detection by using conventional radiographs is still not much effective, so there arises a need for alternative image processing techniques to improve the efficiency of detection. Image segmentation is considered as the first step in processing an image. Further analysis of the image would be made more effective if segmentation is efficient. There exits many segmentation algorithms based on clustering and thresholding approaches. In this paper, a bimodal, optimized and modified k-means algorithm is developed to segment the chest images.***

Keywords: Chest Radiographs, Segmentation, Thresholding, Clustering.

I. INTRODUCTION

Radiography is the branch of imageology which deals with the study and usage of various ionizing and non-ionizing radiations including x-rays and gamma rays. X-rays are the most common, effective and non-invasive efficient mechanism for diagnosis and treatment. It plays an important role in diagnosing serious lung disorders as well. Detection of lung cancer at an early stage is a challenging problem. Most people die within one year after the onset of clinical symptoms, so screening and early detection plays a crucial role in saving the life of a patient [1]. Many techniques such as Chest Radiographs (X-ray images), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scan exist to diagnose lung cancer. Most of these techniques are expensive and are advised only at a later stage that lung cancer detection is made possible only at their advanced stages. Pulmonary X-rays use a very small dose of ionizing radiation which is not harmful and helps to monitor treatment for a variety of lung conditions. They create images of sections inside the lungs which are useful for evaluating the possibilities of various lung disorders and illness. Although x-rays were used frequently from early days, they have their own limitations. Examination using x-rays may not necessarily help to find all problems in the chest. Certain conditions of the chest like small cancer cells, a blood clot in the lungs, pulmonary embolism and similar situations as such cannot be detected using conventional chest x-ray images. Image processing techniques provide better solutions for improving the manual analysis using chest radiographs. Further studies are necessary in the area of image processing for clarifying the results of a chest x-ray and to look into the abnormalities that are not visible. Image segmentation is considered the first and the most important step in image processing which assists doctors to diagnose and make decisions [2]. There exists a variety of techniques for lung segmentation of which thresholding is the simplest one. A much simpler method is to select the threshold using histogram equalization [3]. Segmentation based on classification of image pixels into different groups is termed clustering.

A. Segmentation

Methods that are used for extracting and representing the information from an image are termed segmentation techniques [4]. It serves as the pre-processing step for feature extraction. It divides an image into multiple set of pixels depending on their intensity values. To find out the abnormalities in the lungs, the primary task is to delineate the lungs from the chest x-rays. For further analysis of the image, an efficient segmentation method is required which makes the task easier. Prasantha et.al. [5], performed a study on various medical image segmentation techniques and found histogram based methods to be very efficient in terms of computation complexity and computation time, when compared to other image segmentation methods.

B. Thresholding

Thresholding is a method used in image processing to select an optimal gray-level value which separates objects of interest in an image. Binary images can be created from the gray-level ones by using thresholding technique.

Manuscript published on January 30, 2020.
* Correspondence Author
Mary Jaya V J*, Department of Computer Science, Assumption Autonomous College, Changanacherry, Kerala, India. Email: jaya.mary@gmail.com
S. Krishnakumar, Department f Electronics, School of Technology and Applied Sciences, MG University Research Centre, Edappally, Kochi, Kerala, India. Email: drkrishsan@gmail.com

This is achieved by converting the pixels in a gray-level image below the specified threshold value to zero and those above this value to one. At some global threshold T, let g(x, y) be the threshold version of f(x, y). Then the threshold version can be defined [6] as,

$$g(x, y) = 1;\ \text{if } f(x, y) \geq T$$
$$= 0;\ \text{otherwise}$$

Separation of the image from the background based on gray-level value is a difficult task in image processing. For developing image thresholding algorithms, gray-level histogram of image can be used as an effective tool [7]. The two types of thresholding techniques that exists [8] are (i) global thresholding in which the threshold value depends only on gray-level values of the image f(x, y) as in Otsu's method, and (ii) local thresholding where the threshold value depends on both gray-level values of the image f(x, y) and some local property of the image, p(x, y) as in k-means algorithm.

C. Clustering

Clustering is considered as an unsupervised learning problem. A cluster consists of a collection of objects which have similarity between them and are not similar to the objects that belongs to other clusters [9]. Clustering is performed on images by organizing pixels into groups having some similarities based on the given condition. The features that are useful in the given image to be segmented are retained by the algorithm and those attributes which do not contribute to homogeneous clustering in N-dimensional space are rejected [10]. Although there exist several clustering algorithms, due to its simplicity k-means is the one that is used mostly for segmentation.

In this work, images are acquired from the available public database. These images are then enhanced using histogram equalization. The enhanced images are then segmented to delineate the lung portion from the chest images. The main drawback of the existing k-means algorithm is that the centroid values are randomly chosen and the final quality of the image solely depend on the selection of the initial centroids. Random selection of initial centroids provides different results. A better k-means technique, which selects suitable centroid values for providing a segmented image with better quality is to be developed. With this intention, a modified k-means algorithm has been designed. This algorithm gives a better quality segmented output image without random selection of input centroid values.

The remaining part of this paper is organized as follows. Section II and III introduces Otsu's and k-means algorithm. Section IV describes about the developed method. Section V discuss about the results obtained. Finally, Section VI concludes the discussion.

II. OTSU'S ALGORITHM

Otsu's method is an optimized thresholding technique for image segmentation. In this method the threshold is selected by minimizing the within-class variance of the two groups of pixels that are separated by the thresholding operator and does not depend on modelling the probability density functions. It assumes a bimodal distribution of gray-level values [11]. The steps for computational algorithm is described below.

Step 1: Compute the histogram, n.

Step 2: Sum up all the histogram values.

$$\sum_{i=0}^{255} n_i = N$$

Step 3: Initialize maximal value as zero.

Step 4: Compute the probability of each intensity level.

$$p_i = \frac{n_i}{N};\ i=0 \text{ to } 255$$

Step 5: Probability θ_0 and mean value μ_0 of the background are,

$$\theta_0 = \sum_{i=0}^{t} p_i$$

$$\mu_0 = \frac{\sum_{i=0}^{t} i p_i}{\theta_0}$$

Step 6: Probability θ_1 and mean value μ_1 of the target are,

$$\theta_0 = \sum_{i=0}^{t} p_i$$

$$\mu_0 = \frac{\sum_{i=0}^{t} i p_i}{\theta_0}$$

Step 7: Variance is calculated as,

$$\sigma^2 = \theta_0.\theta_1(\mu_0 - \mu_1)^2$$

Step 8: Calculate the threshold value by updating the maximal value. Desired threshold corresponds to maximum standard deviation between the class's θ_0 and θ_1.

Step 9: Stop after segmenting the image using the obtained threshold

III. K-MEANS ALGORITHM

It is an unsupervised algorithm used to segment the area of interest from the background. It partitions the given data into k-parts based on the k-centroids and minimizes the sum of squared distances between all points and the cluster center.

Step 1. Choose the number of cluster k

Step 2: Calculate the distance between each data and cluster centre.

Step 3: Assign the data point to the cluster centre, such that its distance from the cluster centre is minimum.

Step 4. Compute and place the new centroid of each cluster.

Step 5. Reassign the data points to the new centroid which is closer. If any reassignment has occurred, go to step 4, otherwise, the model is ready, and the process is stopped.

Although k-means is easy to implement, it has some drawbacks. The quality of the final clustering results depend on the arbitrary selection of initial centroids [12].

Retrieval Number: E6858018520/2020©BEIESP
DOI:10.35940/ijrte.E6858.018520
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Random selection of the centroids, provides different result, so the initial centers are to be carefully chosen to get the desired segmented image. Several studies have been conducted till now to find the optimal centroid values for getting better quality output.

IV. MODIFIED K-MEANS

A new method for segmentation of lungs from chest radiograph images is developed using k-means clustering algorithm. Here the initial centroids are selected by using Otsu's method. In this method centroid values are obtained using mean values of the probability of each intensity level of the two classes. These mean values are then taken as the input centroids of the k-means algorithm, with k=2.

Step 1: Compute the histogram, n.

Step 2: Sum up all the histogram values.

$$\sum_{i=0}^{255} n_i = N$$

Step 3: Compute the probability of each intensity level.

$$p_i = \frac{n_i}{N}; \; i = 0 \text{ to } 255$$

Step 5: Centroids are taken as the probability mean value.

$$C_0 = \frac{\sum_{i=0}^{t} ip_i}{\theta_0} \quad \text{and} \quad C_1 = \frac{\sum_{i=t+1}^{255} ip_i}{\theta_1}$$

Step 6: The absolute value of the difference between each pixel value of the image and centroid value is,

$$d = \| p(x,y) - C_k \|$$

Step 7: Assign all the pixels to the nearest center based on distance d.

Step 8: The algorithm iterates between steps 6 and 7 until the sum of distances between the data points, and their corresponding centroid is minimized.

Step 9: Stop when none of the cluster assignment changes.

V. RESULTS AND DISCUSSION

The algorithm has been tested with about 150 images collected from the public database. Results are analyzed using Structural similarity index method (SSIM), Root mean square error (RMSE), Peak signal to noise ratio (PSNR) and Entropy. SSIM, is a method that is used for comparing quality measurement of one image with the other. It measures the similarity between the two images that are compared. The image that is being compared is regarded to have perfect quality [13]. Higher the value of SSIM, better is the quality of the image. Root-mean-square error (RMSE) is the measure of the differences between predicted values and the observed ones. The magnitudes of the errors in predictions for various times are aggregated into a single measure of predictive power. It is a measure of accuracy. Lower values of RMSE indicate better result. Entropy of a digital image is a statistical measure that expresses the randomness of gray-levels [14]. It specifies the uncertainty in the image values and measures the averaged amount of information required to encode the image values. An element with low entropy is more homogenous than that with a higher entropy and is used in combination with the smoothness criterion. PSNR is the ratio of maximum signal power to the power of corrupting noise. Higher value of PSNR indicates better quality of the reconstructed image.

Table 1 shows the resulting values of the parameters SSIM, RMSE, PSNR and entropy of various images for both the k-means method and the developed method. In the existing k-means method, the centroid values are not fixed as they are randomly selected. Accordingly the output also varies depending on the value of the centroids selected. After the analysis, it has been found that for different values of the two centroids the parameters also yield different results. The table lists the readings of the parameters with the two centroid values 87 and 170 which are selected randomly. As the centroids has to be input each time, these values vary depending on the choice of the programmer. In the developed method, there is no need to input the value of the two centroids. This method automatically takes the centroid values as 125.3341 and 255.0, which are the mean values of the probability of each intensity level for both the background and the target. With these centroid values the parameters for both the k-means method and the developed method is studied and tabulated.

Table 1 SSIM, RMSE and Entropy values

Method	Image	Entropy	SSIM	RMSE	PSNR
k-means	1	5.9014	0.3062	93.1237	8.7496
	2	5.7235	0.2733	93.4605	8.7182
	3	5.9057	0.2489	93.067	8.7549
	4	5.8241	0.2115	84.603	9.5831
	5	5.9067	0.2445	93.4640	8.7179
	6	5.8915	0.2099	92.6657	8.7924
	7	5.9014	0.3425	93.4024	8.7236
Modified k-means method (Developed method)	1	5.8823	0.4271	59.6539	12.618
	2	5.6571	0.3821	59.8261	12.593
	3	5.8702	0.4084	59.7573	12.603
	4	5.796	0.4806	48.4656	14.4221
	5	5.8609	0.3992	60.3618	12.5156
	6	5.8452	0.4050	58.0724	12.8514
	7	5.8592	0.4312	60.5660	12.4862

From the tabulated results of 150 images only a few are included in the table of this paper, as all the results obtained follows the same pattern of SSIM, RMSE, PSNR and Entropy. The segmented gray-level images obtained by k-means method and that obtained applying modified k-means are compared with the original gray-level image. When the results are analysed it is found that the developed method when compared with the existing k-means shows a low entropy range varying between 5.6571 and 5.8823, an increase in SSIM ranging from 0.3821 and 0.4806, RMSE values lowered within range 48.4656 and 50.5660, Rise in PSNR value ranging from 12.6180 and 12.4862. From all these results it can be concluded that the developed method which is so called the modified k-means method, shows a better performance for all these parameters.

Figure 1 shows the resulting images employing both k-means and the developed method. Figure 1(a) shows original images used for testing. Figures 1(b) and 1(d) show the gray-level images obtained by k-means segmentation and the developed method. Figures 1(c) and 1(e) portraits the binary images obtained by k-means segmentation and the developed method

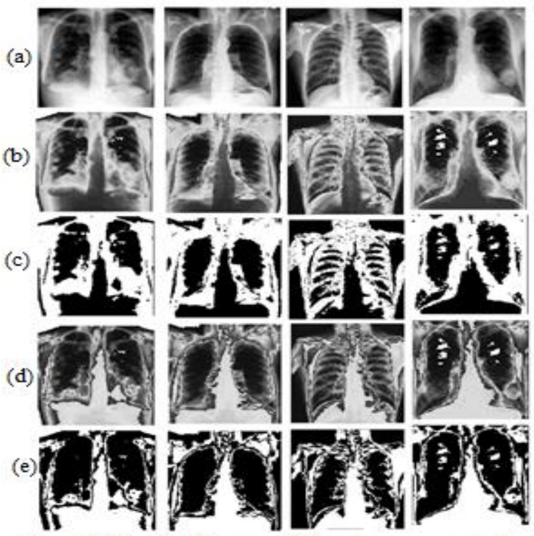

Figure 1 (a) original images (b) k-means segmented gray level images (c) k-means segmented binary images (d) segmented gray level images using developed method (e) segmented binary images using developed method.

VI. CONCLUSION

This paper describes a new optimized and modified k-means algorithm with centroid values of the clusters developed using Otsu's method. This method overcomes the drawback of existing k-means algorithm, where the centroid values has to be input randomly. The developed method has been compared with the existing k-means algorithm using Entropy, SSIM, RMSE and PSNR. The new algorithm developed shows better performance both visually as well as when parameters are analyzed. Initial centroid selection is optimized for better performance. It provides a better quality output. A better result is obtained with cluster value two (k=2).

ACKNOWLEDGMENT

I thank each and every individual who has supported me to carry out this work successfully. I express my sincere thanks to my colleagues for their kind help

REFERENCES

1. E. J. R. van Beek, B. Mullan, and B. Thompson, "Evaluation of a Real-time Interactive Pulmonary Nodule Analysis System on Chest Digital Radiographic Images," Academic Radiology, vol. 15, no. 5, pp. 571–575, 2008.
2. L. Cong, S. Ding, L. Wang, A. Zhang, and W. Jia, "Image segmentation algorithm based on super-pixel clustering," IET Image Processing, vol. 12, no. 11, pp. 2030–2035, 2018.
3. P. Kandhway and A. K. Bhandari, "An optimal adaptive thresholding based sub-histogram equalization for brightness preserving image contrast enhancement," Multidimensional Systems and Signal Processing, vol. 30, no. 04, pp. 1859–1894, 2019.
4. S. S. Chouhan, A. Kaul, and U. P. Singh, "Image Segmentation Using Computational Intelligence Techniques: Review," Archives of Computational Methods in Engineering, vol. 26, no. 03, pp. 533–596, 2018.
5. H. S. Prasantha, H L Shashidhara, K. N. B. Murthy and G Madhavi Lata, "Medical Image Segmentation" International Journal on Computer Science and Engineering vol. 02, no. 04, pp. 1209-1218, 2010.
6. Rogowska, Jadwiga, "Overview and Fundamentals of Medical Image Segmentation," Handbook of Medical Imaging, Academic Press Inc. ISBN: 978-0-12-077790-7, pp. 69–85, 2000.
7. Hetal J. Vala and Astha Baxi," A Review on Otsu Image Segmentation Algorithm", International Journal of Advanced Research in Computer Engineering & Technology, vol. 02, no. 02, pp. 387-389, 2013.
8. D. Liu and J. Yu, "Otsu Method and K-means," Ninth International Conference on Hybrid Intelligent Systems, Shenyang, pp. 344-349, 2009.
9. L. AbedAllah and I. Shimshoni, "k Nearest Neighbor Using Ensemble Clustering," Lecture Notes in Computer Science book series, vol. 7448, pp. 265–278, 2012.
10. G.B. Coleman and H.C. Andrews, "Image segmentation by Clustering", Proceedings of the IEEE, vol. 67, no. 05, pp. 773-785, 1979.
11. M. Basu, "Gaussian-based edge-detection methods-a survey," IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 32, no. 3, pp. 252–260, 2002.
12. N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm," Procedia Computer Science, vol. 54, pp. 764–771, 2015.
13. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.
14. Esley Torres, Yasel Garces, Osvaldo Pereira and Roberto Rodriguez, "Behavior study of entropy in a digital image through an iterative algorithm of the mean shift filtering", International Journal of Soft Computing, Mathematics and Control, vol. 4, no. 3, pp. 1-25, 2015.

AUTHORS PROFILE

Mary Jaya V J, Assistant Professor in Electronics, Department of Computer Science, Assumption College, Changanacherry. Doing Research in medical image processing at Mahatma Gandhi University Regional Centre, Edappally, Kochi-24, Kerala, India. Qualified M.Tech (Electronics) with specialization in Digital Electronics in 2005 from Cochin University of Science and Technology. Qualified M.Sc (Electronics) with specialization in Computer Technology in 2001 from Cochin University of Science and Technology. Areas of interests include image processing, neural networks.

Dr. S. Krishnakumar, completed his M.Sc. in Physics with Electronics specialization in 1987 from Mahatma Gandhi University, Kottayam. He was awarded Ph.D. in Thin Film Devices in 1995 from Mahatma Gandhi University, Kottayam. He got M.Tech. in Computer Science from Allahabad Agricultural Institute – Deemed University (renamed as Sam Higginbottom Institute of Agriculture, Technology and Sciences) in 2006 and also obtained MCA from IGNOU in 2010. He qualified AICTE-UGC special examination for PG Engineering in 2018. He has more than 25 years of teaching experience in Electronics and Computer Science subjects for graduates and post graduates. He served as the Regional Director at the School of Technology and Applied Sciences (STAS), during the period 2013-17. Currently he is working in the STAS, Mahatma Gandhi University Research Centre, Edappally, Kochi-24, Kerala, India. His research interest fields include ANN, Analog circuit design and Image processing. Dr. S. Krishnakumar is an Associate member of Institute of Engineers, India. He was a member of Board of studies of University of Calicut and a member of Academic council of Mahatma Gandhi University, Kottayam.