ORIGINAL ARTICLE

CT Lung Images Segmentation Using Image Processing And Markov Random Field

Khairul Azha A Aziz[1,2], M Iqbal Saripan[1], Fathinul Fikri Ahmad Saad[3], Raja Syamsul Azmir Raja Abdullah[1]

[1] Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2] Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, 76100 Melaka, Malaysia
[3] Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

Introduction: In this study, the performance of computed tomography lung image segmentation using image processing and Markov Random Field was investigated. Before cancer segmentation and analysis, lung segmentation is an important initial process. Thus, the aim of this study is to find the optimal Markov Random Field setting for lung segmentation. **Methods:** The Centre for Diagnostic Nuclear Imaging at UPM provided 11 anonymous sets of cancerous lung CT images for this study. The thresholding technique is an effective method for medical image segmentation when the priori information for the region of interest is known, such as the Hounsfield Unit value of lung. Due to the large differences in grey levels in the image, the thresholding approach is difficult to apply in segmentation, especially for lung. Thus, for the segmentation process, this study used multilevel thresholding with Markov Random Field with three settings; Iterated Condition Mode, Metropolis algorithm, and Gibbs sampler. The images then went through image processing procedures which were binarization, small object removal, lung region extraction and lung segmentation. The output from the experiments were analyzed and compared to determine the ideal lung segmentation setting. **Results:** The Jaccard index average values; Markov Random Field -Metropolis = 0.9464, Markov Random Field -ICM = 0.9499 and Markov Random Field -Gibbs = 0.9512. The Dice index average values; Markov Random Field - Metropolis = 0.9743, Markov Random Field - ICM = 0.9724 and Markov Random Field - Gibbs = 0.9749. **Conclusion:** Markov Random Field using Gibbs sampler delivered the best results for lung segmentation.

Keywords: Computed Tomography, Image Processing, Lung Segmentation, Markov Random Field

Corresponding Author:
Khairul Azha A Aziz, PhD
Email: khairulazha@utem.edu.my
Tel: +612-7295525

INTRODUCTION

Cancer is among the most fatal diseases in the world, with lung cancer having the highest mortality rate. Lung cancer also has the lowest survival rate following medical diagnosis, with the number of deaths rising every year. Cigarette smoking is responsible for more than 75% of lung cancer cases (1). According to a GLOBOCAN study, there were 19.3 million new cancer cases in 2020, with 52% of the patients dying from the disease. In comparison to 2018, the number has increased by 1.2 million. Hence, lung cancer is the leading cause of cancer deaths (2 – 4).

Imaging plays an important role in cancer detection and monitoring. Since Roentgen discovered x-rays in December 1895, radiography has been the most established of all imaging techniques (5)(6). However, over the past 40 years, new imaging techniques have emerged, including ultrasound, positron emission tomography (PET), computed tomography (CT), and magnetic resonance imaging (MRI). In the United States, the use of CT and MRI has considerably enhanced the detection and characterisation of soft-tissue lesions that may not be detectable on radiographs (7).

The motivation of developing a Computer Aided Diagnosis (CAD) tool is to assist medical practitioners in cancer screening process (8). Computer Aided Diagnosis (CAD) systems is a significant task in processing medical images required for analysis. The development of CAD system has led to better analysis for PET or CT images to find cancer or region of interest, which has helped to solve the problem of cancer detection (9 – 10).

The lung image basically consists of lung tissue, heart, air outside the lung tissue, and other elements. The portion of the lung that is examined is lung tissue, which is a component of the lung wall. The lower the Hounsfield Unit (HU) value, the lower is the lung tissue density. The image's higher HU value indicates a higher lung

wall density surrounding lung tissue. The initial stage in the computer-aided design (CAD) pulmonary diagnosis based on CT scans is to separate the lung area from the CT images (11).

Lung segmentation can be performed using a certain range value of Hounsfield Unit (HU) thresholds obtained from the CT volume histograms that define the lung regions (12). Image segmentation is a pre-processing technique for separating the lung region from the surrounding tissue in a scanned image. In the case of lung cancer detection, a mask image created during the preceding process is used to ensure that nodule identification is limited to the lung area. As a result, processing time and the number of false positives will be minimized (8).

Thresholding is a method used to segment an image by extracting the data from grayscale pixels that are higher or lower than a constant discrete level. The image's information that is required is kept, while the information that is irrelevant is ignored (13). To calculate optimal threshold, Otsu introduced between-class variance thresholding as a discriminant function to obtain an ideal threshold from the image histogram by segmenting an image into nearly uniform areas. The thresholding segmentation method has the disadvantage of not being functioning with multi-channel images and characters that aren't particularly different from the images. Grayscale images are similarly difficult to interpret since there are no noticeable distinctions or a larger overlap area of grey value images (11).

Thresholding is difficult in lung segmentation due to the large differences in grey levels in the image. Although the total CT value of lung parenchyma is lower, the density of lung wall around lung parenchyma is higher, resulting in a higher CT value in the image. To improve the normal thresholding technique, multilevel thresholding is applied for lung area border restoration using the mathematical morphology approach. Studies showed that the proposed method is reliable to automatically identify the lung regions from chest CT images (11).

MATERIALS AND METHODS

Methodology
CT is commonly used among many modalities for diagnosis and detection of cancer nodules because of its ability to represent human body scans in 3D form and its sensitivity (14). The first step to determine the lung area is by applying basic image processing technique. The next step is the segmentation of the image to extract the lung from the input image. The Centre for Diagnostic Nuclear Imaging at UPM provided 11 anonymous sets of cancerous lung CT images for this study. Based on the experimental results of the mathematical morphology technique, multilevel thresholds for lung area border restoration showed that the algorithm can automatically

and accurately separate lung regions from chest CT images (11).

Figure 1 shows the steps involved in the lung segmentation process in this study. For segmentation using Markov Random Field (MRF), Demirkaya et al. used Step 1 to Step 3 for region of interest segmentation (15). The input image is subjected to multilevel thresholding process, with the output serving as the MRF's initial segmentation. This study used three MRF settings; Iterated Condition Mode, Metropolis algorithm and Gibbs sampler. In addition, this study also followed steps 4 to 7 for the segmentation.
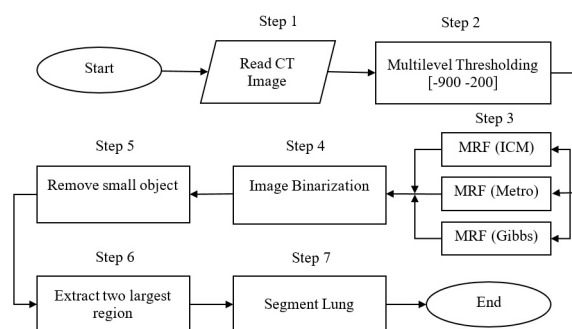
Figure 1: Lung Segmentation Process Flow

Thresholding technique was used because the priori information, such as the Hounsfield Unit value of lung was known. Thresholding is a method used for segmenting lung images by extracting the data from pixels in a grayscale image that are higher or lower than a constant discrete level. The image's information that is required is kept, while the information that is irrelevant is ignored. Table I shows the HU values of lung from various literatures. In this study, multilevel thresholding was used with the values -900 and -200 chosen as T1 and T2 for the settings.

Table I: Lung Hounsfield Unit (HU) Value

Organ	HU
Lung [14]a novel computer aided detection (CAD	-500
Lung [12]	-1023 to -400
Lung [16]	-900 to -200

Markov Random Field
MRF incorporates spatial data, eliminating both the issue of overlapping clusters as well as the impact of noise on clustering, impacting the clustering results. MRF is able to manage dynamic interdependence between data cases, offering a high degree of precision for segmentation tasks (17) (18). A Markov Random Field is a random field X of I that labels the sites of I with respect to the Ns if:

$P(X=x)>0$ for all $x \in \Omega$ (1)
$P(X=x_i | X=x_k, k \neq i) = P(X=x_i | X=x_k, k \in N$ (2)

P(X) described as MRF model,

$$P(X = x) = \frac{1}{Z}\exp\{-\beta U(x)\} \qquad (3)$$

P(Y | X) is the intensity distribution individual region,

$$P(Y = y \mid X = x) = \frac{1}{\sqrt{2\pi\sigma_s^2}}\exp\left\{-\frac{(y-\mu_s)}{\sigma_s^2}\right\} \qquad (4)$$

Where μs and σs are the parameters of distribution of the class xs.

The Iterated Conditional Modes (ICM) algorithm is an approximation of the MAP (maximum a posteriori). The ICM solves the issue of minimization by successively updating labels by lowering the following equation at each pixel:

$$\hat{x} = arg \quad \min_{x_s \in L}\left\{\frac{y-\mu_s}{\sigma_s} + \frac{1}{2}\log(2\pi\sigma_s^2) + \beta U(x_s)\right\} \qquad (5)$$

U(xs) is the number of pixels in the neighbourhood that have the colour xs

$$P(X = x_s) \infty \exp\left(\sum_{s,r \in C}\beta_{sr}\delta(x_s, x_r)\right) \qquad (6)$$

$$\text{where } \delta(x_s, x_r) = \begin{cases} 1 & x_s = x_r \\ 0 & x_s \neq x_r \end{cases} \qquad (7)$$

The main advantage of Metropolis algorithm is that the optimization's performance is no longer dependent on initial labelling. The next 6 steps involved were taken from (15), with the aim to minimize:

$$(\log\sigma_s + \mu_s)/\sigma_s + \beta\, U(x_s))/T \qquad (8)$$

where T is the temperature parameter.
Gibbs sampler: Gibbs sampler has the same advantage as Metropolis algorithm. There are 5 steps involves. For probability P,

$$P = \frac{\exp(-E(x_s)/T)}{\sum_{r \in N_s}\exp(-E(x_s)/T)} \qquad (9)$$

$$\text{where } E(x_s) = \frac{y-\mu_s}{\sigma_s} + \log(\sigma_s) + \beta U(x_s) \qquad (10)$$

Table II shows the MRF setting for CT lung segmentation used in this study. Figure 2 illustrates the CT images and the MRF process outputs for the three settings. The original image is shown in (a), the output from MRF-ICM is shown in (b), MRF-Metropolis is shown in (c), and the MRF-Gibbs is shown in (d).

Performance
The Jaccard similarity index and Dice similarity index were used as performance indicators for CT lung segmentation in this study based on Matlab documentation in (18).

Jaccard Similarity Coefficient
The Jaccard similarity coefficient of two sets A and B (also known as intersection over union or IoU) is expressed as:
jaccard(A,B) = |intersection(A,B)|/|union(A,B)| (11)

where |A| represents the cardinal of set A. The Jaccard
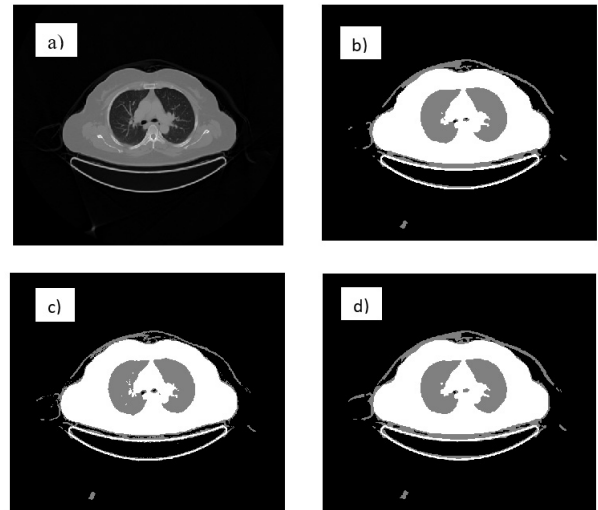
Figure 2: (a) CT Image, (b) MRF-ICM, (c) MRF-Metropolis and (d) MRF Gibbs

Table II: MRF Setting for CT Lung Segmentation

Markov Random Field	Number Of Class	β
ICM	3	4
Metropolis algorithm	4	3
Gibbs sampler	4	2

index can also be expressed in terms of true positives (TP), false positives (FP) and false negatives (FN) as:
jaccard(A,B) = TP / (TP + FP + FN)   (12)

Dice Similarity Coefficient
The Dice similarity coefficient of two sets A and B is expressed as:
dice(A,B) = 2 *|intersection(A,B)|/(| A | + | B |)   (13)

where |A| represents the cardinal of set A. The Dice index can also be expressed in terms of true positives (TP), false positives (FP) and false negatives (FN) as:
dice(A,B) = 2 * TP / (2 * TP + FP + FN)   (14)

RESULTS

Table III shows the Jaccard index similarity and Dice index similarity for CT lung segmentation. The averages for Jaccard index were MRF-Metropolis = 0.9464, MRF-ICM = 0.9499 and MRF-Gibbs = 0.9512. For the Dice index, the average values were MRF-ICM = 0.9743, MRF-Metropolis = 0.9724 and MRF-Gibbs = 0.9749. Table IV shows the comparison for lung segmentation using MRF with previous study done using FCM, Modified FCM and FPCM. Figure 3 shows the image for binarization, small object removal, extraction of two largest region and lung segmentation process.

DISCUSSION

Referring Table III, both similarity indexes showed that MRF-Gibbs produced the best results. Using MRF-Gibbs

Table III: Jaccard and Dice similarity index

Patient	Jaccard Similarity (JI)			Dice Similarity (DI)		
	ICM	Metro	Gibbs	ICM	Metro	Gibbs
1	0.9568	0.9512	0.9609	0.9779	0.9750	0.9801
2	0.9449	0.9499	0.9531	0.9716	0.9743	0.9760
3	0.9587	0.9492	0.9600	0.9789	0.9739	0.9796
4	0.9632	0.9482	0.9579	0.9813	0.9734	0.9785
5	0.9373	0.9268	0.9363	0.9676	0.9620	0.9671
6	0.9531	0.9436	0.9478	0.9760	0.9710	0.9732
7	0.9593	0.9670	0.9549	0.9792	0.9832	0.9769
8	0.9404	0.9322	0.9293	0.9693	0.9649	0.9633
9	0.9679	0.9724	0.9685	0.9837	0.9860	0.9840
10	0.9344	0.9475	0.9581	0.9661	0.9730	0.9786
11	0.9330	0.9227	0.9359	0.9654	0.9598	0.9669
Average	0.9499	0.9464	0.9512	0.9743	0.9724	0.9749

Table IV: Comparison with other methods

Method	JI	DI
Standard FCM (19)such as radiography, computed tomography (CT	0.8603	0.9249
Modified FCM (19)such as radiography, computed tomography (CT	0.8950	0.9446
Fuzzy Possibilistic C-Means Clustering FPCM (19)such as radiography, computed tomography (CT	0.9250	0.9610
MRF ICM	0.9499	0.9743
MRF Metropolis	0.9464	0.9724
MRF Gibbs	0.9512	0.9749
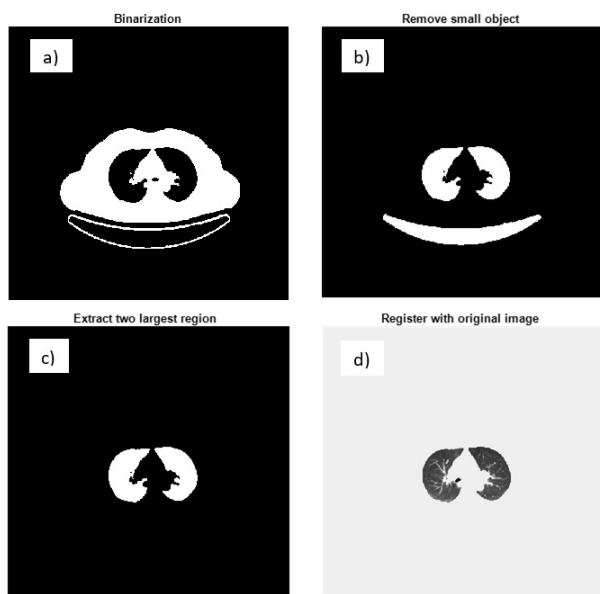
Figure 3: (a) Binarization, (b) Small object removal, (c) Two largest region extraction, (d) Image registration

to calculate Jaccard index similarity yielded an average value of 0.9512. This was 0.0013 higher than MRF-ICM and 0.0048 higher compared to MRF-Metropolis. As for Dice index similarity, MRF-Gibbs gave an average value of 0.9749. This value was 0.0006 higher than MRF-

ICM and 0.0025 higher compared to MRF-Metropolis. For MRF-Metropolis, the method tended to choose the nearest distance between nearby pixels. As discussed by I. B. Sullivan et. al. (19), the disadvantage of this approach is that it can become stuck at a local minimum. In comparison to the previous studies, results in Table IV clearly illustrate that using MRF with all three settings produced better results. This was in line with D. L. Pham's suggestion that MRF could be used to efficiently segment medical images with inhomogeneous intensity (20). As can be seen in Figure 3, the algorithm was able to successfully segment the lung region perfectly. This proved that the new algorithm improved the normal thresholding technique and can automatically and accurately segment lung regions from the chest CT images.

CONCLUSION

The aim to study the performance of computed tomography lung image segmentation using image processing and Markov Random Field was achieved. A CT image for lung segmentation using image processing and MRF was presented. Overall, the results were better than previous reported studies for CT lung segmentation. As shown in the results section, the algorithm managed to get the similarity index for all the three settings to exceed 0.9400. Moreover, MRF using Gibbs sampler gave the best results with the Jaccard index of 0.9512 and the Dice index of 0.9749 compared to MRF-Metropolis and MRF-ICM. The results from this study will benefit the lung cancer analysis research and computer aided diagnosis development.

ACKNOWLEDGEMENTS

REFERENCES

1. D. Sharma and G. Jindal, "Identifying Lung Cancer Using Image Processing Techniques," Int. Conf. Comput. Tech. Artif. Intell., pp. 115–120, 2011.
2. Ms. Hyuna Sung, PhD 1; Jacques Ferlay, MSc, ME2; Rebecca L. Siegel, MPH 1; Mathieu Laversanne and P. Isabelle Soerjomataram, MD, MSc, PhD2; Ahmedin Jemal, DMV, PhD1; Freddie Bray, BSc, MSc, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA CANCER J CLIN, vol. 0, no. 0, 2021.
3. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA. Cancer J. Clin., vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.
4. L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-

tieulent, and A. Jemal, "Global Cancer Statistics, 2012," CA a cancer J. Clin., vol. 65, no. 2, pp. 87–108, 2015, doi: 10.3322/caac.21262.

5. J. E. Adams, Z. Mughal, J. Damilakis, and A. C. Offiah, Radiology. 2012.

6. L. E. Swischuk, "Emergency Pediatric Imaging: Current Status and Update," Semin. Ultrasound, CT MRI, vol. 28, no. 2, pp. 158–168, 2007, doi: 10.1053/j.sult.2007.01.010.

7. H. Carty, F. Brunelle, D. A. Stringer, and S. Kao, Imaging Children. 2nd ed. Edinburgh: Elsevier Churchill Livingstone; 2005. .

8. K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," Med. Image Anal., vol. 13, no. 5, pp. 757–770, 2009, doi: 10.1016/j.media.2009.07.001.

9. D. S. Elizabeth, C. S. Retmin Raj, H. K. Nehemiah, and A. Kannan, "Computer-aided diagnosis of lung cancer based on analysis of the significant slice of chest computed tomography image," IET Image Process., vol. 6, no. 6, pp. 697–705, 2012, doi: 10.1049/iet-ipr.2010.0521.

10. F. Ciompi et al., "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," Med. Image Anal., vol. 26, no. 1, pp. 195–202, 2015, doi: 10.1016/j.media.2015.08.001.

11. A. Karthikeyan and M. Valliammai, "Lungs Segmentation using Multi-level Thresholding in CT Images," Int. J. Electron. Comput. Sci. Eng., vol. 1, no. 3, pp. 1509–1513, 2012.

12. C. Ballangan, X. Wang, D. Feng, S. Eberl, and M. Fulham, "Lung segmentation and tumor detection from CT thorax volumes of FDG PET-CT scans by template registration and incorporation of functional information," IEEE Nucl. Sci. Symp. Conf. Rec., pp. 5349–5353, 2008, doi: 10.1109/NSSMIC.2008.4774441.

13. S. Kalaivani, P. Chatterjee, S. Juyal, and R. Gupta, "Lung cancer detection using digital image processing and artificial neural networks," Proc. Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2017, vol. 2017-Janua, pp. 100–103, 2017, doi: 10.1109/ICECA.2017.8212773.

14. T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," Med. Image Anal., vol. 14, no. 3, pp. 390–406, 2010, doi: 10.1016/j.media.2010.02.004.

15. O. Demirkaya, M. H. Asyali, and P. K. Sahoo, Image Processing With Matlab. Applications in Medicine and Biology. 2009.

16. M. S. Fasihi and W. B. Mikhael, "Overview of Current Biomedical Image Segmentation Methods," Proc. - 2016 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2016, pp. 803–808, 2017, doi: 10.1109/CSCI.2016.0156.

17. A. Norouzi et al., "Medical image segmentation methods, algorithms, and applications," IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India), vol. 31, no. 3, pp. 199–213, 2014, doi: 10.1080/02564602.2014.906861.

18. "Jaccard and Dice similarity index," Mathworks. https://www.mathworks.com/.

19. I. B. Sullivan, "The Metropolis Algorithm," Comput. Sci. Eng., vol. 2, no. 1, pp. 65–69, 2000.

20. D. L. Pham, C. Xu, and J. L. Prince, "CURRENT METHODS IN MEDICAL IMAGE SEGMENTATION," Annu. Rev. Biomed. Eng., vol. 2, pp. 315–337, 2000.