Indian Journal of Science and Technology, Vol 9(9), DOI: 10.17485/ijst/2016/v9i9/80313, March 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Improved Hybrid Algorithm for Accurate Determination of Parameters of Lung Nodules with Dirichlet Boundaries in CT Images

G. Niranjana[1*] and M. Ponnavaikko[2]

[1]SRM University, Chennai - 603203, Tamil Nadu, India; niran_janag@yahoo.com
[2]Bharath University, Chennai - 600073, Tamil Nadu, India; ponnav@gmail.com

Abstract

Background/Objectives: Segmentation of lung nodules with irregular boundaries in CT images remains a challenge. This work aims to advance the accuracy of segmentation using Random walker and watershed algorithm. **Methods/Statistical Analysis**: The input image is considered as a graph representing each pixel as a node. Two seed points which are user-defined (pre-labeled) pixels given as labels, one for the foreground and the other for the background. The gradient of the seed points are calculated. Then the probability of reaching the labeled pixels from each of the unlabeled pixels is obtained and a vector of probabilities is defined for each of the unlabeled pixels. **Findings:** The calculated vector of probabilities for each unlabeled pixel is combined and they can be assigned to one of the labels using the watershed algorithm to obtain tumor segmentation. We used 23 images for validating our method and our experiment compared the original random walker algorithm, random walker with improved weights and watershed segmentation results. Resulted images have maximum of DSC values as 0.92 for Random Walk, 0.94 for Random Walk with Improved Weight and 0.97 for Watershed combined. **Applications/Improvements:** Accurate segmentation of nodules with dirichlet boundaries in CT images using minimum number of seed points.

Keywords: CT Images, Dirichilet Boundaries, Lung Nodules, Segmentation

1. Introduction

Lung cancer is one of the most common deadliest diseases. According to the latest survey reported[1], a total of 159,260 people had died due to lung cancer in US. In India every year 63,000 new lung cancer cases are being reported. In the latest survey of WHO, the mortality rate of lung cancer is higher than any other cancer**.** It is very difficult to analyze the cancer at its early stage. Various Computer Aided Diagnosis (CAD) systems as reported in[1] have been designed for the early diagnose of lung tumor. Early detection of the lung tumor can increase the survival rate of 1 to 5 years. Hence a proper method for detection and classification of lung tumor is the need of the hour.

Most segmentation methods have an automatic implementation. However, automatic segmentation technique doesn't always provide accurate results, and since the tumor size[2] and position can be distinct with different pixels range[3,4]. Since interactive segmentation methods use the user's guidance, segmentation results tend to be more accurate. Hence interactive segmentation techniques are used in medical image processing[5,6].

An improved hybrid approach[6,7] which segments the image interactively using the random walker algorithm with modified weights and watershed algorithm is proposed in this paper. In the random walker algorithm proposed by Leo Grady[8], N number of pixels is selected by the user as labels for different objects. For each of the unlabeled pixels the propability of reaching each of the

*Author for correspondence

labeled pixels are calculated. The proposed method differs from the original random walker algorithm in the way that the proposed method obtains a N-tuple vector of probabilities for each unlabeled pixel and is combined with watershed algorithm. The resultant image produced is segmented using the watershed algorithm with more accurate. The resulting image produced is segmented more precisely between the objects and boundaries using the watershed algorithm[9]. The accuracy of the segmentation can be evaluated using Dice Similarity Coefficient (DSC). It is used as a statistical validation metric to evaluate the difference between manual segmentation and automated segmentation of images.

Compared to original Random walker our approach has the following advantages:

- Accurate segmentation of nodules with dirichlet boundaries.
- Use of constant value instead of free parameter β.
- Accuracy is increased with minimum of seed points.

Rest of the paper is structured as follows. The proposed method is described in the next section. Materials and methods for the proposed work are discussed in Section 3 followed by Results and Conclusion in Section 4 and 5.

Segmentation is a crucial important step in a CAD system. There are numerous segmentation algorithms available and they are basically classified into 4 categories: 1. Thresholding based methods, 2. Region based methods, 3. Learning based methods and 4. Boundary based methods[10]. This paper addresses a graph – based segmentation approach based on the code of random walks combined with the watershed algorithm. This paper focuses on region based random walker segmentation and boundary based watershed segmentation algorithm.

In region based segmentation methods the object boundaries are determined based on the homogeneity of the image. For detecting boundaries the region-based segmentation methods uses the intensities of the image. The region-based methods are divided into two categories: Region Growing and Graph based methods[6]. In Region growing technique spatial information is also included along with the intensity information[11,12]. The algorithm starts at the user defined seed point and the mean and the standard deviation of the connected pixels are calculated.

Based on the calculated values, the connected pixels are either included or excluded in the segmentation. Another parameter, homogeneity metric, is used to find the difference between the new pixel and of the region already selected and to decide the new pixel is also still included in the segmentation[13]. The same procedure is repeated until the entire region of interest has been segmented or otherwise the segmented region does not change anymore. Region growing methods work well in identical regions with properly set intensity identical parameters, but it failed to segment non identical structures. Region growing fairly depends on the homogeneity parameter that is set in the algorithm. When it is not properly set the algorithm may fail even for adequately homogenous uptake regions[14,15].

Graph-based approaches may be supervised or unsupervised. It is superior over other segmentation methods in the way it efficiently recognizes the segmentation with the use of foreground and background seed points to segment the objects in the image[11,14]. To get optimal segmentation results these seed points act as vital constraints and the global information is combined with local pair wise pixel similarities. The most commonly used graph based methods are Graph cuts and Random Walk techniques[7,10,11].

The graph cuts[9] technique is a method developed for interactive and seeded segmentation. In this method the image is viewed as graph and the intensity changes are represented as weights. Some nodes are marked as foreground and the rest are treated as background. The algorithm aims at finding the minimum weight cut between the two nodes by performinga max-flow/min-cut analysis. With enough user input an arbitrary segmentation may be obtained using this algorithm[16]. The limitation with this technique is that if there are minimum number of seeds this technique may experience a "small cut" problem[17]. Small cut problem occurs when the algorithm returns the cut that minimally separates the seeds from the rest of the image. To overcome this problem we need to place seed continuously. Also, the K-way graph cuts problem is NP-Hard and it requires the use of a heuristic for its solution. When the foreground and background are different in an image then multiple small cuts may exist. Hence a small amount of noise will affect the segmentation result drastically.

In[8] proposed a semi supervised random walk approach over a weighted graph for interactive image segmentation. Here the unlabeled pixels are assigned the label of the node to which it can easily reach. This algorithm works well on different types of images but the placement of the labels in the image influences the performance of the algorithm.

2. Proposed Method

The flowchart of the proposed algorithm is as given in Figure 1. Median filtering technique is used initially to remove the noise in the CT image. The preprocessed image is then segmented to extract the lung region using global thresholding technique. Using the user defined input as the labels, each unlabeled pixel is assigned a vector of probabilities using Random walk algorithm. Combining these probability vector a label is assigned to each of the unlabeled pixel using Watershed algorithm for tumor region extraction.

2.1 Preprocessing and Lung Extraction

The CT images used are noisy with obscure edges. Median filtering technique is used to improve the segmentation of region of interest. The goal of median filtering is to remove the noise that has degraded the image. It is a statistical approach. It is the frequently used nonlinear operation to reduce "salt and pepper" noise[11]. The effectiveness of median filter is it simultaneously reduces noise by preserving edges.

Segmentation stage is to separate the objects and borders (lines, curves) in an image. Global Thresholding technique is used to segment and extract the lung region. Thresholding is a non linear operation which converts a gray scale image to binary image based on a specific threshold value. The pixels are assigned either 0 or 1 that are below or above the threshold value. Global image thresholding is done using Ostu's method[18]. The threshold selection is based on statistical criteria. Ostu suggested that an optimal threshold may be obtained by minimizing the weighted sum of the within class variances of the object and background pixels.

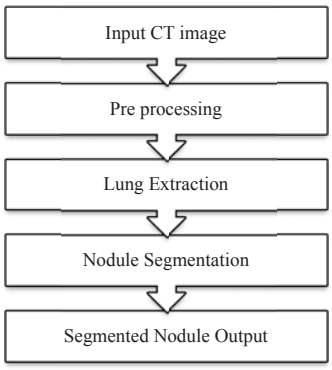

Figure 1. Flowchart of the proposed algorithm.

2.2 Nodule Segmentation

To ease the inconsistency for feature extraction, the essential step is to precisely define the lung nodules. Exactly defining the lung tumors is also crucial for optimal radiation oncology[6]. The following algorithm is used for segmenting the lung nodule.

Algorithm:

- Userdefined input seed points are given to mark the tumour and non tumour regions.
- With the user defined pixels as labels, the probability of reaching each labeled pixel from each unlabeled pixel is calculated using the Random Walker algorithm.
- For each unlabeled pixel a vector of probabilities is defined.
- The calculated vector of probabilities for each unlabeled pixel is combined and they can be assigned to one of the labels using the watershed algorithm to obtain tumor segmentation.

2.2.1 Overview of Random Walker Algorithm

In image segmentation, the relationship between random walks and Dirichlet problem is established in clustering the respective sub-regions according to the users' inputs[1]. The algorithm[19] computes the probability that each unlabeled pixel reaches the labeled pixels or seed points. A final segmentation is acquired by assigning the most probable seed point as label to each unlabelled pixel.

The input image may be viewed as a graph that consists of a pair G = (V,E) with vertices v ∈V representing pixels and edges e ∈ E ⊆ V × V representing the relationship between pixels. An edge e, that connects any two vertices, v_i and v_j, is denoted by e_{ij}. A weight can be assigned to each edge in the graph called as weighted graph. w(i,j) or w_{ij} represents the weight of the edge e_{ij}. The degree of a vertex is defined as the sum of weights of all edges incident on it denoted as $di = \sum w(e_{ij})$. Since w_{ij} is the the bias affecting the choice of labeling the pixel in Random walker's algorithm[20] we require that We also assume that our graph is undirected and connected.

Given a set of foreground seeds, F, and background seeds, B, where set of nodes $S=F \cup B$ and $F \cap B= \varphi$, the probability of a random walker, x_i starting at node v_i first reaches a seeded node, v_S is equivalent to the solution to the Dirichlet problem of finding the harmonic function based on its boundary values.

$$L_U X = -B^{\mathbf{T}} X_M \quad (1)$$

Where L_U, the unseeded node sin Laplacian, is one component to the decomposition of the combinatorial Laplacian matrix, Equation (2), is the boundary conditions at the locations of the seeded points, X_M.

$$L_{ij} = \begin{cases} d_i if i = j, \\ -w_{ij} if v_i and v_j are incident node \\ 0 \qquad otherwise \end{cases} \quad (2)$$

With a defined set of seeds X_U, the belongingness of an unlabelled node x_i to the seed v_S with label ,where $S=(F,B)$ can be identified when its probability, to reach v_S with label is higher.

$$v_i = s, \text{ since } pr_i = \max(s) \quad (3)$$

The Gaussian weighting function is represented by,

$$w_{ij} = \exp\left\{-\beta(g_i - g_j)^2\right\} \quad (4)$$

Where represents the intensity at pixel i. The value of β is treated as free parameter in this algorithm.

2.2.2 Modification in Weighting Function

In the novel Random walk algorithm, specified in Equation (4) parameter β is a free parameter defined by the user. The modified method uses distance between adjacent nodes as a parameter in the place of constant free parameter[6,14]. The weighting function in Equation (4) is written as

$$w_{ij} = \exp\left\{\frac{-(g_i - g_j)^2}{h_{ij}}\right\} \quad (5)$$

Where the parameter is the Euclidean distance between adjacent pixels i and j and setting. In the original method the calculation of probability is based on the gradient between the pixels only and not directly on their intensity. In order to boost the grouping of pixels[21,22] having homogenous intensity a likelihood of probability is added to each class. The improvements in the algorithm facilitate the use of local information[23].

2.2.3 Combining Probability Vector with Watershed Algorithm

The vector of probabilities can be combined into single value by taking the product of all probabilities in the vector so as to obtain the resultant image R. It can be denoted as

$$R = \prod_j x_u^j \quad (6)$$

In the regions having equal probability the resultant image will have maximum values. i.e., when an unlabeled pixel has equal probability to reach any labeled pixel the region will have maximum value. A local ridge formation can be observed around each labeled pixel in the resultant image R generated by Equation (6) as the probability of an unlabeled pixel to reach the labeled pixel decreases when it moves away from it. The labeled pixel region is expanded in all directions in the image R until their corresponding ridge locations are reached. This results in the expansion of seed points to region of interest. To end with, the resultant image R is inverted and a marker controlled watershed transform is performed on the image where the markers are the labeled pixel region. This results in tumor segmentation with more accurate distinction between the object and the boundaries. Figure 2 shows the original image. Figure 3 is the output of extracted lung region and Figure 4 shows the segmented nodule region.

Figure 2. CT image of lungs.

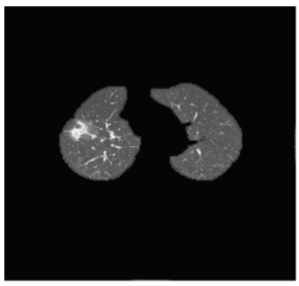

Figure 3. Image of Extracted lung region.

3. Experimental Results

We tested our method with 23 CT images. A sample of 5 images is shown in Figure 5. Nodules are segmented using Random Walk (RW), Random Walk with Improved Weight (RWIW) and Random Walk with Improved Weight combined with Watershed (RWIW-WS). Segmentation accuracy is calculated based on the boundary descriptors such as its area, major axis, minor axis, eccentricity and perimeter[24–26]. Results obtained from the tested CT images are given in Table 1.

Figure 4. Detected nodule region.

The DSC is calculated as

$$DSC = 2(M \cap A) / (M + A) \quad (7)$$

Where M is the manual segmentation of the nodule and A is the proposed method segmentation of the nodule.

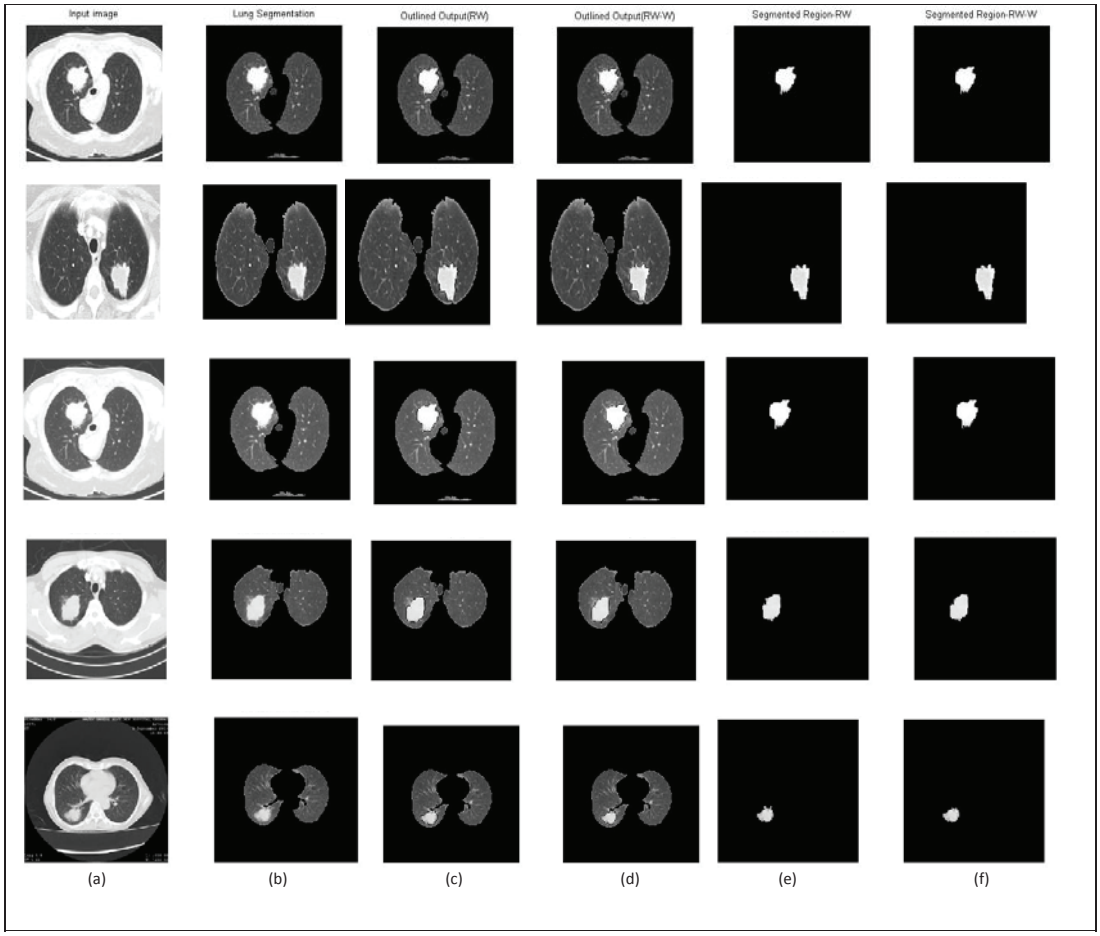

Figure 5. Example segmentation of 5 CT images out of 23 images taken for testing in the study. **(a)** Sample input images. **(b)** Extraction of lung region in processing stage. **(c)** Outlined output of nodule segmentation with improved weight in Random walker algorithm. **(d)** Outlined output of nodule segmentation with Random walker and watershed algorithm combined. **(e)** Segmented nodule with improved weight in Random walker algorithm. **(f)** Segmented nodule with Random walker and watershed algorithm combined..

Table 1. Comparison of calculated parameters using random walker, with improved weights and watershed algorithm

Image ID	Area			Major Axis			Minor Axis			Perimeter			Eccentricity			Equivalent Diameter		
	RW	RWIW	RWIW - WS	RW	RWIW	RWIW - WS	RW	RWIW	RWIW - WS	RW	RWIW	RWIW - WS	RW	RWIW	RWIW - WS	RW	RWIW	RWIW - WS
1	104	84	84	21.048	14.45	14.45	7.633	7.714	7.714	52.63	35.8	35.8	0.932	0.846	0.846	11.507	10.403	10.403
2	61	58	58	9.551	9.432	9.43	8.307	8.032	8.032	25.9	25.9	25.9	0.494	0.524	0.524	8.956	8.593	8.593
3	27	29	29	9.083	9.069	9.069	4.02	4.307	4.307	18.83	19.41	19.41	0.897	0.88	0.88	6.412	5.754	5.754
4	34	34	37	9.01	9.01	8.797	5.009	5.009	5.545	20.24	20.24	20.83	0.831	0.831	0.776	6.58	6.58	6.956
5	92	92	92	11.5	11.5	11.50	10.317	10.317	10.317	32.73	32.73	32.73	0.442	0.442	0.442	10.823	10.823	10.823
6	68	68	68	12.245	12.245	12.25	7.371	7.371	7.371	29.66	29.66	29.66	0.79	0.79	0.79	9.305	9.305	9.305
7	162	166	171	23.269	23.852	24.039	9.414	9.472	9.762	54.63	57.21	57.21	0.915	0.918	0.914	14.184	14.184	14.755
8	69	71	71	12.805	12.803	12.803	7.066	7.249	7.249	30.49	30.49	30.49	0.834	0.824	0.824	9.441	9.508	9.508
9	1115	1130	1118	48.049	48.134	47.911	30.058	30.351	30.186	140.43	138.43	139.01	0.78	0.777	0.777	37.678	37.931	37.729
10	168	168	168	17.793	17.793	17.793	12.306	12.306	12.306	49.31	49.31	49.31	0.722	0.722	0.722	14.625	14.625	14.625
11	233	230	228	24.301	24.022	23.836	12.872	12.848	12.857	62.63	62.04	61.46	0.848	0.845	0.842	17.224	17.113	17.038
12	56	59	56	9.349	10.078	9.348	8.063	7.905	8.063	27.07	28.49	27.07	0.506	0.62	0.506	8.444	8.667	8.444
13	36	36	35	11.566	11.566	11.006	5.005	5.005	4.175	23.83	23.83	22.83	0.945	0.945	0.925	6.936	6.936	6.676
14	352	352	353	28.532	28.532	29.782	15.021	15.021	16.069	82.73	82.73	83.94	0.892	0.892	0.842	21.098	21.098	21.2
15	97	97	99	12.902	12.902	13.932	8.275	8.275	9.265	33.53	33.53	35.9	0.877	0.877	0.747	10.367	10.367	11.227
16	1521	1521	1519	65.897	65.897	61.817	34.979	34.979	33.779	185.33	185.33	182.33	0.871	0.871	0.838	44.978	44.978	43.978
17	29	29	29	7.063	7.063	7.0629	5.406	5.406	5.406	17.66	17.66	17.66	0.644	0.644	0.644	6.077	6.077	6.077
18	1257	1257	1257	46.125	46.125	46.125	36.709	36.709	36.709	164.71	164.71	164.71	0.605	0.605	0.605	40.006	40.006	40.006
19	175	175	175	16.383	16.383	16.383	13.938	13.938	13.938	50.04	50.04	50.04	0.526	0.526	0.526	14.927	14.927	14.927
20	151	151	151	17.377	17.377	17.377	12.962	12.962	12.962	57.46	57.46	57.46	0.666	0.666	0.666	13.868	13.868	13.868
21	185	185	180	20.57	20.57	19.206	14.41	14.41	13.414	59.28	59.28	57.21	0.826	0.826	0.716	17.024	17.024	15.139
22	41	41	41	12.035	12.035	12.035	4.755	4.755	4.755	26.14	26.14	26.14	0.919	0.919	0.919	7.225	7.225	7.225
23	551	536	507	29.775	29.775	29.693	24.431	24.431	22.732	106.67	106.67	100.08	0.517	0.572	0.643	26.487	26.124	25.407

4. Conclusion

The proposed method is used to identify lung nodules from CT images with irregular boundaries with user defined seed points. The Random Walker Watershed algorithm proposed for tumor detection works on the basis of combining the probabilities of each unlabeled pixel obtained using Random Walks to get a resultant image which is then segmented with Watershed algorithm. The advantage is that any pixel in the region of interest can be taken as seed points and can accurately segment and describe the region of interest precisely. The Random walker algorithm is improved since the weight fuction is not only the gradient of the pixel but also the Euclidean distance between adjacent pixels. Results show that the proposed method improves the accuracy of segmenting nodules with dirchlet boundaries.

5. References

1. Sluimer I, Schilham A, Prokop M, Ginneken BV. Computer analysis of computer tomography scans of the lung: A survey. IEEE Transactions-on-Medical Imaging. 2006; 25(4):385–405.
2. Wang N, Huang LL, Zhang B. A fast hybrid method for interactive liver segmentation. IEEE 2010 Chinese Conference on Pattern Recognition (CCPR); chongging. 2010. p. 1–5.
3. Qiu G, Yuen PC. Interactive imaging and vision-Ideas, algorithms and applications. Pattern Recognition. 2010; 43(2):431–3.
4. Grady L. Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006; 28(11):1768–83.
5. Ram S, Rodriguez JJ. Random walker watersheds: A new image segmentation approach. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Vancouver, BC. 2013. p. 1473–7.
6. Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. Computers in Biology and Medicine. 2014; 50(25):76–96.
7. El-Baz A, Beache GM, Gimel'farb G, Suzuki K, Okada K, Elnakib A, Soliman A, Abdollahi B. Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. International Journal of Biomedical Imaging. 2013 Jan; 29:46
8. Day E, Betler J, Parda D, Reitz B, Kirichenko A, Mohammadi S, Miften M. A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients. Medical Physics. 2009; 36(10):4349–58.
9. Bagci U, Yao J, Caban J, Turkbey E, Aras O, Mollura D. A graph-theoretic approach for segmentation of PET images. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC; Boston, MA. 2011. p. 8479–82.
10. Kuhnigk JM, Dicken V, Bornemann L. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. IEEE Transactions on Medical Imaging. 2006; 25(4):417–34.
11. Dehmeshki J, Amin H, Valdivieso M, Ye X. Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach. IEEE Transactions on Medical Imaging. 2008; 27(4):467–80.
12. Diciotti S, Picozzi G, Falchini M, Mascalchi M, Villari N, Valli G. 3-D segmentation algorithm of small lung nodules in spiral CT images. IEEE Transactions on Information Technology in Biomedicine. 2008; 12(1):7–19.
13. Goodman LR, Gulsun ML, Washington L, Nagy PG, PiacsekK L. Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. American Journal of Roentgenology. 2006; 186(4):989–94.
14. Helen R, Kamaraj N, Selvi K, Raman VR. Segmentation of pulmonary parenchyma in CT lung images based on 2D Otsu optimized by PSO. 2011 International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT); Tamil Nadu. 2011. p. 536–41.
15. Onomaa DP, Ruan SS, Thureau S, Nkhalia L, Modzelewskia R, Monnehan GA, Vera P, Gardin I. Segmentation of heterogeneous or small FDG PET positive tissue based on a 3D-locally adaptive random walk algorithm. Computerized Medical Imaging and Graphics. 2014; 38(8):753–63.
16. Boykov YY, Jolly MP. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. Proceedings IEEE International Conference on Computer Vision, ICCV'01; Vancouver, BC. 2001. p. 105–12.
17. Yin LK, Rajeswari M. Random walker with improved weighting function for interactive medical image segmentation. Bio-Medical Materials and Engineering. 2014; 24(6):3333–41.
18. Wang N, Huang LL, Zhang B. A fast hybrid method for interactive liver segmentation. 2010 Chinese Conference on Pattern Recognition (CCPR); chongging. 2013. p. 1–5.
19. Adams R, Bischof L. Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1994; 16(6):641–7.

20. Hu S, Xu C, Guan W, Tang Y, Liu Y. Texture feature extraction based on wavelet transform and gray-level cooccurrence matrices applied to osteosarcoma diagnosis. Bio-Medical Materials and Engineering. 2014; 24(1):129–43.
21. Clausi DA. An analysis of co-occurrence texture statistics as a function of grey level quantization. Canadian Journal of Remote Sensing. 2002; 28(1):45–62.
22. De Martinao M, Causa F, Serpico SB. Classification of optical high resolution images in urban environment using spectral and textural information. Proceedings of 2003 IEEE International Geoscience and Remote Sensing Symposium, IGARSS'03; 2003; 467–9.
23. Dharmarajan A, Velmurugan T. Lung cancer data analysis by k-means and farthest first clustering algorithms. Indian Journal of Science and Technology. 2015 Jul; 8(15). DOI: 10.17485/ijst/2015/v8i15/73329.
24. Indira KP, Hemamalini RR. Evaluation of choose max and contrast based fusion rule using DWT for PET, CT images. Indian Journal of Science and Technology. 2015 Jul; 8(16).
25. Theresa MM, SubbiahBharathi V. CAD for lung nodule detection in chest radiography using complex wavelet transform and shearlet transform features. Indian Journal of Science and Technology. 2016 Jan; 9(1). DOI: 10.17485/ijst/2016/v9i1/75243.
26. Min-CheolJeon, Han M-S, Lim H-S, Jang J-U, Lee S, Kim Y-K, Lee I-H. The guidelines of laser on lung biopsy under CT guidance. Indian Journal of Science and Technology. 2015 Jan; 8(S1). DOI: 10.17485/ijst%2F2015%2Fv8iS1%2F59437.