ORIGINAL ARTICLE

Segmentation of lung nodule in CT data using active contour model and Fuzzy C-mean clustering

Ezhil E. Nithila*, S.S. Kumar

Noorul Islam Centre for Higher Education, Kumaracoil 629180, India

Received 24 November 2015; revised 25 March 2016; accepted 4 June 2016
Available online 14 June 2016

1. Introduction

Of the other cancers (colon, breast and prostate cancer), the most serious death causing cancer is Lung cancer. Recently, the United States has diagnosed 2,28,190 peoples with lung cancer [1]. Most of the people were diagnosed with lung cancer having non-small cell lung cancer rather small cell lung cancer. Early diagnosis and diagnosis are important for the increase in survival rate up to 50% [2].

Nodule identification is one of the fundamental problems in medical image processing. Pulmonary nodules are small tissue in the lung and most of them are benign [3]. To assist the specialist, with the information regarding nodules in lung parenchyma computer-aided detection (CAD) system was developed. An automated CAD system was designed to detect the nodule sized between 3 and 30 mm. CAD has more advantages in terms of speed, accuracy [4], detection of nodules in pulmonary CT images and a reduction in miss rate [5,6]. Segmentation is an essential step in many applications involving CAD. It partitions the image into segments corresponding to the anatomical objects in the image. In the recent years, a lot of pulmonary nodule segmentation methods have been proposed, which can be categorized as thresholding method [7,8], morphological method [9], deformable model [10], clustering method [11–13], graph cut method [14,15], Markov random field, region growing [16], watershed, neural networks, fuzzy logic, active contours [17] and histogram based segmentation [18].

Among various segmentation methods, active contour was one of the most popular and successful one. Many researchers are devoted to study the detection of pulmonary nodules...
attached to vessels and the pulmonary wall. El-Bazl et al. [19] proposed a segmentation technique using Markov random field, consisting of two stages. The first stage is to select the optimum decision level to create an initial labeling image, and the second one is to extract the lung tissues from each slice. Using this methodology, about 50 subjects were evaluated. Out of this 40 subjects were normal and ten subjects had abnormalities in their CT scans.

Tong et al. [20] used a three step process to detect lung nodules. Firstly, an adaptive threshold algorithm was used to segment the lung region. Secondly, active contour model (ACM) was used to remove lung vessel and finally a Hessian matrix (selective shape filter) was used to detect the suspicious nodules. This method was able to produce an overall detection rate of 85%. Marten et al. [21] evaluated and compared features such as nodule size, position, margin, matrix characteristics, vascular and pleural attachments with gold standard. Some authors use manually segmented lesion as the gold standard and some other uses specialist references as the gold standard. Azimifar et al. [22] used active contour modeling for segmentation and produces an overall detection rate of 89%. Dehmeshki et al. [23] proposed volumetric measurement for the detection of lung nodule. A new region growing method for segmentation in combination of fuzzy connectivity, distance and intensity information as the growing mechanism and peripheral contrast as the halting criterion has been used. It was found that this method is highly reproducible for various types of nodules from various data protocols. All these works aim to detect lung nodule automatically from CT images.

Most of them produce segmentation results with less accuracy which is undesirable for clinical usage. Hence it is necessary to develop a suitable technique which segments lung nodules with good accuracy, reduced computation time and less segmentation error rate per image.

The present work aims to segment the lesion automatically with the help of CT images. For lung reconstruction, Selective Binary and Gaussian Filtering with new signed pressure force function (SBGF-new SPF) and segmentation of lung nodule, 3-class FCM was used.

2. Methodology

Automatic nodule detecting scheme using CT scans helps the physicians to reduce the load and to improve detection quality. The methods proposed for the detection of lung nodule consist of the CT lung acquisition and the segmentation of lung nodules. The main aim of this process was to remove the portions that are part of the CT image other than lung lesion. Fig. 1 shows the various stages of segmentation scheme.

![Figure 1 Main diagram of segmentation scheme.](image-url)

2.1. CT lung acquisition

The Lung Image Database Consortium image collection (LIDC-IDRI) consists of diagnosing and lung cancer screening thoracic CT scans with marked-up annotated lesions [24]. It is a web-accessible resource for development, training, and evaluation of CAD methods for lung cancer detection and diagnosis.

2.2. Reconstruction of lung parenchyma

The objective of this stage was to eliminate the mediastinum and thoracic wall and to separate the parenchyma using region based ACM. There exist many region based models such as Geodesic active contour (GAC) model [25], Chan–Vese (C–V) model [26], Piecewise Smooth (PS) model [27], Local Binary Fitting (LBF) [28] model and so on. Some models suffer due to weak edges and others due to evolutionary time. But SBGF-new SPF performs well under weak edges, with less computational time and increased efficiency.

2.2.1. The proposed method

Based on the work inspired by [29] the proposed method segments the lung nodule and has the advantage of easy implementation, speed, less segmentation error and accuracy.

In order to increase the efficiency, the algorithm was modified by the inclusion of contour constants in SPF function. Let $\Omega \in R^3$ be the image domain, and $I(x)$ be the input image. Let $c_1$, $c_2$ be the constants inside and outside the contour. The proposed method uses selective step and no re-Initialization is required. It first penalizes level set function to be binary, and then uses a Gaussian filter to regularize it. The Gaussian filter can make the level set function smooth and the evolution more stable. The balloon force $\alpha$ is used for shrinking or expanding the contour. The Signed Pressure Force (SPF) function is used to control the direction of evolution. The function ranges in $\{-1, 1\}$. It uses local image information. It modulates the signs of the pressure force which shrinks of contour inside the object and expands of contour outside the object. A Gaussian kernel is used to regularize the level set function which not only regularizes it but also removes the need of computationally expensive re-initialization. If the object boundary has gaps, the contour tends to leak and hence boundary leakage problem arises. This can be solved by applying an edge stopping function, which pulls the contour to the boundary [30]. The existing methodology uses more iteration. The proposed methodology does not change the functionality of evolution but reduces the number of iterations.

Therefore, the energy function of the proposed system was given by

$$\frac{\partial \phi}{\partial t} = c_1 \ast c_2 \ast \text{spf}(I(x)) \ast z|\nabla \phi|, \ x \in \Omega$$

$$\text{spf}(I(x)) = \frac{I(x) - \frac{\max}{\max (|I(x) - \frac{\max}{\max})}}{\max (|I(x) - \frac{\max}{\max})}, \ x \in \Omega$$

$$c_1(\phi) = \frac{\int_\Omega I(x) \ast H(\phi) dx}{\int_\Omega H(\phi) dx}$$

$$c_2(\phi) = \frac{\int_\Omega I(x) \ast (1 - H(\phi)) dx}{\int_\Omega (1 - H(\phi)) dx}$$
Using Eq. (1) the energy function can be formulated. The parameters such as sigma (σ), number of iterations, delta and alpha (α) can be varied to obtain the desired lung parenchyma. Decreasing σ results in incomplete lung lobe. On the other hand increasing σ results in large time consumption.

2.2.2. Reconstruction of lung border

The reconstruction of the lung border is an important step that aims to recover lung nodules that are attached to the thoracic wall. The juxtapleural nodule can be recovered using Adaptive border marching [31], greedy snake algorithm [32], ray casting, vector quantization [33], and chain codes [34]. The present work employs a technique called rolling ball, which uses morphological closing operations with a circular structuring element along the contour of the lung, causing the reconstruction of the concavities where this element cannot enter. The structuring element was used in a disk with radius equal to 12.

2.3. FCM

Clustering algorithms achieve region segmentation by partitioning the image into sets or clusters of pixels that have strong similarity in the feature space. In hard clustering, data are divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster [35,36].

Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters. One of the most widely used fuzzy clustering algorithms is the FCM Algorithm [37]. This can be used in many fields such as pattern recognition, fuzzy identification, and feature extraction.

Let \( U \in M_c \) be a fuzzy \( c \) partition of \( X \), and the FCM function [42] is defined as

\[
J_m(U, v) = \sum_{k=1}^{n} \left( \sum_{i=1}^{c} (u_{ik})^m d_{ik} \right)^2
\]

\[
d_{ik} = \| x_k - v_i \|^2
\]

The FCM algorithm produces a fuzzy \( c \) partition of the data set \( X = \{x_1, x_2, \ldots, x_n\} \). The 3-class FCM algorithm is used to segment the region of interest from the reconstructed lung. The region of interest (ROI) contains nodules, blood vessels and bronchi. In order to separate nodule from these structures, morphological operations are performed.

3. Results and discussion

The proposed system uses LIDC for the evaluation on lung CT. For benignity the nodule size is 3–30 mm and > 30 mm for malignancy. The original lung CT, reconstructed lung parenchyma, FCM output and segmented nodule detection output for different candidates are shown in Fig. 2. It has been found that the segmented output can be achieved at faster rate with less number of iterations. The number of iterations required is 140 with a response time of less than 1 min. The parameters are used in the proposed algorithm for separating the parenchyma from the mediastinum and the thoracic wall and nodule detection are \( \sigma \) equal to 0.3, 1 and 2, number of iterations to segment is equal to 150, delta equal to 1, \( \alpha \) equal to 4 and number of clusters is equal to 3. It was noted that if the size of parenchyma is large, the number of iterations can changed to higher value. No other parameter needs to be changed.

3.1. Quantitative metrics for evaluation

The images that are segmented manually are termed as Gold Standard (GS). S denotes the automated segmented image. Let \( e \) be the error measure and \( n \) be the number of samples used. Evaluation of segmented (S) against GS images was made for following measures:

a) Volume error, \( VE = \frac{2(S - GS)}{(S + GS)} \) (6)

b) Coefficient of Similarity = \( 1 + \frac{(GS \cap S)}{GS} \) (7)

c) Spatial overlap = \( \frac{2(GS \cap S)}{(S \cup GS)} \) (8)

d) RMSE = \( \sqrt{\frac{1}{n} \sum e_i^2} \) (9)

e) Under segmentation rate, \( U = |GS - (S \cap GS)| \) (10)

f) Over segmentation rate, \( V = |S - (S \cap GS)| \) (11)

The average volume error is 0.968%. For clinical usages, less than 5% volume error is more likely suitable [38]. Figs. 3 and 4, show the comparison of coefficient of similarity and spatial overlap between LBF and proposed method. Coefficient of similarity shows the resemblances between the segmented and gold standard images. Spatial overlap is the accurate measure of spatial properties of segmented images. The best coefficient of similarity and overlap fraction is 0.914 and 0.584 and worst is 0.074 and 0.089. According to [41] the spatial overlap of LIDC database can be reported from 0.51 to 0.66. The proposed work shows a spatial overlap fraction of 0.584 which lies within the specified range. Experimental results show that the proposed method shows increased performance when compared with LBF model. In the previous literature Paik et al. [31] suggested an average over segmentation and under segmentation ratio was 0.43% and 1.63%. In the present study, the observed results (average over segmentation and under segmentation ratio) were 0.63% and 0.015% respectively. It was found that the proposed methodology shows better performance when compared with the previous literature results [31].

Table 1 depicts the comparison of various segmentation methods with the proposed method. The root mean square error (RMSE) and overlap measure describe the similarities between segmented and gold standard image. The proposed method achieves RMSE of 0.10 mm and accuracy measure of 98.95%. In research studies the performance of RMSE has been used as a standard statistical metric. The proposed method shows very less error rate when compared with other methods. Under segmentation rate defines the proportion of the unsegmented lesion area. Over segmentation rate is defined as the ratio of the segmented non-lesion area. Figs. 5 and 6...
show the over segmentation and under segmentation ratio for LBF and proposed model. The average over segmentation and under segmentation ratio was 0.63% and 0.015% respectively.

3.2. Advantage of proposed model over the LBF model

Compared with LBF, proposed model utilizes global segmentation with less number of iterations and low computation time. LBF uses local segmentation, with more iteration and very high computation time. LBF is a region based active contour model which uses signed distance function (SDF) and requires re-initialization. Evolution converges at 100 iterations with the required time of 454 s whereas the proposed model utilizes 140 iterations for 18 s. The best similarity and overlap fraction are 0.914 and 0.584 and worst is 0.074 and 0.089. The average over segmentation and under segmentation ratio was 0.63% and 0.015% respectively.
4. Conclusion

Computer-aided diagnosis of lung cancer is used to segment pathologically changed tissues fast and accurately. The proposed algorithm SBGF-new SPF and FCM successfully segment the lung nodule from the CT. For lesion the average volume error obtained is 0.968%. The coefficients of similarity, spatial overlap, RMSE, average over and under segmentation ratio are 0.914, 0.584, 0.10 mm, 0.63% and 0.015% respectively. The experimental results indicate that the newly proposed algorithm could segment blood vessel adhesion, pleura adhesion fast and exactly performs better than traditional segmentation effects, with executive efficiencies and decreased rate of errors.

Acknowledgment

We express our sincere thanks to DST, New Delhi for the financial support under INSPIRE fellowship scheme for the above research work.

References


![Table 1 Performance comparison with other models.](image)

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hu et al., [39]</td>
<td>Iterative threshold, Morphological Operations</td>
<td>Root mean square Difference = 0.54 mm</td>
</tr>
<tr>
<td>Yim et al., [40]</td>
<td>Region growing, Connected component</td>
<td>Root mean square Difference = 0.32 mm</td>
</tr>
<tr>
<td>Badura et al. [41]</td>
<td>Modified relative Fuzzy connectedness</td>
<td>Spatial Overlap = 0.6 Accuracy = 98%</td>
</tr>
<tr>
<td>Deep et al. [42]</td>
<td>Rotation invariant LBP</td>
<td>Root mean square Difference = 0.10 mm</td>
</tr>
<tr>
<td>Proposed method</td>
<td>SBGF-new SPF, FCM</td>
<td>Spatial Overlap = 0.584 Accuracy = 98.95%</td>
</tr>
</tbody>
</table>

![Figure 5 Under segmentation rate of various images in databases.](image)

![Figure 6 Over segmentation rate of various images.](image)


[16] Li Wang, Lei He, Arabinda Mishra, Chunming Li, Active contours driven by local Gaussian distribution fitting energy, Signal Processing 89 (2009) 2435–2447.


