Modified approaches on Lung Cancer Cell Extraction and Classification fromComputerized Tomography Images

ANTONY JUDICE A*[1], K. Parimala Geetha[2], R. Krishnan Thampi[3]

[1]Research Scholar, ANNA UNIVERSITY, CHENNAI, INDIA
[2]Professor and Head, Department of ECE, PJCE, INDIA
[3]Dean, PJCE
pravinhireling@gmail.com, geetham_1969@yahoo.co.in, krishnanthampir@gmail.com

Abstract: Lung Tumor Delineation is a critical aspect in radiotherapy treatment for cancer. It is usually performed with the anatomical images of a Computerized Tomography(CT) scan. An image processing techniques and Computer Aided Diagnosis systems has demonstrated to be an effectual system for an improvement of Radiologists, Diagnosis, especially in the case of Medical Image Processing. In this paper, we present an Automatic Computer-Aided Diagnosis system for an Early Detection of Lung Cancer by an analyzing chest Computed Tomography (CT) images**.** One of the main difficulties limiting the segmentation of Lung Tumors by CT images is the noise due to the patient's Respiratory movements.A detection of the Lung Cancer in its early stage can be helpful for medical treatment to limit the danger. Most traditional medical diagnosis systems are founded on huge quantity of training data and takes long processing time. So for reducing these problems the Hidden Markov Model is proposed. This method will increase the diagnosis confidence and also reduce the time utility. Histological data are used only for validation and comparison of segmentation technique.
[ANTONY JUDICE A, K. Parimala Geetha, R. Krishnan Thampi. **Modified approaches on Lung Cancer Cell Extraction and Classification from Computerized Tomography Images.** *Life Sci J* 2013;10(2):1621-1626]. (ISSN:1097-8135). http://www.lifesciencesite.com. 224

Keywords: HMM, CT, Segmentation, Feature Extraction, Image classification

1. INTRODUCTION

Lung cancer is the most common cause of death due to cancer in both men and women through-out the world. The American Cancer Society estimates that 219,440 new cases of lung cancer in the U.S. will be diagnosed and 159,390 deaths due to lung cancer will occur in 2009[1] [2]. Lung cancer was initially thought to be infrequent in India. Lung cancer constituted 14.4% of all cancers in a review of 9210 consecutive autopsies by Banker17.The National Cancer Registry Programme of the Indian Council of Medical Research, which collected data from six different parts of the country, both rural and urban areas, showed varying figures in different areas 22. While cancer of the trachea, bronchus and lungs was the most common form of malignancy in males in 1989 from Bombay, Delhi, and Bhopal. India males, the same is 16.1 to 33.3 - 0 to 3.7 in females).

Since lung cancer tends to spread or metastasize very early after it forms, it is a very life-threatening cancer and one of the most difficult cancers to treat. While lung cancer can spread to any organ in the body. Something very difficult to interpret and very time consuming to radiologists, which may cause high false-negative rates for detecting small lung nodules, and thus potentially miss a cancer[3][7]. The first step for lung lobe segmentation is lung segmentation, in this method lung regions are extracted by using threshold and left and right lungs are segmented using dynamic programming [4].The pre requisite for the Fissure Detection to segment the lung lobes is Lung segmentation[12] in this paper segmentation by region glowing is used. Lung lobes are segmented in clinical CT image using sweeping for finding the fissure region and using DWT to enhance the fissure [5][8].

Some of the CAD system uses Fuzzy Logic, Neural network algorithm. Their disadvantages are time consumption and needed a lot of data for training. So the Hidden Markov Model is introduced for getting more advantage. A conventional method for summarizing sequential data is to use Hidden Markov Models (HMMs), which are a generalization of mixture models. In addition to several mixture components, they include a state variable that indicates the mixture component active at the current time step and a set of transition probabilities that indicates the likelihood that each component will become active given the previously active component. there exists an expectation-maximization (EM) algorithm known as the Baum-Welch algorithm, for finding maximum likelihood estimates of the parameter values of an HMM. This algorithm is easy to understand and easy to implement.

2. PROPOSED METHOD

2.1 CT SCAN:

A computerized axial tomography scan is an X-ray procedure that combines many x-ray images with the aid of a computer to generate cross-sectional views and three-dimensional images of the internal organs and structures of the body. Computerized axial tomography is more commonly known by its abbreviated names, CT scan or CAT scan.

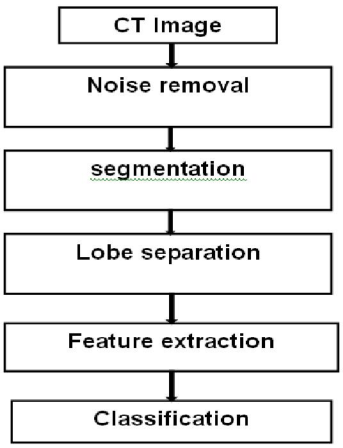

Figure 1:Proposed system

CT scan is used to define normal and abnormal structures in the body and/or assist in procedures by helping to accurately guide the placement of instruments or treatments [6].

A large donut-shaped x-ray machine takes x-ray images at many different angles around the body. These images are processed by a computer to produce cross-sectional pictures of the body. In each of these pictures the body is seen as an x-ray "slice" of the body, which is recorded on a film. This recorded image is called a tomogram. "Computerized Axial Tomography" refers to the recorded tomogram sections at different levels of the body.

Imagine the body as a loaf of bread and you are looking at one end of the loaf. As you remove each slice of bread, you can see the entire surface of that slice from the crust to the center. The body is seen on CT scan slices in a similar fashion from the skin to the central part of the body being examined. When these levels are further added together, a three-dimensional picture of an organ or abnormal body structure can be obtained.

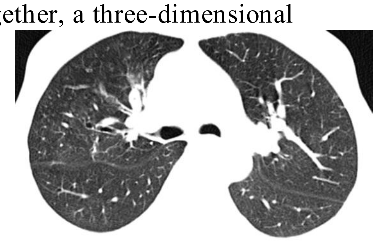

Figure 2: CT Image

2.2 NOISE REMOVEL

The most important technique for removal of blur in images due to linear motion or unfocussed optics and also due to vibrations. From a signal processing standpoint, blurring due to linear motion in a photograph is the result of poor sampling. Each pixel in a digital representation of the photograph should represent the intensity of a single stationary point in front of the camera. Unfortunately, if the shutter speed is too slow and the camera is in motion, a given pixel will be an amalgram of intensities from points along the line of the camera's motion. This is a two-dimensional analogy to

$$G(u,v)=F(u,v).H(u,v)$$

where F is the Fourier transform of an "ideal" version of a given image, and H is the blurring function.

In the real world, however, there are two problems with this method. First, H is not known precisely. Engineers can guess at the blurring function for a given circumstance, but determination of a good blurring function requires lots of trial and error. Second, inverse filtering fails in some circumstances because the sinc function goes to 0 at some values of x and y. Real pictures contain noise which becomes amplified to the point of destroying all attempts at reconstruction of a Fest.

The best method to solve the second problem is to use Wiener filtering. This tool solves an estimate for F according to the following equation:

$$\mathrm{Fest}(u,v=|H(u,v)|^2.G(u,v)/(|H(u,v)|^2.H(u,v)+K(u,v))$$

K is a constant chosen to optimize the estimate. This equation is derived from a least squares method.

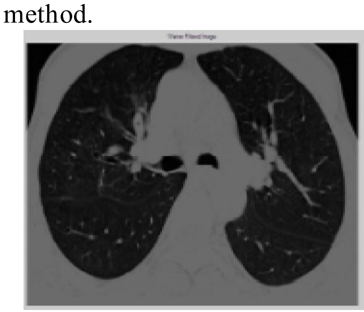

Figure 3: Filtered Image

In closing, it should be noted that Weiner filters are far and away the most common deblurring technique used because it mathematically returns the best results. Inverse filters are interesting as a textbook starting point because of their simplicity, but in practice Wiener filters are much more common. It should also be re-emphasized that Wiener filtering is in fact the underlying premise for restoration of other kinds of blur.

2.3 SEGMENTATION

Techniques employed are thresholding, Region Growing, Edge Detection, Ridge Detection, Morphological Operations, fitting of geometrical models or functions and dynamic programming. On the other hand, there is another approach used in lung regions extraction process based on pixel classifications, where each pixel in the CT image is classified into an anatomical class (usually lung or background, but in some cases more classes such as heart, mediastinum, and diaphragm). Classifiers are various types of neural networks, or markov random field modeling, trained with a variety of local features including intensity, location, and texture measures [7]. CADs can be divided into two groups [8]: density-based and model-based approaches.

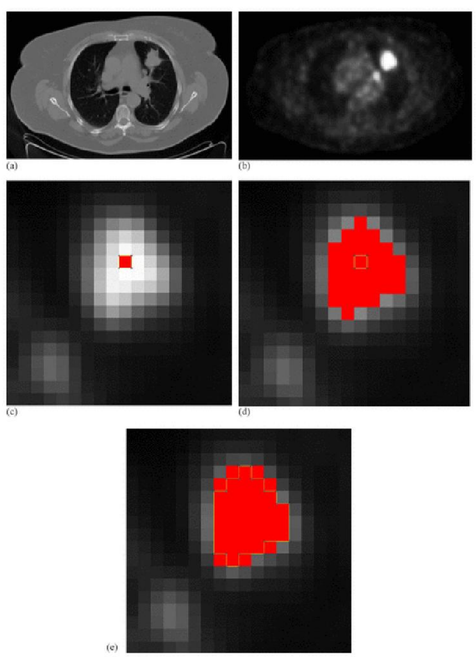

Figure 4: Segmentation on CT image

Considering the fact that lung nodules have relatively higher densities than those of lung parenchyma, density-based detection methods employ techniques such as multiple thresholding, region-growing, locally adaptive thresholding in combination with region growing, opening and closing, using the histogram, the top 20% gray values considered as initial cancerous candidate regions, using the histogram the normal tissues are removed, then elliptical-shaped regions, which is in general represent abnormalities, are detected, and fuzzy clustering used to identify nodule candidates in the lungs. False-positive results can then be reduced from the detected nodule candidates by employing a priori knowledge of small lung nodules. After getting the segmentation results, different features should be extracted to be used in the diagnosis phase where sets of rules are formulated to distinguish between true and false cancerous candidates.

Figure 5: Segmented Lung image

In some approaches uniformity, connectivity, and position features. In the features such as size, circularity, and mean brightness of region of interests (ROIs) were extracted. Area, thickness, we are using region glowing technique When viewed on transverse CT slices, the anterior and posterior junctions between the left and right lungs may be very thin with weak contrast.

2.4 LOBE SEPERATION

The goal of the lobe separation step is to locate these junction lines and completely separate the right and left Lungs. Using a technique similar to that employed in dynamic programming is applied to find the maximum cost path through a graph with weights proportional to pixel gray-level. The maximum cost path corresponds to the junction line position. However, we use a different strategy to find the dynamic programming search regions. In our method, a search region is found on a 2-D slice to successive slices. Because of the smooth pulmonary anatomy, the junction line position varies slowly through the data set.

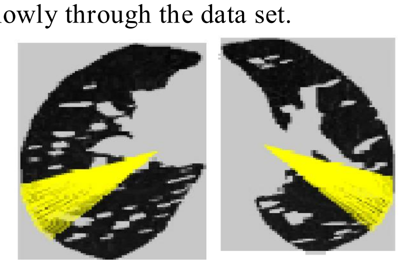

Figure 6: Separated Lung Lobes

To further reduce computation time, we only apply the lung separation step to those slices that contain a single, large, connected lung component. A conditional dilation is then used to restore the approximate original boundary shape, without reconnecting the two lungs again.

The fissure search technique conducts

pixel-by-pixel analysis and automatically places anchor points at a distance of 5 pixels apart along the identified fissures. Following the fissure search, the algorithmuses a fissure verify technique, which validates the correctness of a current fissure by comparing its anchor points with their counterparts on a previous adjacent fissure.[12]

The Lobe segmentation algorithm works well for the automatic detection of the fissure locations and curvatures for bothleft and right oblique fissures[12] Horizontal fissure presents moreof a challenge, because of the large variations in its pattern. Radiologists and surgeons also have difficulties in identifying the horizontal fissure inisotropic CT images. Zhang et al. [9] and Zhou et al. [10] segmented the horizontal fissure, but provided no specified segmentation accuracy for it.

2.5 FEATURE EXTRACTION

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately.

Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples.

3. PROPOSED CLASSFICATION ALGORITHM

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network.

HMMs are composed of states, which are traversed according to transition probabilities. The sequence data is viewed as a series of observations emitted by the states, where an emission distribution over observations is associated with each state. Formally, an HMM is characterized by three stochastic matrices, called the initial, transition and observation matrices.

The transition matrix, A, is a square matrix that holds the probabilities of transitioning from each state to any other.

The probability of transitioning from state i to state j is denoted by $\mathbf{a}_{ij}$. The initial distribution vector, π, is a column vector that stores the probabilities of starting in each state at the beginning of the sequence. π_i denotes the probability of starting in state i. Finally, the observation matrix, B, defines the probabilities of observing each base pair for every state. The probability of observing observation k in state j is denoted by $b_j(k)$.

In a regular Markov model, the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a Hidden Markov model, the state is not directly visible, but output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states[14][15]. Note that the adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model; even if the model parameters are known exactly, the model is still 'hidden'.

Most selected Training algorithm used is the Baum-Welch re-estimation Formulas. The Hidden Markov Model is a finite set of states, each of which is associated with a (generally multidimensional) probability distribution. So according to the training of HMM with regarding the features it will report the accrue out more effectively[16]. An artificial neural network is a structure which will attempt to find a relationship i.e. a function between the inputs, and the provided output(s), in order that when the net be provided with unseen inputs, and according with the recorded internal data (named "weights"), will try to find a correct answer for the new inputs.

The main difference could be this: In order to use a markov chain, the process must depend only in it´s last state. For use a neural network, you need a lot of past data.

The Baum-Welch algorithm can be used to train an HMM to model a set of sequence data. The algorithm starts with an initial model and iteratively updates it until convergence. The algorithm is guaranteed to converge to an HMM that locally maximizes the likelihood (the probability of the training data given the model).

Since the Baum-Welch algorithm is a local iterative method, the resulting HMM and the number of required iterations depend heavily on the initial model. Of the many ways to generate an initial model, some techniques consider the training data while others do not.

4. CONCLUSIONAND RESULT

CAD applications will have high sensitivity but also have high specificity to avoid cost-intensive, inconvenient, or even harmful follow-up procedures to rule out misclassified lung lesions. Fissures cannot be seen by naked eyes because CT images are taken by low resolution scanners and as the number of image increases, efforts taken by the medical experts to analyze the image consumes much time. Because of this reason, there is a proposal for developing an Computer automated Diagnosis system to assist the medical experts. This proposed reducing the consumption of time and increasing diagnosis confidence to the patient.

Computer Tomography Image is normally taken for detection of lung cancer because CT scan is able to detect the small nodule in lung. CT is especially effective for diagnosing lung cancer at its earlier stage, which is shown in Figure 7.

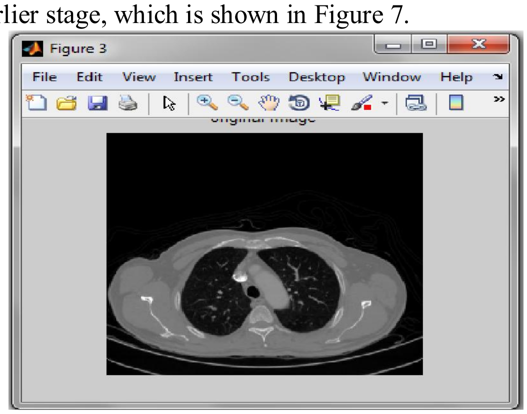

Figure 7 Original CT image

Preprocessing Stage

The preprocessing is very important in isotropic CT Lung images, since isotropic CT images contain more noise than their clinical counterparts, that can be removed by Wiener filter which gives effective results.

To perform the wiener filter, noise reduction can be done through image blurring, blurring is removed by Gaussian smoothening model. Figure 8represents the noise removal stage using wiener filter.

Figure 8 Filtered Image

{(a)=Original image, (b)=Blurred image,(c) =Denoised image, (d)=Wiener filter output}

The Figure 9 illustrate the noise reduced image

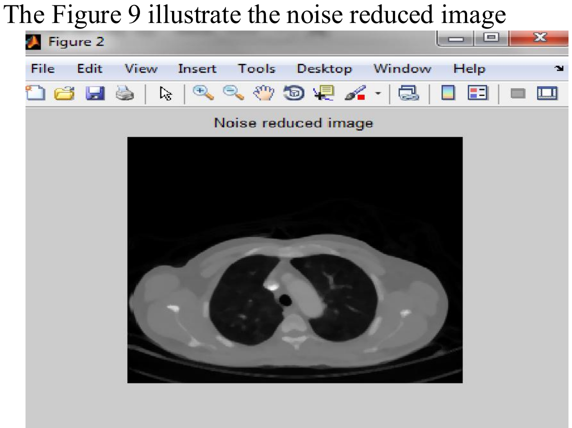

Figure 9 Noise reduced image

After noise reduction the histogram of the CT image is found out. The histogram of the input CT image is shown in the Figure 10

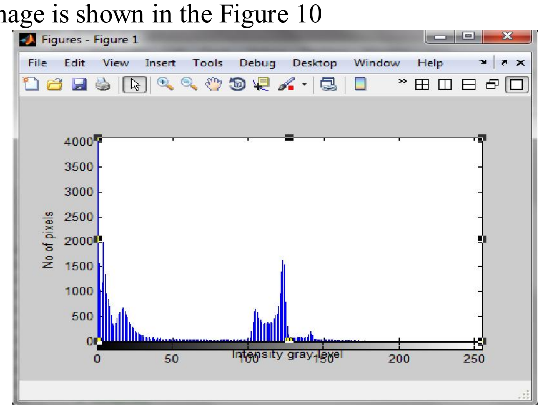

Figure 10 Histogram of the input CT image

Segmentation Stage

Segmentation divides an image into its constituent regions or objects. Edge detection is shown in the Figure 10(a). After edge detection, the lung region are cropped that is illustrated in Figure 10(b)

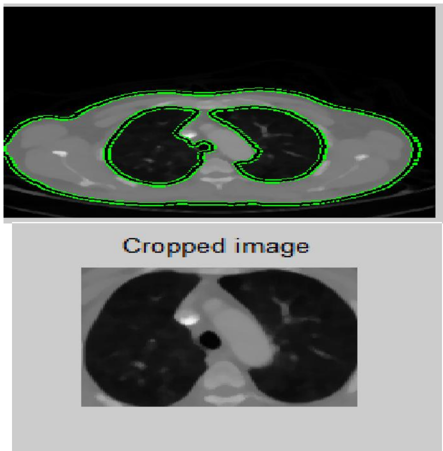

Figure 10(a)Edge Detection (b) Cropped Image

The Figure 11 illustrates the global region based segmentation.

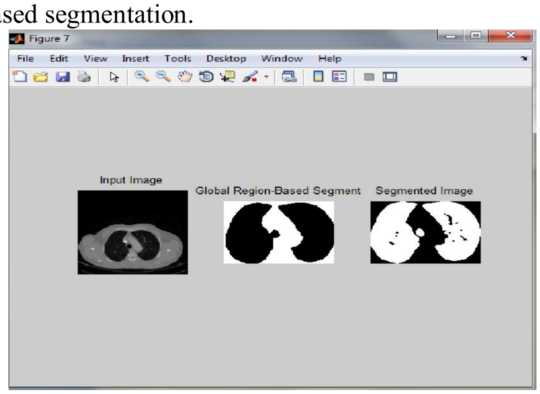

Figure 11. Segmentation stage Lobe Separation

The lobe separation is carried out by setting suitable thresholding values,so that left and right lung are separated.

Figure 12 illustrate the left and right lung with threshold value 0.0555

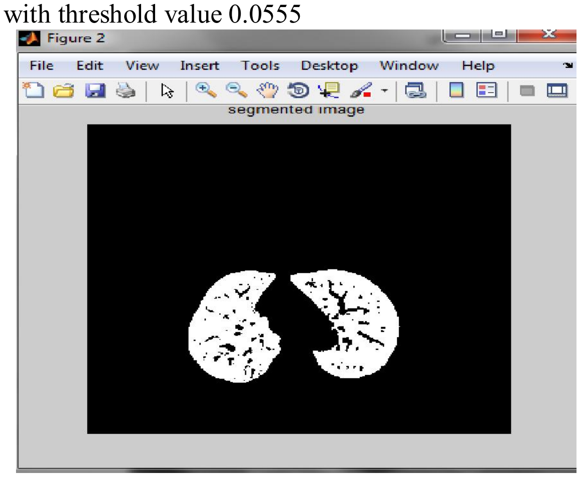

Figure 12 Lobe separation

5. FUTURE SCOPE

By this process tie complexity is reduced and diagnosis confidence is increased. This process reduces the time complexity and increases the diagnosis confidence. The collected data contain noise the noises are removed and then segmentation of the Lung images and after that the image is separated. The Lobes want to be separated and according to the features of the defected part we can conclude the cancer is present or not and the patient is in which stage. For calculation the output image is trained by using the HMM model and the diagnosis is made from the output.

6. REFERENCE

[1] World Health Organization, Cancer 2009.

[2] National Cancer Institute, A snapshot of lung cancer 2006.

[3] R. Wiemker, P. Rogalla, "Computer-aided detection (CAD) and volumetry of pulmonary nodules on high-resolution CT data", (2003).

[4] Hu S HoffmanE.A and. Reinhardt J.M, (2001), 'Automatic lung segmentation for accurate quantization of volumetric X-ray CT images', IEEE Trans.

[5] Ukil S. Sonka M, and Reinhardt F.C (2006) 'Automatic segmentation of pulmonary fissures in X-ray CT image using anatomic guidance', Proc SPIE (Med. Imag.), vol. 6144, pp. 213–223.

[6] K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu "Computer-aided diagnosis for pulmonary nodules based on helical CT images", Compute. Med. Image Graph, vol. 22, no. 2 (1998), pp. 157-167.

[7] A. El-Baz, A. A. Farag, PH.D., R. Falk, M.D. and R. L. Rocco, M.D., "detection, visualization, and identification of lung abnormalities in chest spiral CT scans: phase I", Information Conference 2002.

[8] Goldman.L.H, "Principles of CT: Radiation dose and image quality," J. Nucl. Med. Technolol. 35, pp. 213–225, 2007.

[9] B.V. Ginneken, B. M. Romeny and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey", IEEE, 2001.

[10] Wang J Betke M and. Ko J. P. (2002), 'Segmentation of pulmonary fissures on diagnostic CT Preliminary experience' in Proc. Int. Conf. Diagn. Imag.Anal, pp. 107–112.

[11] Wei, Q., Hu, "Segmentation of Lung Lobes in High resolution Isotropic CT images" IEEE Trans. Med. Imag.,vol. 56, no. 5, pp. 1–11, may 2009.

[12] Q Hu Y. Macgrego J. H r, and Gelfand G. (2008), 'Segmentation of lung lobes in clinical CT images', Int. J. CARS, vol. 3, pp. 151–163.

[13] Porat, S. & Feldman, J. A. (1991), 'Learning automata from ordered examples', Machine Learning7, 109–138.

[14] L. R. Rabiner, A tutorial on Hidden Markov Models and selected applications in speech recognition, Proc. IEEE. 77(1989) 257-286.

[15] Ghahramani, Z., & Jordan, M. I. (1997). Factorial hidden Markov models. Machine Learning, 29, 245–273.

5/22//2013