Implementation of Lung Cancer Nodule Feature Extraction using Threshold Technique

T. Siva Sakthi[1], Kumar Parasuraman[2], Arumuga Maria Devi[3]

Department of Information Technology and Engineering, Manonmaniam Sundaranar University, India[1, 2, 3]

Abstract: Image Processing and Analysis can be defined as the act of examining images for the purpose of identifying objects and judging their significance. In current days, image processing techniques are widely used in many medical areas for improving earlier detection and treatment stages, especially in various cancer nodules such as the lung cancer, breast cancer, and brain cancer and so on. Segmentation refers to the process of partitioning a digital image into multiple segments known as super-pixels. Image segmentation is typically used to locate objects and boundaries in images. In this work we are uses mean filter and median filter for image preprocessing. For image segmentation, Otsu's thresholding and marker controlled Watershed segmentation approach are used to segment the lung of CT image. The main objective of this paper is to implement lung nodule segmentation and feature extraction using digital image processing for the classification of the disease stages to avoid serious stages early and to reduce lung cancer percentage distribution.

Keywords: Segmentation, Mean filter, median filter, thresholding.

I. INTRODUCTION

Digital image processing is now a very important component in many industrial and commercial applications and a core component of computer vision applications. Image processing techniques also provide the basic functional support for document image analysis and many other medical applications. The field of digital image processing is continually evolving. Transform theory plays a key role in image processing. Image and signal compression is one of the most important applications of wavelets.

The field of image processing continues, as it has since the early 70s, on a path of dynamic growth in terms of popular and scientific interest and number of commercial applications. Considerable advances have been made over the past 30 years resulting in routine application of image processing to problems in medicine, manufacturing, entertainment, law enforcement, and many others. Examples include mapping internal organs in medicine using various scanning technologies (image reconstruction from projections), automatic fingerprint recognition (pattern recognition and image coding).

The lungs are a pair of sponge-like, cone-shaped organs [1]. The right lung has three lobes, and is larger than the left lung, which has two lobes. Anatomy of lung is shown in Fig.1. Lung cancer is a disease of abnormal cells multiplying and growing into a nodule. Fig.2 describes the beginning of the cancer. The types of lung cancer are divided into four stages. In stage I, the cancer is confined to the lung. In stages II and III, the cancer is confined to the chest (with larger and more invasive tumors classified as stage III).Stage IV cancer has spread from the chest to other parts of the body. Of all types of cancer, lung cancer is the most common cause of deaths, accounting for 1.3

million deaths annually. An estimated 159,260 people are expected to become from lung cancer in 2014, accounting for approximately 27 percent of all cancer. Early detection of lung cancer can increase the chance of survival among people.

There are many techniques to diagnose the lung cancer, such as Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. However, most of these techniques are expensive and time consuming. Therefore, there is a great need for a new technology to diagnose the lung cancer in its early stages. Image processing techniques provide a good quality tool for improving the manual analysis[13]. The use of image processing techniques can assist radiologists and doctors in diagnosing diseases and to offer a rapid access to medical information gained importance in a short time. In this paper, MATLAB has been used through every procedures made.

In technical literature done by A.Amutha and R.S.D Wahidabanu [3], Level Set-Active Contour Modeling was used as a method in diagnosing lung tumor. First step was removing noise from image using kernel based non-local neighborhooddenoising function and done feature extraction based on histogram to classify between normal and abnormal classes. At the final step or in tumor detection, level set-active contour modeling with minimized gradient to theimage was introduced. In another study [4], Auto enhancement, Gabor filter and Fast Fourier transform (FFT) were used to enhance the image and used Thresholding and Watershed segmentation to segment the image. While for feature extraction, Binarization and Masking approach were applied. N.A. Memonet. al [5] proposed thresholding method which

select the threshold based on the object and background pixel means. Region growing is used then to extract the exact cavity region with accuracy.

II. SEGMENTATION SYSTEMS

There are many techniques to diagnose the lung cancer, such as Chest Radiograph, Computed Tomography, Magnetic Resonance Imaging and Sputum Cytology. Image processing techniques provide a good quality tool for improving the manual analysis. The use of image processing techniques can assist radiologists and doctors in diagnosing diseases and to offer a rapid access to medical information gained importance in a short time.

In this paper, MATLAB has been used through every procedures made. General description of lung cancer nodules segmentation and feature extraction system that contains five basic steps. The first step starts with taking a collection of CT images from the available database. The second step applies image preprocessing, to get best level of quality and clearness. The third step is image segmentation which plays an effective role in image processing steps, and the fourth step contains feature extraction. The final step gives image diagnosis result.

A. Image Acquisition

The foremost step in medical image processing is image acquisition. The images are collected from a database of Lung Image Database Consortium [6] and also from other hospitals. The medical data is usually in DICOM format, which is the standard for storage and transfer of medical images [7]. The lung CT images having low noise when compared to X-ray scan image and MRI image. Thus the CT images are taken for detecting of the lungs. The main advantages of computed tomography image have better clarity, low noise and distortion[15]. Lung CT images are given as input. Dimensions of images are 512x512 pixels in size. The input CT image contains noises such as white noise, salt and pepper noises etc. So, image preprocessing stage is needed to eliminate noises

B. Image Preprocessing

Image preprocessing is one of the categories of image processing, attempt to make diagnostic more obvious. Image preprocessing is a way to improve the quality of image, so that the consequential image is better than the original one. In this paper, mean filter and median filter are presented for preprocessing of selecting the CT images[15]. The median filter is a non-linear tool, while the average filter is a linear one. Mean filtering is a simple, intuitive and easy to implement of smoothing images, i.e. reducing the amount of intensity variation between one pixel and the next. The median filter is normally used to reduce salt-and-pepper noise in an image. It often does a better job than the mean filter of preserving useful detail in the image. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value [8]. If the

neighborhood under consideration contains an even number of pixels, the average of the two middle pixel values is used. Both mean and median filters are used to remove noise. This preprocessing image is used as the input for image segmentation.

C. Image Segmentation

Is an essential process for most image analysis subsequent tasks. Segmentation divides an image into its constituent regions or objects [8]. The goal of segmentation is to make simpler or change the representation of an image into something that is more meaningful and easier to analyze.

III.METHODOLOGY

A. Marker-Controlled Watershed Segmentation Algorithm

Image segmentation is the first and one of the most important steps in image processing. The goal of segmentation is to find and isolate desired objects from the image background. The results of image segmentation are important in further analysis, especially in retrieval multimedia data, extraction of region-specific indexing features, in identifying regions of interest or in processing of medical images. The segmentation algorithms are based on two basic properties of intensity values discontinuity and similarity [9]. First category is to partition an image based on abrupt changes in intensity, such as edges in an image. Second category is based on partitioning an image into regions that are similar according to a set of predefined criteria. In this paper, marker-controlled watershed segmentation and Otsu's thresholding method are used.

Separating touching objects in an image is one of the more difficult image processing operations. The watershed transform treats the intensity as a function defining 'hills' and 'valleys' and attempts to detect the valleys.In simple watershed, the problem of over segmentation is faced [9].

To reduce this problem, marker-based watershed segmentation is used. Marker-controlled watershed is an improved form of watershed. In marker-based watershed segmentation, markers are used. A marker is a connected component belonging to an image. Marker-controlled watershed approach has two types: external associated with the background and internal associated with the objects of interest. The marker based watershed segmentation can segment unique boundaries from an image.Marker-controlled watershed approach has two types: external associated with the background and internal associated with the objects of interest.

The marker based watershed segmentation can segment unique boundaries from an image. Marker-driven watershed segmentation extracts seeds indicating the presence of objects or background at specific image locations. Marker locations are then set to be regional minima within the topological surface (typically, the gradientof the original input image), and the watershed algorithm is applied [10].

g(x,y) = { 1 f(x,y) > T
0 f(x,y) ≤ T

Where, g(x,y)= output image
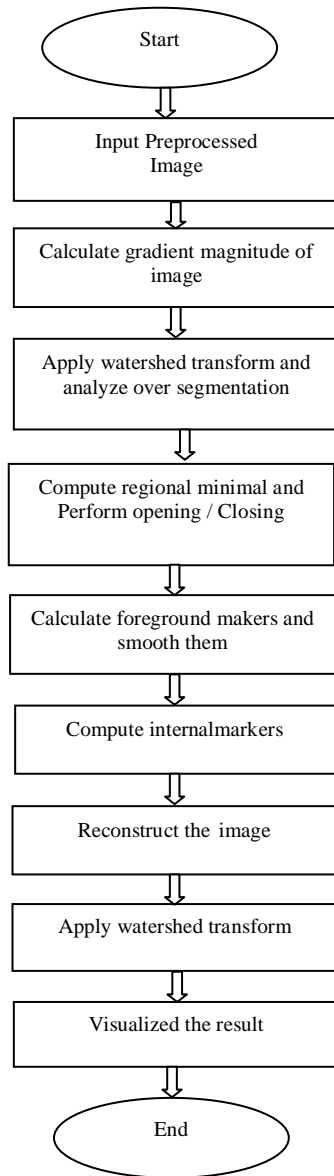f(x, y) = input image
T = threshold value

Figure1 Segmentation Flow Chat

B. Otsu's Thresholding Segmentation

Otsu's thresholding a non-linear operation that converts a gray-scale image into a binary image where the two levels are assigned to pixels those are below or above the specified threshold value [12]. The two levels in a binary image are assigned to pixels below or above the particular threshold. It is based on threshold range by statistical. Otsu suggested minimizing the weighted sum of within-class variances of the object and background pixels to establish an optimum threshold. Recall that minimization of within-class variances is equivalent to maximization of between-class variance. Threshold value based on this method is

between 0 and 1. Otsu's thresholding method is based on selecting the lowest point between two classes. Frequency and Mean value are the following equations to be calculated
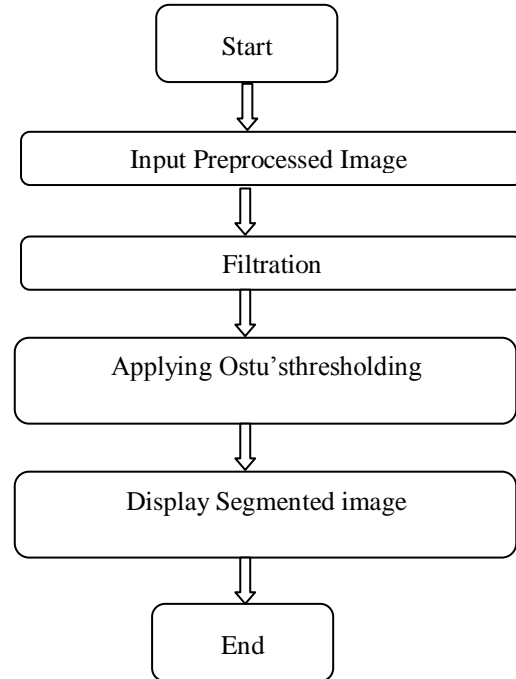

Figure2 Architecture for Ostu'sThresholding

Otsu's thresholding method is based on selecting the lowest point between two classes (peaks). Frequency and Mean value are the following equations to be calculated.

Mean:

$$\mu = \sum_{i=0}^{T} iP(i) \ / \ w \qquad [3]$$

The variation of the mean values for each class from the overall intensity mean of all pixels: between-classes variance $\sigma b2$,

$$\sigma_b^2 {}^{w}_0 (\mu_0 - \mu_t)^2 + {}^{w}_1 (\mu_0 - \mu_t)^2 [4]$$

$$\mu_t = {}^{w}_0 \mu_0 + {}^{w}_1 \mu_1, \qquad [5]$$

$$\sigma_b^2 = {}^{w}_0 {}^{w}_1 (\mu_1 - \mu_0)^2 \qquad [6]$$

w_0, w_1, w_0, w_1 stands for the frequencies and mean values of two classes, respectively. Derived from this method, threshold value represents between 0 and 1 and the segment of image will be achieved.

C. Feature Extraction

After the segmentation is performed on lung region, the segmented nodules are used for feature extraction.A feature is the basic property of any shape or object whichdistinguishes it from other objects. Feature Extraction is the process to extract the various features from an image whichdiscriminate the most important

characteristics of the image[12].The main endeavor of Feature Extraction is to lessen thenumber of false positives in detection of Region of interest (ROI). The features that are used here in this paper are as follows:

A. Shape Features
a) Circularity - It specifies the roundness of the objects i.e. how close the shape of object approaches to circle. It is given by:

$$Circularity = \frac{4\pi*(AREA)}{PERIMETER)^2}$$

b) Eccentricity - It is the ratio of the distance betweenthe Foci of the ellipse and its Major axis length. Thevalue lies between 0 and 1.
c) Solidity - It is the proportion of the pixels in theConvex Hull that are also in the ROI.
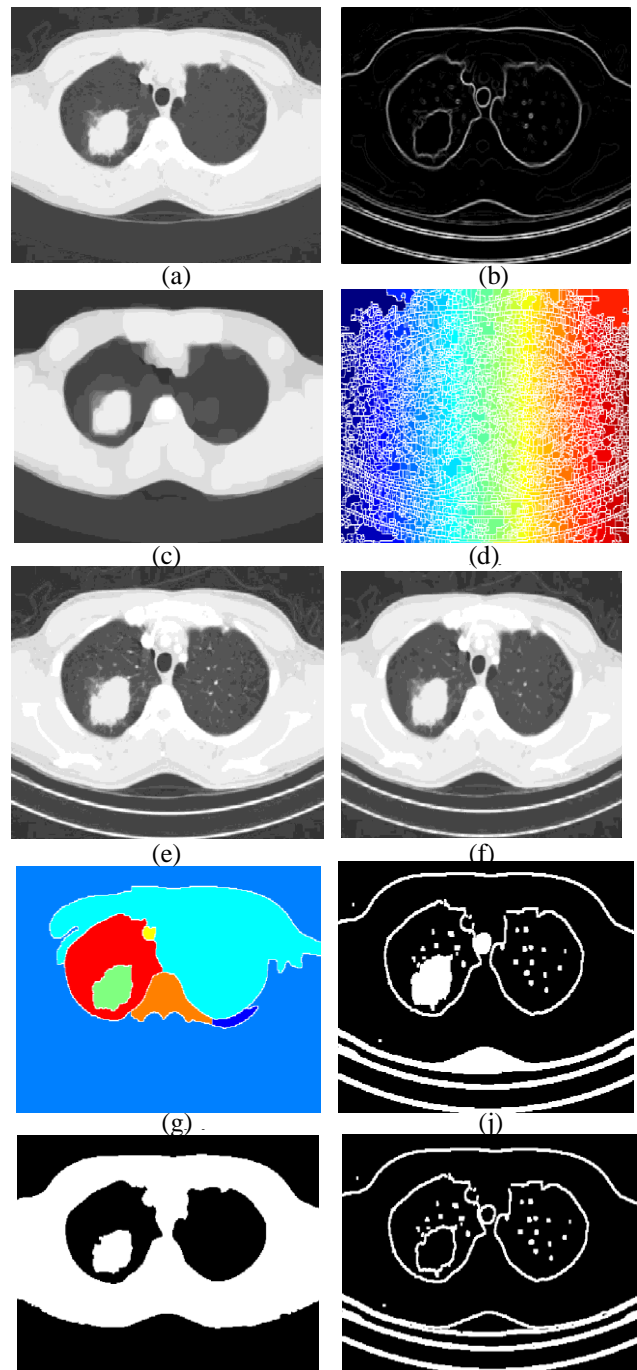Feature extraction is one of the most important steps in this system. A feature is a significant piece of information extracted from an image which provides more detailed understanding of the image. A feature is defined as a function of one or more measurements, the values of some quantifiable property of an object, computed so that it quantifies some significant characteristics of the object.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In the preprocessing step, both the mean filter and the median filter can be used to remove noise from an image. According to the outputs of image preprocessing images in Fig.8, median filter is more suitable than the average filter for this research because the main advantage of median filtering is that even after pixel intensity values are changed the edges of the images are preserved. The increasing mask size is more effective in minimizing the impact of noise. The mask size of these figure results is 15x15. Moreover, the median filter has the smallest MSE and the highest PSNR .So, the median filter is chosen for this research. According to the outputs of the segmentation step, marker-controlled watershed segmentation approach has higher accuracy and better quality than thresholding approach. After segmentation of image, morphological operation is used to get individual lung and to eliminate unnecessary parts.

By doing morphological operations, it gets not only the individual lung but also apparent the lung nodule. For feature extraction step, feature extracted from the image is used for determining and classifying the suspicious area of the lung (nodule). Features are calculated from their formulas. Calculated features and nodule size and nodule location will help doctors to take the correct decision. Finally, the performance of the system is evaluated. Three performance evaluations are accuracy, sensitivity and specificity for the image processing techniques. This system offers sensitivity of 100%, specificity of 83.33% and accuracy of 90.909%. The result of this system and the analysis of doctor together increase the accuracy of detecting malignant and benign lung nodules. If the

malignant nodule is detected in its earlier stage, then the chance of surviving the patient increases.

(a)   original   image,   (b)preprocessing,   (c)gradient madnitude, (d)water transform of gradient magnitude, (e)opening by reconstruction, (f)otsu threshold image, (g)boundary line., (h)colour watershed lable, (i)fill holes, (j)segmented image

IV.CONCLUSION

The mortality rate of lung cancer is the highest among all other types of cancer; it can be detected early by detecting the lung nodules. In this paper, image pre-processing and

image segmentation are implemented to obtain the diagnosis result. By using these steps, the nodules are detected and some features are extracted. The extracted features can be used for classification of disease stages. Determining the nodule features can provide to know more information of the condition of lung cancer at the early stages. This technique helps the radiologists and the doctors by providing more information and taking correct decision for lung cancer patient in short time with accuracy. Therefore, this method is less costly, less time consuming and easy to implement.

REFERENCES

[1] Non-Small Cell Lung Cancer, [online available], http://www.katemacintyrefoundation.org, Adapted from National Cancer Institute (NCI) and Patients Living with Cancer (PLWC), 2007, accessed on 13 July 2013.

[2] Anatomy of lung picture and beginning of cancer, [online available], www.allreferhealth.com, accessed on 4 June 2013.

[3] A.Amutha and R.S.D Wahidabanu, "A Novel Method for Lung Tumor Diagnosis and Segmentation using Level Set- Active Contour Modelling", European Journal of Scientific Research, Vol.90, No.2, November 2012, pp.175-187

[4] Mokhled S. Al-Tarawneh, "Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies, Issue 20, January- June 2012, p.147-158. ISSN 1583-1078

[5] N.A. Memonet. al, "Segmentation of Lungs from CT Scan Images for Early Diagnosis of Lung Cancer," World Academy of Science, Engineering and Technology. 2006.

[6] Lung Cancer Database Consortium, [online available], www.eddie.via.cornell.edu.com, accessed on 22 July 2013.

[7] Digital Imaging and Communications in Medicine (DICOM) Part 1: Introduction and Overview, National Electrical Manufacturers Association, 2006, pp. 11.

[8] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 35, no. 8, pp. 1798–1828, 2013.

[9] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of alzheimer's disease with deep learning," in Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on. IEEE, 2014, pp. 1015–1018.

[10] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in Medical Image Computing andComputer-Assisted Intervention–MICCAI 2013. Springer, 2013, pp. 403–410.

[11] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, E. I. Changet al., "Deep learning of feature representation with multiple instance learning for medical image analysis," in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014, pp. 1626–1630.

[12] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for ct images," Computer methods and programs in biomedicine, vol. 113, no. 1, pp. 202–209, 2014.

[13] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative et al., "Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis," NeuroImage, vol. 101, pp. 569–582, 2014.

[14] T. A. Lampert, A. Stumpf, and P. Gancarski, "An empirical study into annotator agreement, ground truth estimation, andalgorithm evaluation," arXiv preprint arXiv: 1307.0426, 2013.

[15] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images", Computer methods and programs in biomedicine, pp. 202-209, 2014.

[16] A. Tariq, M. U. Akram, M. Y. Javed, "Lung nodule detection In CT images using Neuro Fuzzy Classifier", Proceedings of the fourth International workshop on Computational Intelligence in Medical Imaging(CIMI), pp. 49-53, 2013.

BIOGRAPHIES

Siva Sakthi T received M.Sc. degree in Information Technology from Sri Ram Nallamani Yadava college, Tenkasi, affiliated by Manonmaniam Sundaranar University, Tirunelveli, India in 2014, and student of M.Phil Information Technology in Manonmaniam Sundaranar University, Tirunelveli, India in 2016

Kumar Parasuraman received M.Sc. degree in Information Technology from Alagappa University, Karaikudi, in 2006, M.Tech. degree in Computer and Information Technology from Manonmaniam Sundaranar University, Tirunelveli, India in 2008 and Ph.D. degree in Information Technology - Computer Science and Engineering, India in 2012 from Manonmaniam Sundaranar University, Tirunelveli. Currently, he is working as Assistant Professor, Center for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli. His research interests include Image Processing, Segmentation and Pattern recognition. He is a Member, IEEE.

T. Arumuga Maria Devi received B.E. Degree in Electronics and Communication Engineering from Manonmaniam Sundaranar University, Tirunelveli India in 2003, M.Tech degree in Computer and Information Technology from Manonmaniam Sundaranar University, Tirunelveli, India in 2005 and also worked as a Lecturer in Department of Electronics and Communication Engineering in Sardar Raja College of Engineering.She received Ph.D in Information Technology - Computer Science and engineering from Manonmaniam Sundaranar University in year 2012 and also the Assistant Professor of Centre for Information Technology and Engineering of Manonmaniam Sundaranar University. Her research interests include Signal and Image Processing, Multimedia and Remote Communication.