LUNG CANCER DETECTION USING IMAGE PROCESSING

Prof M.Z.Khan, Ankita Deshmukh, Kalyani Jog, Krutika Juware, Neha Khorgade
Department of Electronics and Telecommunication Engineering
Anjuman College of Engineering and Technology, Nagpur, Maharashtra, India

Abstract: — Lung cancer seems to be the common cause of death among people throughout the world. Cancer detected at an Early stage can increase the chance of survival among people. It is shown that the overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected at an early stage. X-ray & Computed Tomography (CT) is used to detect the disease. Computed Tomography (CT) is considered to be more efficient than X-ray. However, problem seemed to merge due to time constraint in detecting the presence of lung cancer regarding on the several diagnosing method used. Hence, a lung cancer detection system using image processing is used to classify the present of lung cancer in an CT-images. In this study, MATLAB have been used through every procedures made. In image processing procedures, process such as image pre-processing, segmentation and feature extraction have been discussed in detail. We are aiming to get the more accurate results by using various enhancement and segmentation techniques.

Index Terms- **CT image, segmentation, enhancement, extraction, pre-processing**

I. INTRODUCTION

The image quality of computed tomography (CT,) has increased considerably during the past years. However, the diagnosis made by the radiologist becomes more cumbersome as the amount of data increases. Computers can be a great help analyzing the data. This is called computer-aided diagnosis (CAD): the radiologist or clinician makes the diagnosis, incorporating output from computerized image analysis of the medical image.

In diagnosis of breast cancer, computerized detection of lesions in mammograms is already possible [Giger and Armato III, 2001]. Also for lung cancer screening computer methods are being developed. People who are screened for lung cancer can have small, potentially cancerous nodules (tumors) in their lungs. By means of CAD it is possible to detect those nodules in CT-datasets automatically, which is useful for close observation or biopsy. Such lung nodule detection algorithms assume that the lungs are in the image in a standardized way. A robust algorithm that first detects if and where the lungs (or parts of them) are present in a dataset can increase the robustness of the nodule detection algorithm.

This allows for fully automated detection of lung nodules, even when the scan was not acquired for lung viewing. When the lungs have been detected, a binary volume must be created that can be used in nodule detection algorithms: segmentation. Segmentation also allows for visualization of the tissue inside the lungs, such as blood vessels and nodules. This visualization gives the radiologist a very quick overview of what is inside the lungs of the patient, which facilitates the diagnosis. Since 1989, when the idea of computer assistance for detecting nodules in CT datasets was introduced [Preteux, 1989], a number of articles has been written concerning lung and lung nodule detection in high-resolution CT images. Examples are [Kanazawa et al., 1994], [Brownet al., 1997], [Hu et al., 2001], [Armato III et al., 2001], [Gurcan et al., 2002]. In all these methods, it is assumed that the lungs are completely present in the CT image, and that the lungs are the only large structures in the image within the same gray value range. As is pointed out in the introduction of chapter four, there exist other structures in the body, with the same gray value range as the lungs. So distinction must be made between the lungs and those other structures, such as the colon and the cavities in the head [13].

The goal of this project is to automatically detect if and where (parts of) the lungs are present in a CT dataset and to create a binary volume suitable for nodule detection and visualization. To achieve this, a way must be found to distinguish lungs from other organs, such as the colon. The structure of this report is as follows. First some background knowledge is given about lung anatomy and their appearance in CT datasets. In chapter three some basic morphological operations are described, together with a description of the working environment and the used datasets.

II. LITERATURE SURVEY

Korfiatis et al. have proposed a computer aided scheme for identification and characterization of Interstitial Pneumonia patterns in MDCT [7]. First the segmentation of CT scan has been performed in two stages - lung field segmentation and vessel tree segmentation. The co-occurence matrix for all regions are calculated from which eleven GLCM features are then extracted in five distances across four different directions, which provides the feature vector. This results in high dimensionality, which has been reduced through stepwise discriminant analysis (SDA). Finally, classification of image into three classes has-been done using k-NN classifier.

M.H Fazel Zarandi et al. discussed the classical fuzzy approach to lung segmentation [9]. Optimal smoothening of the input image has been done to reduce the noise followed by fuzzy c means clustering of the required lung tissues.
Arati S Kurani et al. explained the feature extraction using Concurence matrix for 2D images and then moved on to how a similar technique can be used to construct cooccurence matrix for volumetric data [11]. It gives deep understanding of the Co-occurence feature extraction.

Francisco Moreno-Seco et al. proposed the KNN classification method. Hedi scusses the advantages and disadvantages of KNN classification and proposed a modified classifier called K-NSN classifier which is more efficient than KNN classifier.

M. Gomathi et al. explained the fuzzy segmentation for lungs and gives four other varied techniques [14]. The modified fuzzy segmentation technique is made use of in this system. The average weight used for redefining cluster centre gives better performance than the ordinary fuzzy segmentation.

In Nihad Mesanovic , Mislav Grgic, Haris Huseinagic, Matija Males, Emir Skejic, Muamer Smajlovic proposed CT Image Segmentation of the Lungs which uses Region Growing Algorithm The procedure in Region growing algorithm initiates with a seed pixel, it examines other pixels that surrounds it, then determines the most similar one, and, if it meets certain criteria, it is included in the region. This process is continued and the region is iteratively grown by examining all unallocated neighbouring pixels to the region.

In Nikita Pandey, Sayani Nandy came up with a proposed a novel approach for detection of cancerous cells from Lungs CT scan images. This work proposed a method that detect the cancerous cells successfully from the CT scan images by dropping the detection inaccuracy made by the physicians naked eye for medical study based on Sobel edge detection and label matrix. Sobel operator helps to find the edges in an image; it does so by calculating the image gradient. Image gradient is defined as the change in the intensity of the image. Prof. Samir Kumar Bandyopadhyay provides another method using Computer Aided Diagnosis System (CAD) for detection of edges from CT scan images of lung for detection of diseases.

WhereverTimesisspecified,TimesRomanorTimesNewRomanmaybeused.Ifneitherisavailableonyourwordprocessor,pleaseusethefontclosestinappearancetoTimes.Avoidusingbit-mappedfonts.TrueType1orOpenTypefontsare required. Please embed all fonts, in particular symbolfonts, aswell, formath, etc.

III. PROPOSED METHODOLY

3.1 Image Acquisition

The first stage involves image acqusition. It starts with collecting a collection of CT images (normal and abnormal) from the available database. The second stage applies image enhancement which involves Gabor filter, auto enhancement and Fast Fourier transform techniques. The third stage consist of image segmentation algorithms which are applied to extract the affected region from the image. Thresholding and watershed segmentation algorithms which plays an effective role in image processing stages are also used in this stage. In the fourth stage, the general features of an image like average intensity, area, perimeter, eccentricity and texture features are obtained from enhanced segmented image. With the help of classifier in the classification stage tumour is classified as a normal or abnormal.

3.2 Image pre-processing

The first stage in the image Pre-processing stage is image enhancement. Image enhancement is done to provide better input for other automated image processing techniques. For this reason, images have to undergo several preprocessing process. In Image pre-processing process smoothing, enhancement, and segmentation is done.

3.3 Image Enhancement

Enhancement technique is used to improve the perception of information in images for human viewers, or to provide better input for other automated image processing techniques. In image enhancement stage three techniques used: Gabor filter, auto-enhancement and Fast Fourier transform techniques Image enhancement can be classified in two main categories, spatial domain and frequency domain. Here gabor filter is used for enhancement purpose as it gives better result compared to fast fourier and auto enhancement. A Gabor filter is a linear filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function [5]. The Gabor function has been recognized as a very useful tool in computer vision and image processing, especially for texture analysis, reason being its optimal localization properties in both spatial and frequency domain. Auto enhancement, involuntarily adjusts and enhances.

3.4 Image Segmentation

Segmentation divides an image into its constituent regions or objects. The segmentation of medical images in 2D, has many useful applications in the medical field. Image segmentation is an important process for most image analysis tasks. The result of image segmentation is a set of segments that collectively cover the whole image, or a set of contours that are extracted from the image (edge detection). Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, varaince or texture. Adjacent regions may significantly differ with respect to the same characteristic.

In image segmentation, thresholding and watershed segmentation techniques are used. Thresholding is one of the most powerful tools for image segmentation which is a non-linear operation that converts a gray-scale image into a binary image where the two levels are assigned to pixels that are below or above the specified threshold value. The segmented image thus obtained from thresholding has the advantages of smaller storage space, fast processing speed and effortlessness in manipulation, compared with gray level image which usually contains 256 levels [6]. Watershed segmentation extracts seeds indicating the presence of objects or background at specific image locations. The marker based watershed segmentation is capable of segmenting unique boundaries from an image.

3.5 Feature Extraction

The Image features Extraction stage is very important in our working in image processing techniques which using algorithms and techniques to detect and isolate various desired portions or shapes (features) of an image. Feature extraction is an essential stage that represents the final results to determine the normality or abnormality of an image.

These features act as the basis for classification process.Only these features were considered to be extracted; average intensity, area, perimeter and eccentricity. The features are defined as follows:

1) Area: It is a scalar value that gives the actual number of overall nodule pixel. It is obtained by the summation of areas of pixel in the image that is registered as 1 in the
binary image obtained.

2) Perimeter: It is a scalar value that gives the actual number of the outline of the nodule pixel. It is obtained by the summation of the interconnected outline of the registered pixel in the binary image.
3) Roundness Eccentricity: This matric value aroundness or circularity or irregularity index (I) is to 1 only for circular and it is <1 for any other shape. Here it is assumed that, more circularity of the object.

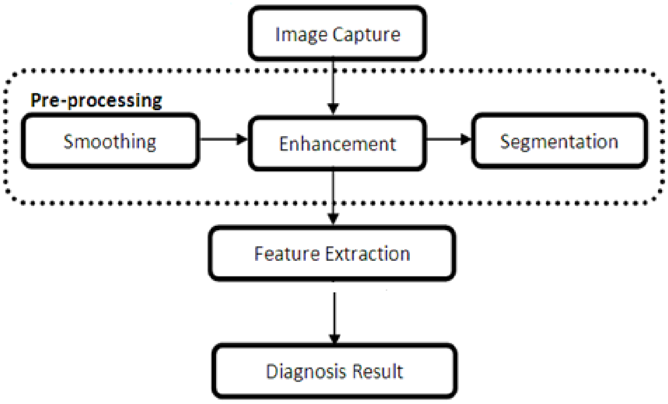

Fig.1

IV. RESULTS

4.1 RESULT OF IMAGE ENHANCEMENT

We can define image enhancement as away to improve the quality of image, so that the resultant image is better than the original one, the process of improving the quality of a digitally stored image by manipulating the image with MATLAB software. It is quite easy, for example, to make an image lighter or darker, or to increase or decrease contrast. MATLAB also supports many filters for altering images in various ways .

Gabor Filter Enhancement Technique

The Gabor filter was originally introduced by Dennis Gabor, we used it for 2D images (CT images). The Gabor function has been recognized as a very useful tool in computer vision and image processing, especially for texture analysis, due to its optimal localization properties in both spatial and frequency domain.

Auto Enhancement Technique

Auto enhancement, automatically adjusts and enhances the image (brightness, color and contrast) to optimum levels, and this is clearly observed by following (Figure 4.2) as you see (a) is the original image and (b) is the image after applying auto enhancement MATLAB code. This method strongly depends on statistical operations such as mean, variance calculation.

Fast Fourier Transform technique

Fast Fourier Transform technique operates on Fourier transform of image. The frequency domain is a space in which each image value at image position F represents the amount that the intensity values in image I vary over a specific distance related to F. Fast Fourier Transform “FFT” is a faster version of the Discrete Fourier Transform (DFT).

4.2 RESULTS FOR SEGMENTATION

Segmentation divides an image into its constituent regions or objects. The segmentation of medical images in 2D, slice by slice has many useful applications for the medical professional: visualization and volume estimation of objects of interest, detection of Abnormalities (e.g. tumors, etc.), tissue quantification and classification, and more.

Thresholding approach:
Thresholding is a non-linear operation that converts a gray-scale image into a binary image where the two levels are assigned to pixels that are below or above the specified threshold value.

Watershed Segmentation Approach

Separating touching objects in an image is one of the more difficult image processing operations, however it has no smoothing/generalization properties. The marker based watershed segmentation can segment unique boundaries from an image.

4.3 RESULTS FOR FEATURE EXTRACTION

Feature extraction is an essential stage that represents the final results to determine the normality or abnormality of an image.
Features estimated for separated nodule has been found as follow:
Average intensity: 226.0;
Area: 1024;
Perimeter: 196.3;

Eccentricity: 0.78;

These features act as the basis for classification process. Regarding on the type of the image processed, which is binary image, the only color presented is black and white. Thus, only three features were considered to be extracted; area, shape and perimeter. The features are defined as follows:

1) Area: It is a scalar value that gives the actual number of overall nodule pixel. It is obtained by the summation of areas of pixel in the image that is registered as 1 in the binary image obtained.

2) Perimeter: It is a scalar value that gives the actual number of the outline of the nodule pixel. It is obtained by the summation of the interconnected outline of the registered pixel in the binary image

3) Eccentricity: This matrix value or roundness or circularity or irregularity index (I) is to 1 only for circular and it is 25mm, and have a faster growth rate. In the normal images nodule size is less than 25mm. And in the abnormal images its size is greater than 25mm. In the segmentation that nodule is detected and than we use feature extraction to extract the features from that segmented image by which we can identify the stages of lung cancer.

Lung nodule show up as round, white opacities on chest x-rays and computed tomography scans. Previous scan x-ray or scan and the current x-ray and ct-scan is used to determine if there is any change in shape, size, or appearance of the nodules. If the nodule do not grow larger after monitoring for a 2 year period, no further treatment is necessary.

V. CONCLUSION

Since lung diseases are dynamic and evolutionary in nature, their detection, treatment will also progress based on the dynamic nature of the disease. The treatment detection technique in lung cancer keeps on changing with time. Therefore, the image processing technique must also progress in a direction of finding cancer as early as possible. The screening techniques in cancer detection must be reliable, robust and must have high- level of diagnostic value. For this purpose, right from the image acquisition to the detection of cancer we must be very careful enough to identify the real indicators of cancer nodules. For this purpose, we have used algorithm inspired by the analogy of filling water in an uneven topology. In our case, also cancerous part of the lung becomes very high or a region having global maxima and non- cancerous part having less intensity as compared to cancerous part or simply the complete lung tissues having typical intensity value with respect to their nature. This process is called watershed algorithm with distance transformation. This algorithm goes highly successful in segmenting or identifying the cancerous nodule after the removal of thorax or background of lungs.

REFERENCES

[1.] M. Weeks, Digital Signal Processing Using Matlab and Wavelets, Infinity Science Press LLC, 2007

[2.] N. Ahmed, Joint Detection Strategies for Orthogonal Frequency Division Multiplexing, Dissertation for Master of Science, Rice University, Houston, Texas. pp. 1-51, Apr, 2000.

[3]. J. D. Poston and W. D. Horne, "Discontiguous OFDM Considerations For Dynamic Spectrum Access In Idel TV Channels," in Proc. IEEE DySPAN, 2005.

[4]. R. Mirghani, and M. Ghavami, "Comparison Between Wavelet-Based And Fourier-Based Multicarrier UWB Systems", IET Communications, Vol. 2, Issue 2, pp. 353-358, 2008.

[5]. R. Dilmirghani, M. Ghavami, "Wavelet Vs Fourier Based UWB Systems", 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, pp.1-5, Sep. 2007.

[6]. S. R. Baig, F. U. Rehman, and M. J. Mughal, "Performance Comparison of DFT, Discrete Wavelet Packet and Wavelet Transforms in an OFDM Transceiver for Multipath Fading Channel,", 9th IEEE International Multitopic Conference, pp. 1-6, Dec. 2005.

[7]. Xue Li, , Vasu D. Chakravarthy, , Bin Wang, and Zhiqiang Wu, "Spreading Code Design of Adaptive Non- Contiguous SOFDM for Dynamic Spectrum Access" in IEEE Journal of Selected Topics In Signal Processing, Vol. 5, No. 1, Feb. 2011.