RESEARCH ARTICLE

PRE-PROCESSING AND SEGMENTATION TECHNIQUES FOR LUNG CANCER ON CT IMAGES

*Shyamala, S. and Dr. Pushparani, M.

Department of Computer Science, Mother Teresa Women's University, Kodaikanal

ABSTRACT

In a Human body the cell that grows out of control forms a tumour called cancer. The tumour found in the lungs is named after it as lung cancer which endangers the human life the early identification increase the patient's survival in this world. For detection of lung nodules Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are the computer aided diagnosis techniques are used. This paper mainly discuss about the pre-processing and segmentation process. Pre-processing step focus to reduce the noises in CT images and segmentation refers to the process of partitioning the pre-processed CT image into multiple regions. Further this paper briefly explains feature extraction and classification.

INTRODUCTION

When normal lung cells sustain genetic damages, it leads to uncontrolled cell proliferation. Like that of all cancers lung cancer cells have the ability to affect the neighboring tissues and spreads to other parts of the body. If left untreated, lung cancer slowly kills that person (Tina, 2005). Lung cancer is an eminent killer in the United States of men and women of all ethnicities. Lung cancer kills more people every year than breast cancer, colon cancer, and prostate cancer. The American Cancer Society estimates that nearly 180,000 new cases of lung cancer are found and lung cancer deaths are nearly 170,000. This shows that every day approximately 480 people are diagnosed with lung cancer and 460 people die because of that disease (Tina, 2005). Lung cancer is the leading cause of cancer death all over the world. The World Health Organization (WHO) had reported that over millions people die due to lung cancer. The lung cancer death increases widely. WHO has clearly states that lung cancer is one of the greatest problems faced by the world in the present centuries (Tina, 2005).

The tumors are staged based on their size and its spreading to any lymph nodes in the affected area or to other parts of the organs (freetobreathe.org

First stage: A tumor found up to 5 cm wide that has not spread to any lymph nodes or other parts of the organs is defined as first stage. These tumors are usually surgically removable. (freetobreathe.org)

Second stage: Cancers which are larger than 5 cm wide that has spread to lymph nodes on the same side of the chest, or might have begun to develop in other structures within the chest. These tumors are usually surgically removable (freetobreathe.org).

Third stage: A tumor which has spread to the lymph nodes of the chest on the same side and the same had not spread to other organs outside the chest is noted as the third stage. The tumors at this stage are unresectable (freetobreathe.org).

Fourth stage: Cancer accompanied by metastasized (spread) or that has pleural effusion to other parts of the body is stated as fourth stage (freetobreathe.org).

---

*Corresponding author: Shyamala, S.
Department of Computer Science, Mother Teresa Women's University, Kodaikanal

In the field of diagnosing the lung cancer the computer-aided diagnosis systems plays an efficient role in detecting and in characterization the various lesions (Mahersia, 2015). Here we are mainly discussing how pre-processing and segmentation process helps in diagnosing the lung cancer and how it briefly explains the features classifications and extractions.

MATERIALS AND METHODS

For the detection of pulmonary nodules different methods are used they may be of automated and semi –automated types. Anyway, to deduct the pulmonary nodule four steps are followed they are i) pre-processing, ii) segmentation, iii) feature extraction and iv) classification. The below shown Figure 1 explains the same (Mahersia *et al.*, 2015).
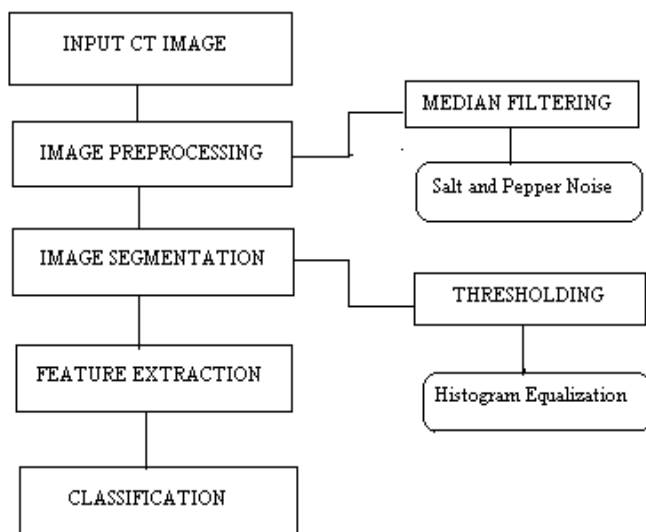

Figure 1. Lung cancer stages

Image preprocessing

For automatic diagnostic of lung pathologies Image preprocessing is the significant starting place. Wherein it simplify the identification and the classification whereby it improves overall the precision of diagnosis (Zakirov *et al.*). Median filters are widely used to minimize the presence of the noise and adjacent organs on CT-Images due to which certain types of random noise produced are reduced thereby they play a role called very good noise reduction capabilities, with considerably less blurring when compared with linear smoothing filters of same size. The median filter is the best known order-statistics filter, the value of a pixel is replaced by the median of the gray level found in the neighborhood of that pixel

$$f^{\wedge}(x, y) = \underset{s,t \,\in s_{xy}}{\text{median}}\{g(s, t)\}$$

In the computation of the median the original value of the pixel is included (Gajanand Gupta, 2011).

The output of the median filter is shown below

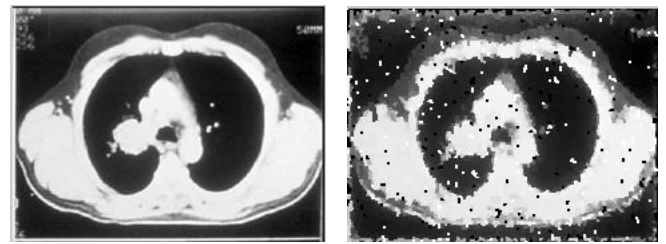The below described algorithm is used to remove salt and pepper noise in the corrupted image.

Figure 2. a)Affected CT imageb) k-median output image

Algorithm

Step I: First of all the two dimensional window (denoted by 3x3 W) of the size 3x3 is selected and it is placed in the center around the processed pixel p(x,y) in the corrupted image.

Step II: Secondly according to the ascending order the pixels are sort in the selected window and find the median pixel value denoted by (Pmed), the minimum pixel value (Pmin) and the maximum pixel value (Pmax) of sorted vector V0. Now the last and first elements of the vector V0 are the Pmax and Pmin respectively and the middle element of the vector is the Pmed.

Step III: Thirdly if the processed pixels are within the range Pmin < P(x,y)<pmax, Pmin > 0 and Pmax <255, it is stated as uncorrupted pixel and left unchanged. Otherwise p(x,y) is stated as corrupted pixel.

Step IV: If p(x,y) is corrupted pixel, then we have the following two cases:

Case i: If Pmin < Pmed < Pmax and 0< Pmed <255, replace the corrupted pixel p(x,y) with Pmed

Case ii: If the condition in case i is not satisfied then Pmed is a noisy pixel. In this case on hand compute the differences between each of the pair of adjacent pixel across the sorted V0 and obtain the difference vector VD. Now find maximum difference in that VD and mark its corresponding to the pixel in the V0 to the processed pixel.

Step V: Step i to step V are replaced until the processing is completed for the entire image (Gajanand Gupta, 2011).

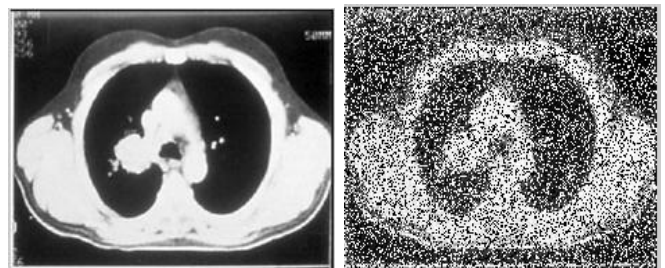### The output of the salt and pepper is shown below

Figure. 3 :a)Affected CT image b)Salt and Pepper output image

Image Segmentation

In the methods processing scheme segmentation of the lung regions is the second stage. The pre-processed CT images are partitioned into multiple regions to separate the pixels corresponding to the lung tissue from the surrounding anatomy. For the image segmentation Thresholding approach has been used (Hiram Madero Orozco, 2015).

Binary Threshold

It is the simplest form of image segmentation is called the Threshold. Based on the grayscale image, thresholding are used to create binary images

Categorizing thresholding Methods

For making the threshold completely automated, it is just and necessary for the computer to select automatically the threshold T. Sezgin and Sankur (2004) wherein it categorize thresholding methods by the following six groups, the algorithm manipulates based on the information:

Histogram shape-based methods are the valleys, curvatures, and peaks of the smoothed histogram were analyzed by

- Clustering-based methods, where the gray-level samples are brought together in two parts as foreground and background (object).
- In entropy-based methods results in the algorithms that use the entropy of the background and foreground regions, the cross-entropy between the binarized and original image, etc.
- In object attribute-based methods searches a measure of similarity between the binarized images and the gray-level, fuzzy shape similarity, edge coincidence, etc.
- In spatial methods use higher-order probabilities distribution and or correlation between the pixels
- To the local image characteristics the threshold value on each pixel are adapted in the local methods. In the local methods, a different T is selected for each pixel in the image.

Analyzing images using image thresholding techniques

Image thresholding is an easiest and effective way of partitioning an image into a background and foreground. The image analysis technique is the type of image segmentation which isolates objects where the grayscale images are converted into binary images. With the high levels of contrast the Image thresholding is most effective. Histogram Equalization algorithm is included in the common image thresholding method. For enhancing the contrast Histogram equalization is used. The contrast will not be increasing throughout the way alone. In certain cases histogram equalization can be worse on such cases the contrast is decreased.

Histogram Equalization

Histogram Equalization is an algorithm where digital images are represented here with two dimensional pixel arrays. At a given point each pixel indicates the brightness or color of that image. In the whole brightness scale Histogram equalization creates an image where the levels of brightness are equally distributed. The MATLAB is a high-performance language in technical computing integrates such as computation, visualization, and programming, which permits algorithms for execution and simulation. For many programming languages MATLAB has similar syntax i.e., allowing one to create a code

parallel to that algorithm. The results of a larger set of data set can be obtained from a small set of data when it is used to test an algorithm in the MATLAB. We chose to implement the algorithm using loops and lower-level commands for portability of VHDL though; MATLAB has a histogram equalization function (histeq). Outputs for the each stage of the histogram equalization algorithm were derived for a 4x4 array. In the MATLAB the algorithm was implemented and it is verified using the 4x4 array and the Graphics Interchange Format (GIF) image from the previous test flight simulation. The algorithm is then coded in VHDL and it is tested using the 4x4 matrix (Stephanie Parker and Kemi Ladeji-Osias).

Algorithm

The histogram equalization algorithm used is established in this first step. There are five steps to perform histogram equalization.

(i) For an NXM image of the G gray-levels (often 256) create an array H of length G initialized with zero values.
(ii) Gives the image histogram: Scan every pixel and increment the relevant number of H—if pixel p has intensity gp, perform

$$H(gp) = H(gp) + 1$$

(iii) Gives the cumulative image histogram Hc:

$$Hc(0) = H(0)$$
$$Hc(p) = Hc(p-1) + H(p), p....2 \text{ to } G$$

(iv) Rescan the image and write an output image with gray-levels gq, setting
$$gq = T(gp) \quad (3)$$

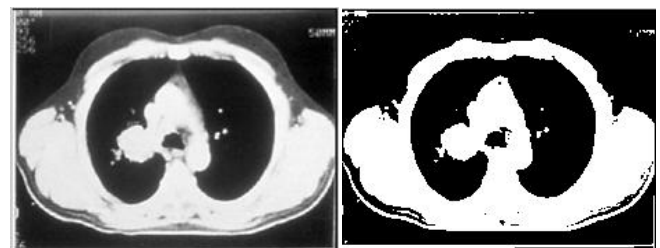The output of the thresholding is seen in Figure 4.

Figure 4. a) Affected CT image, b)Threshold Output

Feature Extraction

In a texture the objects contained inside a CT scan can offer great information to describe the objects in the medical imaging. In an artificial vision implementations Texture plays an important role. The spatial distribution of gray levels in a neighborhood is characterized by Texture. So to that extent the texture cannot be defined at a point. Based on the resolution where an image is observed is determined by the scale where the texture is perceived (Hiram Madero Orozco, 2015).

In an examined organ the most important source of information on the state of the health is offered by Texture in

CT images. The healthy counterparts are smoother whereas the diseased tissue usually has more rough structure, the same can be characterized quantitatively in an automated diagnostic system. For a correct classification the quality of the extracted texture measures is of significant importance, especially when there is difference between two different tissues which becomes minor. In distinguishing malignant from benign nodules it is critical to observe that the texture at the edge of the lung nodules, at the medical point of time (Hiram Madero Orozco, 2015).

Classification

For the probability of a True Positive (TP) each and every region identified is evaluated individually (scoring) after the structure is analyzed. For the classification process several methods are in existence. They are rule based methods, cascade classifier, minimum distance classifier, Bayesian classifier, Radial Basis Function network (RBF),Multilayer perception, Artificial Neural Networks, Fuzzy logic Support Vector Machine (SVM), etc (Shraddha *et al.*, 2015).

RESULT AND ANALYSIS

On application of pre-processing technique on CT Images the noise, blurring of that image is rectified and a clear output is given as shown in the figure 5.Here Median Filter and Salt and Pepper noise are used to give the best result. Median Filter is a very good noise reduction considerably with less blurring when compared with linear smoothing filters of same size. On the output of the Pre-processed technique on CT Images Thresholding is used to convert Gray scale image into Binary image. Here Histogram Equalization algorithm is used, which results in a Thresholding output as seen in the Figure 5.
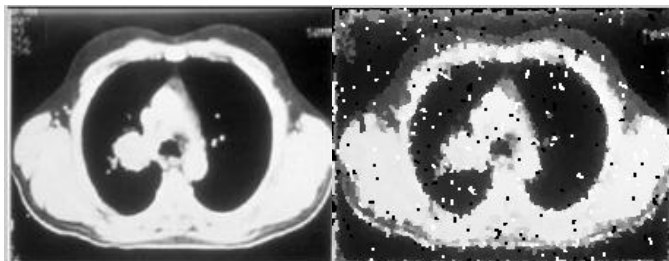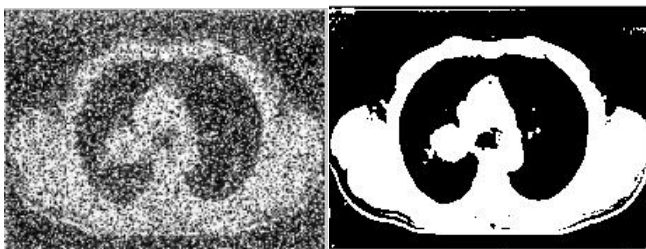

Figure 5. a) Input Image　　　**b) k-median output**

c) Salt and Pepper noise　　　**d) Threshold output**

Conclusion

Medical imaging is very powerful with the key techniques such as Pre-processing and segmentation which improves the imaging modalities and for better diagnosis by helping in integrating, planning and training software. The preprocessing and segmentation is a form of segments that collectively cover the entire image. A number of algorithms have been proposed, due to the importance of lung cancer image segmentation. Based on the images that are inputted, the selection of algorithm and its calculations should be very cautious for getting the good results. Though lung analysis techniques had improved in the past over decade yet still it needs development for better techniques of contrast and for performing evaluation is also needed (Vincent Luboz).

REFERENCES

Free to breathe, " Living With A Diagnosis Of Lung Cancer" freetobreathe.org

Gajanand Gupta, "Algorithm for Image Processing Using Improved Median Filter and Comparison of Mean, Median and Improved Median Filter" *International Journal of Soft Computing and Engineering* (IJSCE) ISSN: 2231-2307, Volume-1, Issue-5, November 2011

H.Mahersia,M.Zaroug and Gabrella, " Lung Cancer Detection on CT Images: A Review on the Analysis Techniques" *International Journal of Advanced Research in Artificial Intelligence,* Vol.4, No.4,2015.

Hiram Madero Orozco, Osslan Osiris Vergara Villegas, Vianey Guadalupe Cruz Sanchez,Humberto de Jesus Ochoa Dominguez and Manuel de Jesus Nandayapa Alfaro' "Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine", Madero Orozco *et al.* BioMedical Engineering Online (2015)

Shraddha G. Kulkarni1, Sahebrao B. Bagal, " Techniques for Lung Cancer Nodule Detection: A Survey" *International Research Journal of Engineering and Technology* (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 09 | Dec-2015 www.irjet.net p-ISSN: 2395-0072

Stephanie Parker, J. Kemi Ladeji-Osias, "Implementing a Histogram Equalization Algorithm in Reconfigurable Hardware" Morgan State University, Department of Electrical and Computer Engineering.

Tina M. St. John, M.D, "With Every Breath: A Lung Cancer Guidebook" Copyright © 2003, 2005 Tina M. St. John, M.D. All Rights Reserved. ISBN 0-9760450-2-8.

Vincent Luboz, "Image Segmentation and Preprocessing" Department of Surgery and Cancer Imperial College London

Wavelet Based Feature Extraction Scheme of Electroencephalogram

Zakirov, A.N., R.F. Kuleev, A.S. Timoshenko and A.V. Vladimirov, "Advanced Approaches to Computer-Aided Detection of Thoracic Diseases on Chest X-Rays" *Applied Mathematical Sciences*, Vol. 9, no. 88, 4361-4369. Hikari Ltd.

****