Chest X-Ray Analysis to Detect Mass Tissue in Lung

Emon Kumar Dey

Institute of Information Technology
University of Dhaka
Dhaka, Bangladesh
emonkd@univdhaka.edu

Hossain Muhammad Muctadir

Institute of Information Technology
University of Dhaka
Dhaka, Bangladesh
muctadir.hossain@gmail.com

Abstract— **Patient death ratio because of lung cancer is much higher than other types of cancer. Early detection is a must to take effective steps towards curing this disease. Medical researchers have proven that in most of the cases nodules or mass tissues in lung is an indication of lung cancer. Detection of these abnormalities at an early stage can help doctors to take proper steps towards curing this disease. The aim of this study is to make an intelligent semi-automated system that can assist detection of lung mass tissues from X-ray of chest. Simple image processing techniques like morphological operations, template matching etc. are used here which made the proposed method easily implementable, resource effective and efficient. A test on JSRT image database using this method shows accuracy and precision of 76.12% and 76.16% respectively.**

Keywords- Chest X-Ray Analysis; Lung Abnormality Detection; Mass Tissue Detection; Lung Area segmentation; DCT Based Matching; Chest X-Ray abnormality; Lung Cancer Cell Detection.

I. INTRODUCTION

Chest x-rays are used to find out abnormalities in heart, rib, especially lung. The goal of this research is to detect abnormal mass tissue in the lung by processing digital x-ray images. A mass tissue or nodule in lung can be a tumor, cancer tissue and infection. In a black and white x-ray image, a mass tissue looks like a gray shadow and sometime becomes very hard to detect. If this kind of tissue can be detected automatically, we can overcome human errors and help diagnose diseases more effectively. A lung mass tissue is defined as a spot on the lung image that is 3 cm or larger in diameter and is more likely to be cancerous. Mass tissue can also point to a number of other diseases. They are deadly and can cause lung failure if not detected and treated as soon as possible. People having smoking habit have more than 50% probability of containing nodule and abnormal mass tissue in their lungs at some point of their life. There are various types of digital x-ray machines in the market but most of them are not capable to generate automated report. CT scan is another advanced x-ray based medical imaging technology which provides better results. But in developing countries CT is not accessible or affordable by the general population. So, automated x-ray result might help. In this paper we investigate a simple scheme to automatically detect nodule and mass tissue in chest x-ray. The scheme uses a simple classical template matching method.

To classify these areas several researches were conducted over a thresholded candidate using classical features like the ones cited in [1] and [2]. A number of systems were proposed for detecting lung nodules. Reducing the false positive detection is generally their common goal. Most of the methods adopt a two stage approach. First stage is for initial preprocessing to detect a set of nodules, and the second stage consists in classifying these detected regions using a pattern recognition technique, which is a combination of a feature extraction process and a classification process. The performance of the classifier depends directly on the ability of characterization of candidate regions by the adopted features.

A method described in [3] detected lung nodules from x-ray image. They removed the false positive detection by a rule based technique and statistical approach. They used a database which contained only 192 images. Local contralateral subtraction method was proposed in [4]. They developed this method resulting in reduction of false positives reported by a computer-aided diagnosis (CAD) scheme for detection of lung nodules in chest x-ray. Their method was based on the removal of normal structures in the regions of interest (ROI), based on symmetrical characteristics between the left and right lungs. They extracted two ROIs, one from the position where a candidate of a nodule is located and other from the anatomically corresponding location in the opposite lung. They applied a wavelet based multi resolution image registration method to match the two ROI. They were able to reduce 44% of false positive. In [6] a method has described where 924 digitized gray scale 500 X 500 chest x-ray images used. They divided the entire search area into 7 x 7 regions of interest (ROIs: 64 x 64 matrix size) and applied their templates to detect nodules.

To detect a chest mass tissue, a region of interest needs to be identified. Obviously the region of interest is lung part of a chest x-ray image. In a plain x-ray image, there are shadows of various intensities. Simple edge detection is not suitable for detecting lung because of noises and other unwanted information in x-ray image. In the proposed method some techniques have been applied before running edge detection to eliminate noise and better output. After extracting lung region the image is convoluted with predefined mask to find mass tissue.

Rest of the paper is organized as follows. Section II describes the proposed method. Section III describes the experimental result, and conclusion of the work is described on section IV.

II. PROPOSED METHOD

A. Preprocessing

X-ray images are formed by the different intensity of the captured radiation signal which is a black and white shadow. Without proper image enhancement technique it is not possible to get a proper output from raw X-ray image. First of all contrast stretching was implemented on the image to make the shadows more visible. Equation for contrast stretching is given below:

$$p_{out} = (p_{in} - c)\left(\frac{b-a}{d-c}\right) + a \qquad (2.1)$$

Here p_{in} is the intensity level of input pixel. c and d are respectively lowest and highest pixel value in the image and a and b are upper and lower intensity limit of an image. For example an 8 bit image has upper bound of 255 and lower bound of 0 pixels. A more robust approach is to first take a histogram of the image, and then select c and d at, say, the 5th and 95th percentile in the histogram (that is, 5% of the pixel in the histogram will have values lower than c, and 5% of the pixels will have values higher than d). This prevents outliers affecting the scaling so much.

A binary image is created after contrast enhancing. A threshold (T) is selected automatically using Otsu's method [9] for creating the binary image.

$$g(x,y) = \begin{cases} 1, & f(x,y) > T \\ 0, & f(x,y) \le T \end{cases} \qquad (2.2)$$

Where $g(x,y)$ is the binary image and $f(x,y)$ is the original image after contrast enhancement. In binary image some noise may be present. For better performance the noise should be eliminated. A morphological method, closing, is applied for removing the unexpected noise from the binary image. Closing is a hybrid morphological operation which is simply a dilation followed by an erosion. A 3x3 structuring element *(s)* is designed to remove the unexpected noise by applying dilation.
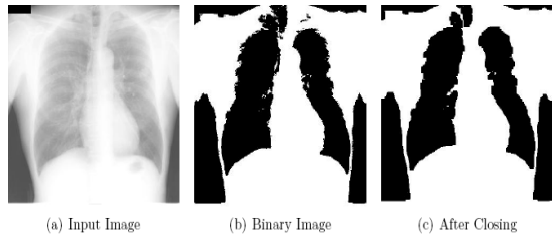
$$f \bullet s = (f \oplus S) \ominus S \qquad (2.3)$$

(a) Input Image (b) Binary Image (c) After Closing

Fig. 2.1 Morphological closing operation in an X-ray Binary image.

Here *f* is the original image, *s* is structuring element and $f \bullet s$ is the resulting image after the closing operation. $\oplus$ and $\ominus$ denote the morphological dilation and erosion.

B. ROI extraction

This work used JSRT X-ray image database [10]. This database contains raw images each of which has a resolution of 2048 x 2048 pixels. The database also contains left and right lung mask manually extracted from the X-ray images. These masks are used to detect lung region which is our ROI.

To detect separate lung area, firstly spinal cord from the x-ray image is detected. It is assumed that in x-ray image spinal cord lies approximately somewhere in the middle of the image.

If we draw a histogram by plotting every position of the binary image width on X-axis and corresponding number of 0's (white pixel) among Y-axis, we shall get a local maximum line near about middle region of the histogram. This indicates the spinal cord of the x-ray image. Resolution of our binary image is 2048 x 2048. So, the histogram contains 2048 points on X axis and each point indicates the number of 0's on corresponding Y-axis. The detection of spinal cord can help the image to divide into two parts. Each part contains a lung area. Each area is then converted into 1024 x 1024 sizes.
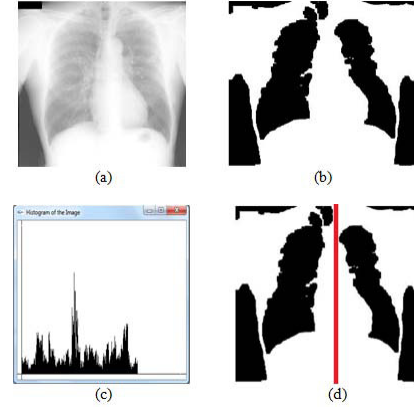

Fig. 2.2. Spinal cord detection and left and right lung separation (a) main x-ray image (b) Binary image (c) histogram of the binary image (d) Left and Right lung area.

The JSRT database contains left and right lung mask manually extracted from the X-ray images. The boundary of each mask is extracted using simple morphological operations previously on converted binary images. Then every binary contour images *B(A)* is stored. Before boundary extraction each of the binary mask images was divided into two equal parts vertically.

$$B(A) = A - (A \ominus B)$$

Where A is the main image and B is a structuring element of size 3x3. All the boundary images are normalized into 1024 x1024 sizes. Now for detection boundary from the test image the following algorithm is applied:

1. For every left lung mask contour image $B(A)$ save the co-ordinates that represents the contour in *Tmp[x,y]*.

2. Read every points of left lung area of the test image and save the co-ordinates *Tst[x,y]*.

3. Compare *Tst[x,y]* with each image of *Tmp[x,y]* pixel by pixel and calculate *Hit* and *Miss* score. Initially *Hit* and *Miss* score is set to zero.

   a) Hit Score: If *Tst(x,y)* = *Tmp(x, y)* = *1*, then increment Hit Score.

b) Miss Score: If *Tst(x,y)* ≠ *Tmp(x, y)*, then increment Miss Score.

4. Calculate the Hit ratio.

   Hit Ratio:

   $$Hr = H / (H + M)$$

   Where, H = Total Hit Score

   M = Total Miss Score

5. Find the highest Hit score for the left lung boundary contour image and plot the boundary into the main binary image of the left lung. This boundary area is the expected ROI for left lung.

6. Apply step 1 to 5 for detecting ROI for right lung.

(a)   (b)

Fig. 2.3 ROI selection (a) Input Binary Image (b) ROI

C. Mass Tissue Detection

Mass tissues are normally of round shape. Some circular shape areas are manually extracted from the binary x-ray images to train the system. We have extracted 70 circular shape areas from the whole JSRT database and also some (more than 25) from other reliable resources which are declared as mass tissue areas by doctors. We have assumed that at the primary stage of cancer circular mass tissue area is small and not greater than 128x128 in size. So, all the extracted circular shape area is less than of size 128x128. Each of these extracted areas are converted into multi sized (8x8, 16x16, 32x32, 64x64, 128x128 etc.) masks. Each of the masks is then convoluted through the whole ROI for left and right lung. Each time a Euclidian distance is calculated for every mask. If the calculated distance satisfied a predefined threshold value *Th* the area is marked and considered as mass tissue area.

Fig. 2.4. Mass Tissue Templates

Mass tissue detection technique from test image is given below:

The test X-Ray binary image is passed through a DCT module. The main interesting point of DCT is that, for a typical image most of the visually significant information about the image is concentrated in just a few co-efficient of the DCT [7].

So only by examining the top points of a DCT converted image the properties of the image can be tested. These following equations are used for DCT calculation as well as transformation of a two dimensional image.

$$D(u, v) = \frac{2}{\sqrt{M * N}} a(u)a(v)x \sum_{M=0}^{M-1} \sum_{N=0}^{N-1} I(m,n)Cos[A]Cos[B] \quad (2.3)$$

Where,

$$A = \frac{(2m+1)u\pi}{2M} \quad \text{and} \quad B = \frac{(2n+1)v\pi}{2N}$$

$$a(u) = \begin{cases} \sqrt{1/M}, & for\ u = 0 \\ \sqrt{2/M}, & for\ u = 1,2,3, \ldots \ldots \ldots M-1 \end{cases}$$

$$a(v) = \begin{cases} \sqrt{1/M}, & for\ v = 0 \\ \sqrt{2/M}, & for\ v = 1,2,3, \ldots \ldots \ldots N-1 \end{cases}$$

Here *D(u, v)* is the transformed image and *I(m, n)* is the main image. The whole test image (1024x1024) is divided into several (ex. 8x8, 16x16, 32x32, 64x64) blocks and in each block EQU (2.3) is applied. For a 8x8 block values are stored in a two dimensional array and then transformed in an one dimensional array by scanning the whole 8x8 array in a zigzag order. In this fashion all of the 8x8 blocks in an image are stored in a one dimensional array. This is similar to sorting the values according to importance of the pixels. High importance coefficients are located on the top-left corner of the block. First top 4 significant values are stored for each 8x8 block. In this way again the DCT is applied into each 16x16 and 32x32 blocks of the x-ray image. In these case first 8 top most values for of each 16x16 module and first 16 top most values for 32x32 modules are stored.
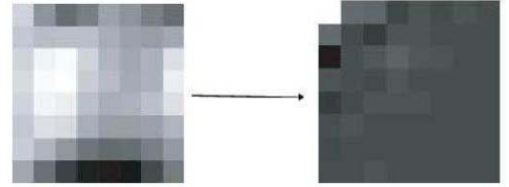

Fig. 2.5. DCT transformations (a) an 8x8 image block, (b) transformed DCT of the 8x8 Block
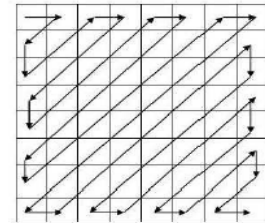

Fig. 2.6. Zigzag Scanning of a Two Dimensional DCT Coefficient

Each variable sized mass tissue area mask is also passed through the DCT module. And first top 4, 8, 16 and 32 values are stored for 8x8, 16x16, 32x32 and 64x64 mass tissue mask. The mask tissue detection classifier then calculates a Euclidian distance only for first top 4 values from the 8x8 test image of the x-ray to the 8x8 mask image of the mass tissue area. If the average distance is greater than a predefined threshold then the system identifies the area as a mass tissue of size 8x8. In the same way this classifier indentifies mass tissue area of size 16x16, 32x32, 64x64, etc.

III. EXPERIMENTAL RESULT

The experiment was done on Ubuntu Linux PC which has Core i5 2.4 GHz processor and 4GB ram. Java was used as a programming language. For image processing purpose OpenCV 2.4.3 was used. Integration between Java and OpenCV was done with the help of JavaCV. Fig 3.1 shows the brief idea of mass tissue detection.
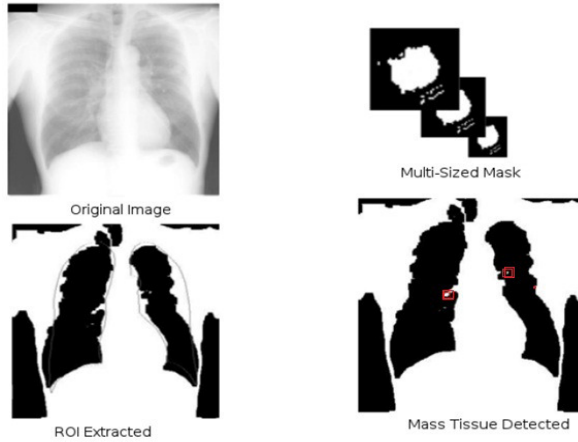

Fig. 3.1. Mass Tissue Detection Technique

The proposed method is applied on the JSRT x-ray images. Before calculating accuracy precision and accuracy false positive (f_p), false negative (f_n), true positive (t_p) and true negative (t_n) is calculated. Accuracy and precision is calculated by the following equations

$$Accuracy\ A = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} * 100\%$$

$$Precision\ P = \frac{t_p}{t_p + f_p} * 100\%$$

In case of mass tissue detection true positive is detection of a mass tissue which is present is a lung x-ray image. A false positive occurs when a mass tissue is detected which is not present in a chest x-ray image. This is also known as false

alarm. If a mass tissue is not detected, it is a false negative result.

Table 3.1 shows the experimental result of the proposed system:

Table 3.1: Experimental result of the proposed system

	Number of Images	Total abnormality	Detected correctly	False Detection	Accuracy	Precision
Case #1	20	23	16	7	69.56	69.56
Case #2	25	31	25	11	69.44	80.64
Case #3	30	26	21	11	79.12	80.76
Case #4	15	19	14	8	86.36	73.68
					76.12	76.16

Figure 3.2 shows some output of the proposed mass tissue detection system.
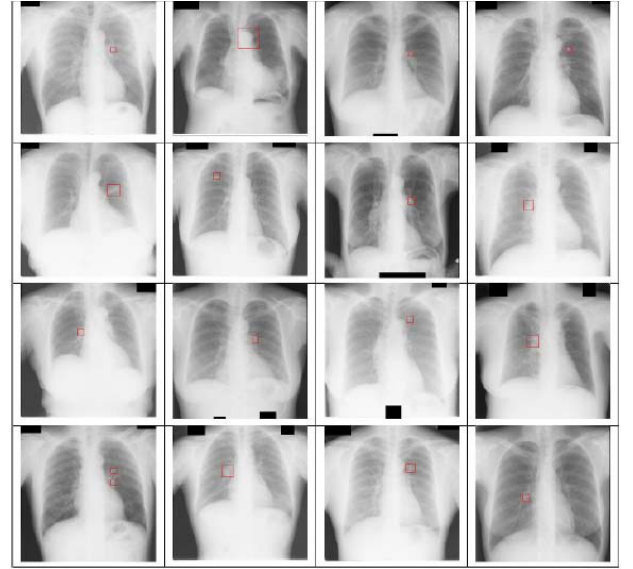

Fig. 3.2. Result of Proposed System

IV. CONCLUSION

This work presents a method for abnormal mass tissue detection on digital x-ray. It adopted the template matching technique for detecting mass tissue. Although various research has done based on template matching for mass tissue detection, this work adopted DCT based template matching which has decreased the matching time. This is suitable for real time x-ray image abnormality detection or detecting mass tissues from video image of x-ray. For simplicity this works has assumed that the mass tissue area will be 8x8, 16x16, 32x32, 64x64 in size. The mass tissue area size can be other

variable sizes. The overall accuracy and precision of this work is 76.12% and 76.16% respectively.

REFERENCES

[1] Gonzalez Rafael C., Digital image processing, Prentice Hall, 2002

[2] Pratt, William K, Digital image processing, John Wiley & Sons, 1991

[3] Hideo Suzuki;Noriko Inaoka;Hirotsugu Takabatake; Masaki Mori;Hiroshi Natori and Akira Suzuki "Experimental system for detecting lung nodules by chest x-ray image processing", *Proc. SPIE* 1450, Biomedical Image Processing II, 99 (July 1, 1991); doi:10.1117/12.44289

[4] Hiroyuki Yoshida and Kuni Doi "Computerized detection of pulmonary nodules n chest radiographs:reduction of false positives based on symmetry between left and right lungs", Proc. SPIE 3979, Medical Imaging 2000: Image Processing, 97 (June 6, 2000); doi:10.1117/12.287732

[5] B. van Ginneken, Bart M. ter Haar Romeny, M. A. Viergever, "Computer-Aided Diagnosis in Chest Radiography:A Survey", IEEE Trans. Med. Imag, Vol. 20, No.12, 2001

[6] Katsuragawa, Shigehiko, and Kunio Doi. "Computer-aided diagnosis in chest radiography." Computerized Medical Imaging and Graphics 31.4 (2007): 212-223.

[7] Dey, Emon Kumar, Mohsin Khan, and Md Haider Ali. "COMPUTER VISION-BASED GENDER DETECTION FROM FACIAL IMAGE." International Journal of Advanced Computer Science 3.8 (2013).

[8] O.Kakde, K. Bhoyar, Color image segmentation based on jnd color histogram, International Journal of Image Processing(IJIP), Volume 3(2010) 283-292

[9] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms."*Automatica* 11.285-296(1975):23-27.

[10] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. American Journal of Roentgenology, 174:71–74,2000.