

Diagnosis of Tuberculosis by using first order statistics applied to chest radiograph

Pallavi T. P., Praveen Kumar P. S.

Department of ECE, Channabasaveshwara Institute of Technology, Gubbi, Tumkur, India.

Correspondence Address: Department of ECE, Channabasaveshwara Institute of Technology, Gubbi, Tumkur, India.

Abstract

In this paper, the Textural properties of normal and tuberculosis posterior and anterior chest radiograph were looked in this investigation. Where the patients especially from rural and remote areas in which there were not having facilities like well-equipped medical centers for blood test, sputum test and skin test, not even radiologist for reading radiologic findings. Here, we first extract the lung field by an interactive lung field segmentation based on active contour models well known called as snake model. After segmentation we analyze the pixel data within region of interest by using a first order statistical approach to extract the textural properties in classifying image as TB and Non TB. For both normal and tuberculosis radiographs, the grey scale intensity distribution within the region of interest was analyzed to study their respective characteristics. Statistically the tuberculosis infected radiographs manifested higher entropy and a lower mean value in their intensity distribution as compared to their normal peers.

Keywords: Tuberculosis (TB), Chest X-ray, CAD, Snake model, and First order statistics

Introduction

Tuberculosis (TB) is a common and deadly infectious disease that can occur at any age. It is commonly caused by bacteria which are known as *Mycobacterium tuberculosis* and it mostly affects the lungs, but can also affects the other organs, including the central nervous system. The basic key to control the spread of TB is the early detection and taking treatment from the physician it is the best prevention for TB. TB bacteria had caused short lifetime to world population. According to annual report 2014, Tuberculosis remains one of the world's deadliest communicable diseases, People over the world infect TB disease about nine

million and die over one million for every year[7]. TB symptoms such as cough, getting fever, losing weight, night sweats can easily known whether people have or do not have TB disease. TB is very continuously speeded widely among the different types of people.

Chest Radiography is nowadays the primary examination tool on TB for medical examination and routine diagnosis. Physicians in general make decision on TB cases mainly based on radiologic findings, combined with clinical data. To resolve the above mentioned, in this paper we introduces a CAD to reduce the workload on radiologists and improve reading efficiency.

Materials and methods

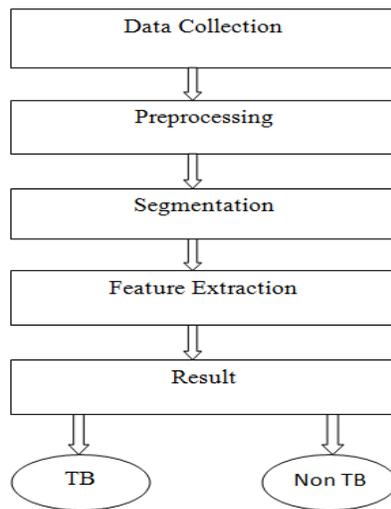


Figure 1: Flow Chart designed for the detection of Tuberculosis.

Data collection

In this study 50 normal and 47 TB infected radiograph images were collected from national institute of TB and respiratory diseases, New Delhi. Each of the confirmed TB cases was diagnosed on the basis of radiologist's reports, doctor consultation, sputum cultural tests, skin test and also blood tests[2].

For this study, radiographs were at first imported to MATLAB environment using built-in function available in the MATLAB image processing toolbox. The imported images were converted to 8-bit grayscale and sub-sampled to the size of 512x512 pixels.

Preprocessing

Most of the pre-processing is done with the help of MATLAB software. Each image sample is scanned, and stored to a size of 512 X 512 pixels. Generally during the scanning, the quality of the image is affected by the different artifacts due to non uniform intensity, variations, motions, shift, and noise[2]. Thus, the pre-processing of the

image aims at selectively removing the redundancy present in scanned images without affecting the details. Most of pre-processing methods use a small neighborhood of a pixel in an input image to get a new brightness value in the output image. Such pre-processing operations are also called filtration.

Here, in this study a Gaussian filter is used to smooth's the image.

Segmentation

Among the preprocessing processes, image segmentation is very important to extract good quality features for classification. Image segmentation are applied in many image preprocessing areas such as recognition of object, object occlusion, estimation of the boundary, editing, or query from image database, image security and image compression. Segmentation algorithms generally are based on one of 2 basis properties of intensity values.

- Discontinuity: To partition an image based on sharp changes in intensity.
- Similarity: To partition an image into the regions those are similar according to a set of predefined criteria.

Snake Model

Here, for segmentation we are using a active contour model well known as a snake model[5]. The snake or active contour models have proven to be very effective tools for image segmentation. An active contour model is essentially a curve that evolves from an initial position toward the boundary of an object in such a way as to minimize some energy functional. We have used snakes successfully for interactive interpretation, in which user-imposed constraint forces guide the snake near features of interest. Its energy function is defined as

$$E_{Snake}^* = \int_0^1 E_{snake}(v(s)) ds$$

$$= \int_0^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))$$

Here, E_{int} represents the internal energy of the snake due to bending. It is used to control the rate of stretch and to prevent the discontinuity in the contour, E_{image} represents the energy which depends upon the image and it gives rise to the image forces, E_{con} gives rise to the external constraint forces introduced by the user.

E_{con} (constraint) and E_{image} (image forces) that pulls it towards object contours and internal forces that resist deformation.

The internal energy is exposed as:

$$E_{int} = E_{cont} + E_{curve}$$

E_{cont} is the continuity of the contour and E_{curve} is smoothness of contour

It can also be written as:

$$E = \int_0^1 \left\{ \frac{\alpha}{2} |x^1(s)|^2 + \beta |x^{11}(s)|^2 + E_{ext}(x(s)) \right\} ds$$

Where $x^1(s)$ and $x^{11}(s)$ are the first and second order derivatives with respect to s , which is controlled by the parameter α and β respectively. Where α and β are the user defined parameters which controls the amount of stretch and curvature respectively. In this we present a three different energy functional which attract a snake to lines, edges and terminations. Then the total image energy can be expressed by these three energy functions as:

$$E_{image} = W_{line} E_{line} + W_{edge} E_{edge} + W_{term} E_{term}$$

Where W_{line} , W_{edge} and W_{term} are the weight of these salient features. Higher the weight indicates larger contribution to image forces.

(a) Line functional

The simplest useful image functional is the image intensity itself. If we set,

$$E_{line} = I(x,y)$$

(4)

Then depending on the sign of W_{line} , the snake will be attracted either to light lines or dark lines subject to its other constraints, the snake will try to align itself with the lightest or darkest near by contour.

(b) Edge functional

The edge functional is based on the image gradient, an image gradient is a directional change in the intensity or colour in an image, used to extract information from images.

Finding images in an image can also be done with a very simple energy functional. If we set

$$Edge = - |\Delta I(x,y)|^2$$

Then the snake is attracted to contours with large image gradients.

(c) Termination functional

In order to find terminations of line segments and corners, we use the curvature of level lines in a slightly smoothed image.

Let $c(x,y) = G^\sigma(x,y) * I(x,y)$ be a slightly smoothed version of the image.

Let $\theta = \tan^{-1}(y/cx)$ be the gradient angle and

Let $G = (\cos \theta, \sin \theta)$ and $n_1 = (-\sin \theta, \cos \theta)$ be the unit vectors along and perpendicular to the gradient direction. Then the curvature of the level contours in $c(x,y)$ can be written as:

$$E_{\text{term}} = \frac{\partial \theta}{\partial n}$$

(6)

This completes the basic snake behaviour, now here we will explain how this model is implemented in our paper.

In our CAD scheme, the user segments the lung field by first placing a number of points around the target of interest (the number of points required by the scheme is not fixed, it depends on the lung shape; but in average it takes about 8 to 12 points to accurately localize the lung), then a polygon is created based on these points, and after that put the specific number of iterations to deform the initial snake contour in accordance to the lung edges as shown in figure(2) and the final figure shows the segmented lung field.

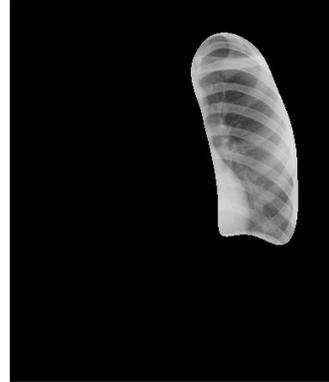
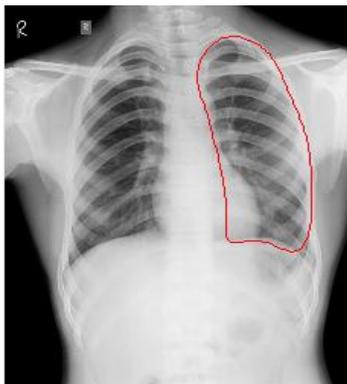
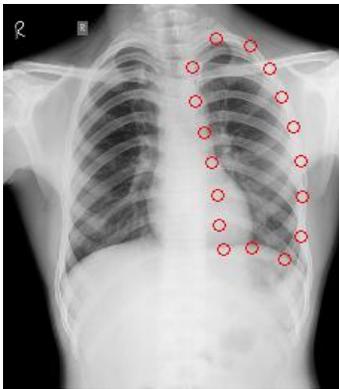


Figure 2: Segmentation of Lung.

The snake algorithm will lead a closed contour (the snake) to converge into some shapes by minimizing the following functional:

By using the Euler's equation, the above equation (3) can be reduced as:

$$\alpha_{x^1(s)} - \beta_{x^{11}(s)} - \nabla E_{\text{ext}} = 0 \tag{7}$$

And $\nabla E_{\text{ext}} = [f_x(x,y) \ f_y(x,y)]$. The term $\alpha_{x^1(s)} - \beta_{x^{11}(s)}$ can be considered as internal force and $-\nabla E_{\text{ext}}$ as external potential force. Hence while a snake progresses to reach its minimum in energy, the term $-\nabla E_{\text{ext}}$ can be considered as the external forces pulling/pushing snake into nearby edges and the internal forces discourages the snake to bend or stretch to bend or stretch during its deformation. The solution to the equation(7) is,

$$X_t = (A_s + \gamma I)^{-1} [\gamma X_{t-1} - f_x(x_{t-1}, y_{t-1})] \tag{8}$$

A_s is a pentadiagonal banded matrix; γ is step-size.

Feature extraction

Feature extraction is the process of defining a set of features, or image characteristics, which will most efficiently or meaningfully represent the information that is important for analysis and classification. When the input data to an algorithm is too large to be

processed and it is suspected to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features (also named features vector). This process is called feature extraction. The issue of choosing the features to be extracted should be guided by the following concerns:

- All features should carry enough information about the image.
- All the features should be easy to compute.
- They should relate well with the human perceptual characteristics

First order Statistics

In this scheme after segmentation we analyze the pixel data within the region of interest by using first order textural statistical properties[8]. The First order statistics are quite straight forward. They are computed from a function that measures the probability of a certain pixel occurring in an image. This function is also known as histogram.

The distribution on i is given as,

$$d(i) = \frac{\varphi(i)}{\Lambda}$$

Where, $\varphi(i)$ be the number of points whose grayscale intensity value is i in the lung field, and

Λ be the total number of points in the region of interest.

From these we get the following properties on the gray scale distribution of both left and right lungs.

Mean: $\mu = \sum_{i=0}^n id(i)$

Entropy: $E = \sum_{i=0}^n d(i) \log d(i)$

Results

Table1 and Table2 tabulate values of the features mean and entropy for the normal and the TB, of both left and right lungs. For the normal person the values of Mean will be high and the entropy will be low. As For the Tuberculosis the values of Mean will be low and the entropy will be high is as shown in below table.

Table 1: Feature values determined on left lungs of normal and tuberculosis.

Features	Normal	Tuberculosis
Mean	38.4857	21.7584
Entropy	0.8464	0.9089

Table 2: Feature values determined on right lungs of normal and tuberculosis.

Features	Normal	Tuberculosis
Mean	28.8556	27.6727
Entropy	0.8725	0.9089

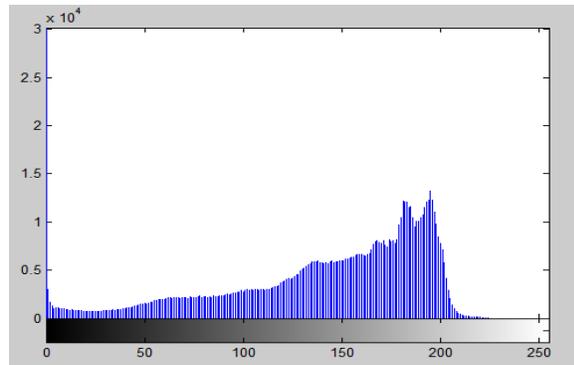
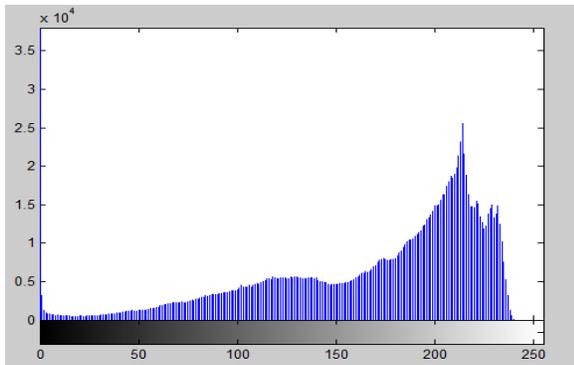


Figure 3: The grayscale intensity distribution of typical normal lung and the TB lung.



(a)



(b)

Figure 4: Two infected TB lungs (left).

Conclusion

This investigation has demonstrated the feasibility in the establishment of visually meaningful and correlated features in the attempt to discriminate normal and abnormal (TB) radiographs. Statistically, the TB infected radiographs offer higher entropy and a lower mean value in their intensity distribution as compared to their normal chest radiographs. TB infected chest radiograph seemed to exhibit greater disparities between left and right on every feature values. Figure 3 shows the histogram plot of grayscale intensity distribution of a typical normal lung and the typical TB lung. It is apparent that, the intensity profile of normal lung approximated normal distribution. And the distribution width was

narrow. On the other hand, TB in usual had a larger spread, with the distribution positively or rightly skewed. Such characteristics indicated that on chest radiograph, TB infected lung had wider range in grayscale value, relatively concentrated at the lower end of the distribution, as a consequence of the hazy presence mostly observed in the upper lobe. The TB lung in figure 4(b) was deemed hazier on the upper lobe as compared to the figure 4(a) as shown.

References

- [1] Abdel-Halim Hafez Elamy, Mrinal Mandal, Behrouz Far, Anup Basu, Irene Cheng, Richard Long, 978-1-4244-5377-1/10/\$26.00 ©2010 IEEE, "An Intelligent CAD System For Automated Detection of Pulmonary Tuberculosis on Chest Radiograph and CT Thorax".
- [2] Arun Chauhan¹, Devesh Chauhan², Chittaranjan Rout, Plos one 9(11):e112980.doi:10.1371/journal.pone.0112980, "Role of Gist and PHOG Features in Computer-Aided Diagnosis of Tuberculosis without Segmentation".
- [3] Acharya UR, Dua S, Du X, Sree VS, Chua KC, IEEE Transactions on Information Technology in Biomedicine, 15(3), 2011, 449-455, "Automated Diagnosis of Glaucoma Using Texture and Higher Order Spectra Features".
- [4] Bram van Ginneken, Shigehiko Katsuragawa, Bart M. Haar Romeny, Kunio Doi, and Max A. Viergever, IEEE Transaction on Medical Imaging, Vol.21, NO.2, February "Automatic Detection of Abnormalities in Chest Radiographs Using Local Texture Analysis".
- [5] Michael Kass, Andrew Witkin and Demetri Terzopoulos, C2465-3/87/0000/025901.00, 1987 IEEE "Snakes: Active Contour Model".

- [6] Poonam Panwar, Neeru Gulati, Journal of Global Research in Computer Science, Volume 4, No. 1, January 2013 "Genetic Algorithms for Image Segmentation Using Active Contours".
- [7] TB Alliance, "2010 Annual Report," 2010.
- [8] Tan JH, Ng EYK Acharya UR, Infrared Physics & Technology 2009; 53:120-126, "Study of normal ocular thermogram using textural parameters".
- [9] Ulas Bagc, Mike Bray, JesusCaban, Jianhua Yao, Daniel J. Mollura, Computerized Medical Imaging and Graphics 36 (2012) 72– 84, "Computer assisted detection of infectious lung diseases".
- [10] Wai Yan NyeinNaing, Zaw Z. Htike, Sheroz Khan, Amir A. Shafie, Electrical & Computer Engineering: An International Journal (ECIJ) Volume 3, Number 4, December 2014, "Computer Aided Tuberculosis Detection in Chest Radiographs".