Journal of Biomedical and Bioengineering,
ISSN: 0976-8084 & E-ISSN: 0976-8092 , Volume 3, Issue 1, 2012, pp.-53-56.
Available online at http://www.bioinfo.in/contents.php?id=87

TEXTURE ANALYSIS OF TB X-RAY IMAGES USING IMAGE PROCESSING TECHNIQUES

PATIL S.A.

Textile & Engineering Institute, Ichalkaranji, India
*Corresponding Author: Email- shrinivasapatil@yahoo.co.in

Received: January 16, 2012; Accepted: March 09, 2012

Abstract- Tuberculosis (TB) was the leading cause of death for all age groups in the Western world from that period until the early 20th century. Finding an effective cure for tuberculosis represents one of the great advances in 20th-century medicine. Fifty years after *Robert Koch*, in 1882, discovered the tubercle bacillus; several antimicrobial drugs were discovered that can cure the disease. TB infection appears on X-ray film as a diffuse opacity representing a patch of consolidation in the lung field. TB infection mostly affects the posterior segment of the upper lobe. On the posterior-anterior (PA) X-ray film it appears as an area of shadowing near the lung apex. A computer algorithm for texture analysis of TB chest radiograph is presented. Algorithm includes important steps, like image acquisition, image pre-processing, lung field segmentation, and features extraction. Total 49 images are used during experiment to estimate 1st and 2nd order texture features. Gray Level Co-occurrence Matrix (GLCM) technique is used to estimate texture features.
Key words- TB, X-ray images, chest radiographs, image processing.

Citation: Patil S.A. (2012) Texture analysis of TB x-ray images using image processing techniques. Journal of Biomedical and Bioengineering, ISSN: 0976-8084 & E-ISSN: 0976-8092, Volume 3, Issue 1, pp.-53-56.

Introduction

Tuberculosis (TB), though no longer the ubiquitous and sinister threat of past decades, is still a problem that must be borne in mind, particularly with immigrants from developing countries and in the immuno-suppressed patients. The primary form of TB infection, which used to be seen almost exclusively in children, is now also being seen in older patients. The primary TB infection appears on X-ray as a diffuse opacity [1] representing a patch of consolidation in the lung field with increased striations extending towards the hilum where the enlarged glands show as rounded opacities. Pleural effusion is also a common manifestation of primary TB infection. The secondary or adult type of TB infection mostly affects the posterior segment of the upper lobe. On the posterior-anterior (PA) film it appears as an area of shadowing near the lung apex often mottled in character. Fig. 1 (a) gives an idea of the spatial distribution of the abnormal areas within the chest radiographs in the TB database. In the vast majority of patients diagnosis of pulmonary TB can be made with confidence by radiological examination of the chest and examination of the sputum. In some patients it is necessary to further radiological examination after a course of treatment with an antibiotic, in order to exclude an acute inflammatory cause for an abnormal radiograph shadow.

Morphology of Tuberculosis

The World Health Organization (WHO) has declared tuberculosis a global emergency. TB is estimated to kill three million people each year. Yet most people infected with the germ, Mycobacterium tuberculosis, do not suffer from TB. Instead the germ becomes trapped in body tissues, sealed up in a calcified nodule, or tubercle. On a radiograph such a nodule can produce a tiny dense shadow at the edge of the lung field, often accompanied by a shadow representing lymph nodes in the center of the field. These two shadows represent the primary complex of tuberculosis and

such an infection is common in children, who often do not suffer from illness and may develop immunity to TB. In adults, TB usually takes a different form.

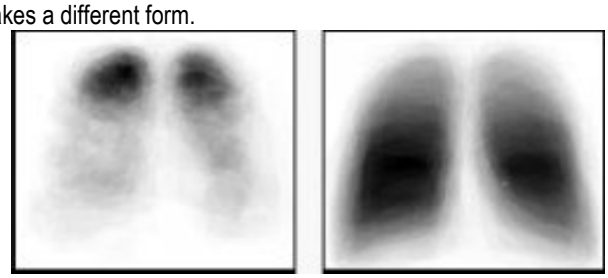


Fig. 1- (a) Distribution of abnormal areas for the TB database, (b) Distribution of abnormal areas for the interstitial disease (ID) database

The bacillus spreads widely in the lungs and forms large cheesy masses that break down the respiratory tissues and develop cavities in the lungs. Once the disease process eats its way into the bronchi, the bacilli can be coughed or breathed out, making TB highly contagious. A blood vessel can be eroded and then the patient coughs up blood. This spreading type of the disease was popularly known as consumption, a name that vividly describes TB's destructive progress. If the patient is treated with drugs, or builds up resistance, the bacilli may be sealed up in the lungs, which by then contain much scar tissue. Consequently, the disease can, in its various stages, give rise to a large variety of signs on an X-ray. Sample X-ray image of TB is included in Fig. 2 [1]. Accounts of tuberculosis can be found as far back as the writings of the ancient Egyptians. TB was the leading cause of death for all age groups in the Western world from that period until the early 20th century. Finding an effective cure for tuberculosis represents one of the great advances in 20th-century medicine. Fifty years after *Robert Koch*, in 1882, discovered the tubercle bacillus; several antimicrobial drugs were discovered that can cure the disease.

Here the approach is to divide the separated lung fields in four different parts and analyze each part separately, with texture features extracted solely from these parts. In this way, the classifier should capture knowledge regarding the normal variation within that particular part. The avg. gray level, standard deviation (second moment), skew (third moment), uniformity, entropy etc., of each filtered image are computed as texture features.

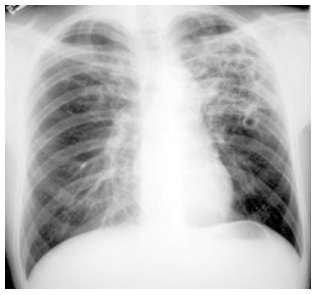


Fig. 2- X-ray image of Tuberculosis

Material and method

The steps followed for analysis and feature extraction from TB X-ray images are described in this section. As an initial step the images are obtained using image acquisition method and then the application of pre-processing algorithms including size normalization and filtering of the image is carried out. The features those are identified to be useful for diagnosis and analysis require separation of the lung fields from the background. Lung field masks are prepared to separate lung fields by estimating peripheral coordinates of lung fields manually, and using region based segmentation technique. Every TB image (images are collected from public database) data is acquired with 256 gray levels (8 bits) and stored as JPEG (.jpg, .jpeg) data. Before extraction of the features from an image, it is necessary to pre-process the image to reduce irrelevant information or noise, and to enhance the image properties, which makes the feature measurement easier, and more reliable. Scanned images are resized to a size of 512 X 512 pixels. Median filter is used to remove the noise or irrelevant information from the images.

An important approach for describing a region is to quantify its texture content [2]. A frequently used approach used for texture analysis is based on statistical properties of the intensity histogram. One class of such measures is based on statistical moments. An expression for the *n*th moment about the mean is given by:

$$\mu_n = \sum_{i=0}^{L-1} (z_i - m)^n \, p(z_i)$$

where z_i is a random variable indicating intensity levels in an image, $p(z)$ is the histogram of the intensity levels in a region, L is the possible intensity levels. A histogram component $p(z_j)$, is an estimate of the probability of occurrence of intensity value z_j, and the histogram may be viewed as an approximation of the probability density function (PDF). GLCM is the technique used to calculate PDF.

$$m = \sum_{i=0}^{L-1} z_i \, p(z_i)$$

Here m is the mean (average) intensity. These moments can be computed using MATLAB function [3] *statmoments*, which is acting as a sub-function in another MATLAB function [3] known as *statxture*. This function is used to calculate first-order statistic texture features like mean, standard deviation, smoothness, third moment, uniformity and entropy.

A measure of average contrast or the standard deviation can be calculated by using following equation, where $\mu_2(z)$ is the second moment.

$$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2}$$

Smoothness measures the relative smoothness of intensity in a region. R is 0 for a region of constant intensity and approaches 1 for region with large excursions in the values of its intensity levels. Smoothness is calculated by using the following equation.

$$R = 1 - 1/(1 + \sigma^2)$$

Skewness of the histogram is also known as third moment. This measure is 0 for symmetric histograms, positive by histograms skewed to the right (about the mean) and negative for histograms skewed to the left. For smooth images this value comes to be negative. Following equation is used to calculate third moment.

$$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 \, p(z_i)$$

When all gray levels are equal, uniformity measures maximum and goes on decreasing from there for the inequality.

$$U = \sum_{i=0}^{L-1} p^2(z_i)$$

Entropy is nothing but the measure of randomness, given by the following equation.

$$e = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$$

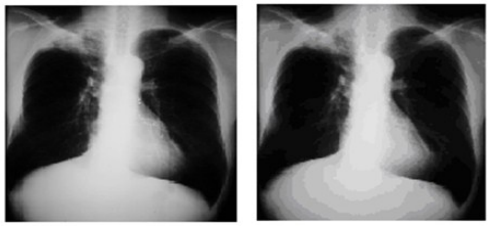
The GLCM [4] functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image. However, a single GLCM might not be enough to describe the textural features of the input image. For example, a single horizontal offset might not be sensitive to texture with a vertical orientation. Therefore it is essential to generate multiple GLCMs with different offset values or at different angles. MATLAB function [3] *graycomatrix* is used to generate such multiple GLCMs. Using multiple GLCMs, second-order statistic features like, Contrast Correlation, Energy, and Homogeneity are estimated.

TB image analysis and results

Chest radiographs have always played an important role in differential diagnosis and determining the extent of tuberculosis, although they are not the only way to diagnose the disease. A problem is that no radiological sign is unique to TB and that an abnormality does not prove the presence of active disease.

Despite the existence of a cheap cure, TB is a leading killer of adults in the world. Mass chest screening can identify cases of active TB to fight the epidemic. The advent of digital chest units could facilitate the application of computer-aided diagnosis to improve the efficiency of mass chest screening. TB may reveal itself in many different radiographic patterns, but in most cases a chest radiograph of a patient with TB contains areas with diffuse abnormalities. In this work we propose a scheme to detect such textural abnormalities.

A main problem in the texture analysis of chest radiographs is the complex "background" of superimposed normal anatomical structures to which the analysis must be somehow insensitive. One way to solve this problem would be to restrict texture analysis to regions of interest. An alternative could be to pre-process the images so as to remove normal background structures. Here the approach is to divide the separated lung fields in parts (4 here) and analyze each part separately, with texture features extracted solely from these parts. In this way, the classifier should capture knowledge regarding the normal variation within that particular part. The avg. gray level, standard deviation (second moment), skew (third moment), uniformity, entropy etc., of each filtered image are computed as texture features.



(a) (b)

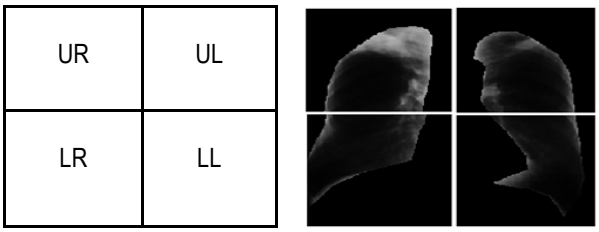
Fig. 3- (a) Original TB image, (b) Filtered TB image

The Fig. 3 (a) and (b) illustrates the original X-ray image of TB and its filtered version respectively. Further the lung field masks have been prepared using region based segmentation technique to separate the lung fields from the other background information. Masks are multiplied with the filtered images to separate the lung fields. Fig. 4 (a) depicts the mask image prepared from the original image shown in Fig. 3 (a). Fig. 4 (b) shows the separated lung fields.

(a) (b)

Fig. 4- (a) Lung field mask, (b) Separated lung fields

According to the morphology of TB, TB infection mostly affects the posterior segment of upper lobes. Therefore for the analysis the separated lung field image is divided in to 4 sections like upper right (UR), upper left (UL), lower right (LR), and lower left (LL) as shown in Fig. 5. Then using Gray Level Co-occurrence Matrix (GLCM) technique the 1st and 2nd order statistic features are estimated for each section.



(a) (b)

Fig. 5- (a) Image divided into 4 sections, (b) Lung fields image divided into 4 sections

The 1st order statistic feature values for the image shown in Fig. 5 (b) are included in the Table 1.

Table 1- 1st order statistic feature values for TB image

	UR	UL	LR	LL
Avg. gray level	96.231	81.723	73.587	25.284
Std. deviation	84.279	89.796	78.815	46.871
Smoothness	0.0985	0.1103	0.0872	0.0327
Third moment	0.3062	3.1721	3.1139	2.674
Uniformity	0.1596	0.2843	0.2423	0.517
Entropy	5.343	4.2395	4.7598	2.931

Tables 2 to 5 depict the 2nd order statistic feature values for 4 sections, for the image included in Fig. 5 (b) respectively.

Table 2- 2nd order statistic feature values for UR section

Second order statistic contrast features	For offset [0 1]	For offset [-1 1]	For offset [-1 0]	For offset [-1 -1]	Avg. value
Contrast	0.392	0.435	0.202	0.489	0.379
Correlatio-n	0.964	0.959	0.981	0.955	0.965
Energy	0.223	0.218	0.225	0.216	0.220
Homogen-eity	0.971	0.960	0.971	0.956	0.964

Table 3- 2nd order statistic feature values for UL section

Second order statistic contrast features	For offset [0 1]	For offset [-1 1]	For offset [-1 0]	For offset [-1 -1]	Avg. value
Contrast	0.351	0.514	0.232	0.315	0.353
Correlation	0.971	0.958	0.981	0.974	0.971
Energy	0.332	0.324	0.334	0.328	0.329
Homog-eneity	0.973	0.960	0.973	0.966	0.968

Table 4- 2nd order statistic feature values for LR section

Second order statistic Contrast features	For offset [0 1]	For offset [-1 1]	For offset [-1 0]	For offset [-1 -1]	Avg. value
Contrast	0.179	0.234	0.070	0.149	0.164
Correlation	0.947	0.930	0.979	0.955	0.931
Energy	0.550	0.545	0.557	0.552	0.537
Homogeneity	0.976	0.965	0.982	0.977	0.971

Table 5- 2nd order statistic feature values for LL section

Second order statistic Contrast features	For offset [0 1]	For offset [-1 1]	For offset [-1 0]	For offset [-1 -1]	Avg. value
Contrast	0.312	0.346	0.177	0.395	0.292
Correlation	0.967	0.963	0.987	0.958	0.969
Energy	0.281	0.288	0.287	0.275	0.280
Homogeneity	0.969	0.969	0.978	0.957	0.968

Table 6- 1st order statistic feature values for TB images

Saples	Avg. gray level	Std. devia-tion	Smoot-hness	Third mom-ent	Unifo-rmity	Entr-opy
TB-1	81.72	89.79	0.11	3.17	0.28	4.24
TB-2	96.23	84.28	0.09	0.31	0.16	5.34
TB-3	73.59	78.81	0.09	3.11	0.24	4.76
TB-4	77.19	73.48	0.08	1.69	0.17	5.39
TB-5	68.38	70.83	0.07	2.85	0.19	5.20
TB-6	40.65	52.54	0.04	2.31	0.32	3.94
TB-7	51.57	54.83	0.04	1.62	0.21	4.66
TB-8	50.62	64.73	0.06	3.81	0.30	4.42
TB-9	76.12	74.64	0.06	2.96	0.27	4.98
TB-10	65.98	80.25	0.09	4.96	0.30	4.39

Table 7- 2nd order statistic feature values (avg.) for TB images

Samples	Contrast	Correlation	Energy	Homogeneity
TB-1	0.432	0.972	0.335	0.961
TB-2	0.354	0.946	0.552	0.972
TB-3	0.368	0.968	0.281	0.963
TB-4	0.401	0.965	0.210	0.951
TB-5	0.453	0.962	0.238	0.963
TB-6	0.215	0.963	0.375	0.980
TB-7	0.198	0.965	0.303	0.973
TB-8	0.294	0.958	0.335	0.956
TB-9	0.342	0.968	0.463	0.973
TB-10	0.301	0.962	0.332	0.970

References

[1] Van Ginneken B. (2001) *PhD Thesis*.
[2] Wei J., Hagihara Y., Shimizu A., Kobatake H. (2002) *Cars*.
[3] Gonzalez R.C., Woods R.E., Eddins S.I. (2004) *Pearson Education*.
[4] MATLAB Handbook, *Mathworks Ltd*.