Proceedings of 2013 ICME
International Conference on Complex Medical Engineering
May 25 - 28, Beijing, China

Lung Nodule Detection Using Multi-Resolution Analysis

[1]Mickias Assefa, [2]Ibrahima Faye, [3]Aamir Saeed Malik
Centre for Intelligent Signal and Imaging Research,
[2]Department of Fundamental and Applied Sciences
[1,2]Departement of Electrical and Electronic Engineering
University Technology PETRONAS
Tronoh, 31750, Malaysia
mikeutp@gmail.com, ibrahima_faye@petronas.com.my

Muhammad Shoaib
Department of Mathematical Sciences
University of Hail
PO Box 2440, Saudi Arabia

***Abstract* - The increase in number of smokers in the 20th century leads for lung cancer being the number one killer among cancer related deaths world-wide with an over-all survival rate of only 15%. According to the latest World Health Organization report released on April 2011, Lung cancer is the number one killer among deaths related with cancer for men in Malaysia. In most cases there are no symptoms or warning signs in the early stages of lung cancer. It is often suspected initially from chest x-ray or Computed Tomography (CT) scans done to evaluate a cough or chest pain. Several research work have been aimed at development of Computer Aided Diagnosis (CAD) systems which enhance early detection of lung cancer nodules. This research work aims to develop a CAD system to detect pulmonary lung nodules from Low Dose CT (LDCT) scan images using template matching algorithm integrated with multi-resolution feature analysis technique in order to enhance the false positive detection rate. 134 out of 165 nodules were correctly detected by our scheme. That results in a detection rate of 81.212%.**

Index Terms* - *Pulmonary nodule, template matching Multiresolution Analysis, Computer Aided Diagnosis

I. INTRODUCTION

In recent years Computer-Aided Diagnosis (CAD) has become a major area of interest in diagnostic radiology. CAD algorithms utilize a combination of one or more techniques of statistics, computer science and/or applied physics and in order to assist the radiologist in medical image interpretation. CAD system usually involves four steps; image pre-processing, Region of Interest (ROI) definition, feature extraction and analysis and finally clustering or classification.

Lung cancer can be defined as a cancer that originates in tissues of the lung, usually in the cells lining air passages. The two main types are small cell carcinoma (SCLC) and non-small cell carcinoma (NSCLC). These types are distinguished by the appearance of the cancer cells under a microscope. Detection of suspicious nodules at the early stage is considered the most effective way to improve the survival rate of lung cancer patients.

Computed Tomography (CT) scans enables radiologists to visualize small and/or low contrast nodules that are nearly impossible to visualize using conventional radiograms. However, a single CT scan will produce a large amount of images that makes it difficult for the radiologist to analyze manually. In addition to that, there are many objects in a serial CT scan images that have the same appearance and Hounsfield Unit (HU) value as pulmonary nodules. CAD systems are being proposed in order to assist radiologists; to be used as a tool to provide a “second opinion”.

Various CAD systems were proposed to detect lung nodules. Sheng et al. proposed a scheme for detection of pulmonary nodules mainly focusing on improving the detection rate of those nodules that overlap with ribs and clavicles using Mass Training Artificial Neural Networks Virtual Dual Energy [1]. GuoXiuhua et al. extracted fourteen texture features based on curvelet transform for lung nodule classification using support vector machine [2]. Ozekeset al. reported Region of Interest (ROI) detection scheme using the density values of pixels in CT images and scanning the pixels in eight directions by using various thresholds [3]. Jia et al. has developed a scheme to detect pulmonary nodules from Hessian matrix constructed from the 2nd derivate of the 3-D image function [4].

Several researchers have used template matching for lung nodule detection. Farag et al. developed active appearance model based templates for lung nodule template matching [5]. Lee [6] has proposed a genetic algorithm based template matching scheme for detection of pulmonary nodules. Dehmeshki et al. has also applied a genetic algorithm template matching scheme with spherical elements for the detection of nodules [7]. Farag et al. developed lung nodule templates using the texture and shape properties of real nodules [8].

The initial nodule detection stage results in a significant number of false positive candidates which necessitates development of false positive reduction scheme. Intensity based features were extracted and used for false positive reduction in [6], [4] and [9]. Memarian et al. proposed hybrid schemes which consist of fuzzy c-means clustering and iterative linear discriminant analysis for false positive reduction [10]. Takemura et al. has applied a logical *AND* operator between consecutive CT slices for false positive reduction [11].

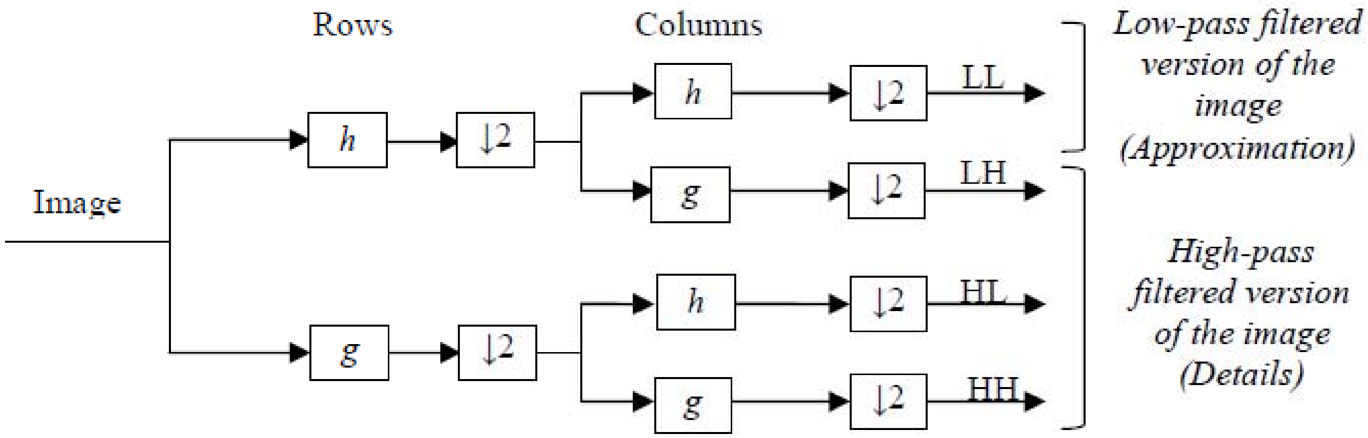

Fig. 1 A Wavelet Decomposition of an Image

In this research work, a multi-resolution based scheme for detection of lung nodules from low-dose CT scans is proposed. The scheme consists of two stages. In the first stage nodular models are developed and used as a reference. Then Multi-resolution and intensity based features were extracted for false positive reduction and improvement of the detection rate.

II. DETECTION SCHEME

Noise removal was done as an initial step on the entire dataset LDCT scans. The performance of the noise filter is evaluated in terms of maintaining the lung nodules, since in some cases averging filters may result in complete or partial removal of important details or alter the location of the lung nodules. A weiner filter with the appropriate filter parameters were designed and applied on the ground truth data set with pre-identifed nodules (Fig. 2A). It has been observed that an adaptive Wiener filtering on 7x7 or 9x9 blocks have given a good performance (Fig. 2B).

Pulmonary nodules in general have spherical structures but they can be perplexed by anatomical structures surrounding it. In this research work, pulmonary lung nodules were classified regardless of medical classification, only depending on their location.

Small nodules appear only on a single slice and relatively larger nodules appear on consecutive slices depending on their size. To detect pulmonary nodules in 2-D slices circular and semi-circular templates were developed. The circular templates are used to detect nodules inside the lung region and the semi-circular nodules are used to detect nodules at the lung boundary.

Since the pulmonary nodules appear as circular structure on LDCT 2-D scans, a circular structure detection scheme is employed as an initial pre-processing step (Fig. 3). This stage helps to optimize the computational time for the template matching process for nodules residing inside the lung.


(A)

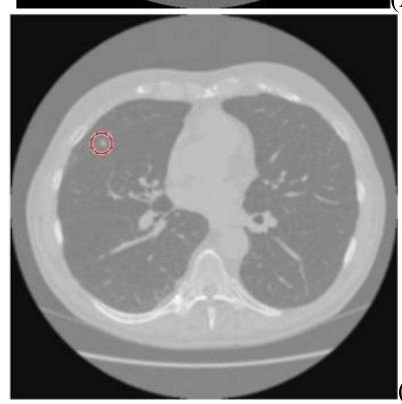
(B)

Fig. 2 Nodule Visibility Test Before (A) and After (B) Applying Wiener Filter

By looking at the intensity HU values in and around the nodules we have observed the nodules intensity values followed a Gaussian-Like distribution. Similar observation has been made by [6] and [8]. As a result, templates with Gaussian-like intensity distribution were developed. The nodular models are obtained by the following formula [8].

$$q(r) = q_{max} e^{-(r/\rho)^2}, 0 \le r \le R \quad (1)$$

$$\rho = R\,(\ln(q_{max}) - \ln(q_{min}))^{-1/2} \quad (2)$$

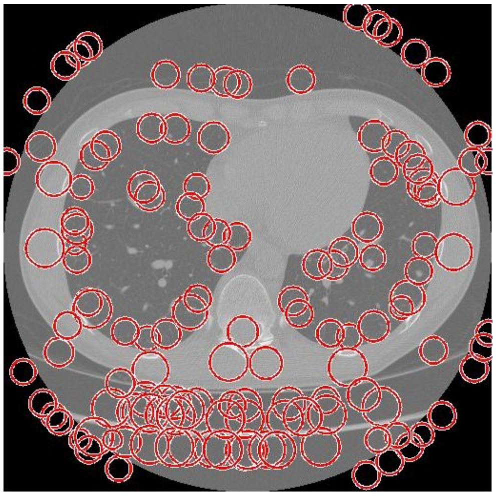

Fig. 3 Result of circular object detection

Where 'r' stands for the distance from the centroid of the nodule and the range q_{min} and q_{max} of the pulmonary nodule intensity distribution values are obtained from the ground truth data.

Normalized cross correlation (NCC) is used as a similarity measure. Let $f(x,y)$be an intensity value of an image and *t(x,y)* be a given template, the normalized cross-correlation can be defined as follows [12] :-

$$NCC = \frac{\sum_{x,y}(f(x,y)-\bar{f}_{u,v})(t(x-u,y-v)-\bar{t})}{\sqrt{\sum_{x,y}(f(x,y)-\bar{f}_{u,v})^2 \sum_{x,y}(t(x-u,y-v)-\bar{t})^2}} \quad (3)$$

$$\bar{f}_{u,v} = \frac{1}{N_x N_y} \sum_{x=u}^{u+N_{x-1}} \sum_{y=v}^{v+N_y-1} f(x,y) \quad (4)$$

Where $\bar{f}_{u,v}$ denotes the mean value of $f(x,y)$ within the area of the template t sifted to (u,v). $\bar{t}$ denotes the mean value of the template image.

III. FALSE POSITIVE REDUCTION

Our scheme detected several false positive nodules per scan in the first stage. As a result two intensity based and seven multi-resolution based features were extracted for reduction of false positive nodules. Details of the features used are described below.

Mean and standard deviation: Mean is a measure of brightness, where $p_{r,s}$ is pixel at location (r,s). Standard deviation is a measure of contrast. These two features are used to differentiate between parts of the bone and skin from the true positive nodules.

$$\mu_p = \frac{1}{n^2} \sum_{r=0}^{n-1} \sum_{s=0}^{n-1} p_{r,s} \quad (5)$$

$$\sigma_p = \left[\frac{1}{n^2} \sum_{r=0}^{n-1} \sum_{s=0}^{n-1} \left(p_{r,s} - \mu_p\right)^2\right]^{1/2} \quad (6)$$

Wavelets are mathematical functions that segment a signal or an image into different frequency components and thus allow us to analyze local properties of a signal or an image. The wavelet transform is therefore useful in analyzing signals and images with large discontinuities and non-periodic signals. The wavelet model for two-dimensional signals can be defined using multi-resolution approximation of $L^2(R^2)$ with one separable two-dimensional scaling function $\varphi(x,y) = \varphi(x)\varphi(y)$ and three separable 2-D wavelets $\psi^H(x,y) = \psi(x)\psi(y)$, $\psi^V(x,y) = \psi(x)\psi(y)$ and $\psi^D(x,y) = \psi(x)\psi(y)$ [13]. $\psi^H(x,y), \psi^V(x,y)$ *and* $\psi^D(x,y)$ are called horizontal, vertical and diagonal wavelets respectively (Figure 1). The separability of the kernels simplifies the computation of the two-dimensional transform by allowing row-column or column-row passes of one-dimensional transform [14].

Daubechies-8 (db8) and symlet wavelet functions were used with a single level decomposition for the multi-resolution based false positive reduction. Then seven statistical features are extracted from the wavelet transform coefficients. The features are as follows:-

$$Energy = \sum_i^M \sum_J^N I^2[i,j] \quad (7)$$

$$Entropy = -\sum_i^M \sum_j^N I[i,j] \log I[i,j] \quad (8)$$

$$Mean = \frac{1}{M*N} \sum_i^M \sum_J^N I[i,j] \quad (9)$$

$$STD = \sqrt{(I[i,j] - \bar{I}[i,j])^2} \quad (10)$$

$$Max\ Probablity = \max I[i,j] \quad (11)$$

$$Inverse\ Difference\ Moment = \sum_i^M \sum_J^N \frac{I[i,j]}{|i-j|^2} \quad where\ i \neq j \quad (12)$$

$$Homogeneity = \sum_i^M \sum_J^N \frac{I[i,j]}{1+|i-j|^2} \quad (13)$$

IV.

V. EXPERIMENTAL RESULTS

This work is based on the Early Lung Cancer Action Program (ELCAP) public database [15], which contains 50 sets of low-dose CT lung scans taken at a single breath-hold with 1.25 mm slice thickness. The ELCAP database has a resolution 0.5x0.5 mm. The locations of the ground truth nodules detected by an experienced radiologists are also provided. In this research work, a subset database containing 165 nodules is used.

The template matching algorithm at stage one, extracted 139 true positive nodules (out of 165 nodules) and 496 False Positive Nodules. The classification after False Positive Reduction stage reduced the number of False Positive Nodules to 58 but at the same time rejected five true positive nodules. Thus the final detection rate becomes 81.212% with detection of 134 true positive nodules out of 165. The overall False Positive rate is 35.15%.

Majority of the missed nodules consist of nodules that have low contrast and nodules that are attached to the mediastinum region. The low contrast nodules exhibit an intensity distribution that is significantly different from the templates used. We have also noticed it is extremely difficult to detect nodules that are attached to the lung vessels because of the anatomical similarity between the vessels and the nodules on LDCT scan images. Lung vessels constitute the majority of the false positive candidates. The number of false positive may decrease by analysing consecutive CT slices before and after the false positive candidate.

VI. CONCLUSION

We proposed a nodule detection scheme based on template matching and multi-resolution based false reduction scheme. A novel circular object detection scheme is applied as a pre-processing step to identify potential nodule candidate and optimize the template matching process. Nodular models with a Gaussian distribution are used as a reference image. Intensity based and multi-resolution based features were extracted for the false positive reduction stage. Our system performed at a rate of approximately 81%. It was observed that lung vessels make up the majority of the false positive candidates, which results in a relatively high false positive rate. We are working on using additional multi resolution features in order to tackle this problem.

VII. ACKNOWLEDGMENT

The authors gratefully acknowledge Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme (FRGS, Grant No: 15-8200-191) and Universiti Teknologi Petronas for providing research facilities.

The author M. Shoaib would like to thank the Deanship of Scientific Research at the University of Hail, Saudi Arabia for financial support under the project SM2 for the year 1433 Hijri (2012-2013).

VIII. REFERENCES

[1] S. Chen and K. Suzuki, "Computerized Detection of Lung Nodules by Means of “Virtual Dual-Energy” Radiography," 2012.

[2] S. T. Guo Xiuhua, Wu Haifeng, He Wen, Liang Zhigang,Zhang Mengxia,Guo Aimin1 and Wang Wei, "Support Vector Machine Prediction Model of Early-stage Lung Cancer Based on Curvelet. Transform to Extract Texture Features of CT Image," *World Academy of Science, Engineering and Technology,* vol. 17, p. 71, 2010.

[3] S. Ozekes and A. Y. Camurcu, "Automatic Lung Nodule Detection Using Template Matching," in *Advances in Information Systems*. vol. 4243, T. Yakhno and E. Neuhold, Eds., ed: Springer Berlin Heidelberg, 2006, pp. 247-253.

[4] T. Jia*, et al.*, "A lung cancer lesions dectection scheme based on CT image," in *Signal Processing Systems (ICSPS), 2010 2nd International Conference on*, 2010, pp. V1-557-V1-560.

[5] A. Farag*, et al.*, "An AAM-based detection approach of lung nodules from LDCT scans," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, 2012, pp. 1040-1043.

[6] L. Yongbum*, et al.*, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *Medical Imaging, IEEE Transactions on,* vol. 20, pp. 595-604, 2001.

[7] J. Dehmeshki*, et al.*, "Automated detection of lung nodules in CT images using shape-based genetic algorithm," *Computerized Medical Imaging and Graphics,* vol. 31, pp. 408-417, 2007.

[8] A. A. Farag*, et al.*, "Data-Driven Lung Nodule Models for Robust Nodule Detection in Chest CT," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 2588-2591.

[9] S. K. V. Anand, "Segmentation coupled textural feature classification for lung tumor prediction," in *Communication Control and Computing Technologies (ICCCCT), 2010 IEEE International Conference on*, 2010, pp. 518-524.

[10] N. Memarian*, et al.*, "Computerized Detection of Lung Nodules with an Enhanced False Positive Reduction Scheme," in *Image Processing, 2006 IEEE International Conference on*, 2006, pp. 1921-1924.

[11] S. Takemura*, et al.*, "Enhancement and detection of lung nodules with Multiscale filters in CT images," in *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP '08 International Conference on*, 2008, pp. 717-720.

[12] K. Briechle and U. D. Hanebeck, "Template matching using fast normalized cross correlation," pp. 95-102, 2001.

[13] I. Faye*, et al.*, "Digital Mammograms Classification Using a Wavelet Based Feature Extraction Method," in *Computer and Electrical Engineering, 2009. ICCEE '09. Second International Conference on*, 2009, pp. 318-322.

[14] I. Faye, "A Random Feature Selection Method for Classification of Mammogram Images," in *Intelligent Systems, Modelling and Simulation (ISMS), 2012*

Third International Conference on, 2012, pp. 330-333.
[15] *ELCAP public lung image database*. Available: http://www.via.cornell.edu/lungdb.html