April 2014

COMPUTER AIDED ANALYSIS OF LUNG CT BASED ON TRANSFORM DOMAIN FEATURES

V. MINNAL

Department of Information Technology, MIT, Anna University, Chennai, India, minnal.v@gmail.com

Follow this and additional works at: <https://www.interscience.in/ijcsi>

Part of the [Computer Engineering Commons](#), [Information Security Commons](#), and the [Systems and Communications Commons](#)

Recommended Citation

MINNAL, V. (2014) "COMPUTER AIDED ANALYSIS OF LUNG CT BASED ON TRANSFORM DOMAIN FEATURES," *International Journal of Computer Science and Informatics*: Vol. 3 : Iss. 4 , Article 5.

DOI: 10.47893/IJCSI.2014.1153

Available at: <https://www.interscience.in/ijcsi/vol3/iss4/5>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer Science and Informatics by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

COMPUTER AIDED ANALYSIS OF LUNG CT BASED ON TRANSFORM DOMAIN FEATURES

V.MINNAL

Department of Information Technology, MIT, Anna University, Chennai, India
E-mail : minnal.v@gmail.com

Abstract - As Many CADx systems have been developed to detect lung cancer based on spatial domain features that process only the pixel intensity values, the proposed scheme applies frequency transform to the lung images to extract frequency domain features and they are combined with spatial features so that the features that are not revealed in spatial domain will be extracted and the classification performance can be tuned up. The proposed CADx comprises of four stages. In the first stage, lung region is segmented using Convexity based active contour segmentation. At second stage ROIs are extracted using spatially constrained KFCM clustering. Followed by standard wavelet transforms is applied on ROI so that transform domain features are extracted with shape and haralick histogram features. Finally neural network is trained by combined feature set to identify the cancerous nodules. Our proposed scheme has shown sensitivity of 95% and specificity of 96%.

Keywords - convexity; active contour; KFCM; wavelet features;neural network.

I. INTRODUCTION

Recently computer programming plays a vital role in solving most of the problems in medical applications such as health monitoring, intensivesurgery due to their fastness and accuracy. Especially image processing techniques are considered to be mostly preferred in diagnosing abnormalities on medical images like X-ray, CT. Computer Aided Diagnosis is the procedure that involves imaging technologies followed by the image processing techniques where as the various images of human body obtained by imaging technologies are processed to identify the illness or abnormalities if present. If CADx developed devoid of drawbacks then it would be a perfect second opinion to the radiologists in medical image interpretation. In last two decades research on development of CADx has grown exponentially, and review of that survey is shown in next section.

As various technologies emerging, several works have been done to identify lung cancer from CT images. Lung cancer diagnosis has been an active research domain since 1980's. In literature, very recent advanced works in diagnosing lung cancer has been reviewed. The development of new schemes for image acquisition, such as high resolution CTs, has improved the detection and diagnosis of cancer significantly.

Even if the research on automated computer aided diagnosis of lung cancer has been started since 1990's list of approaches have been proposed since 2006. In 2006, Antonelli *et al* [1] have proposed Lung Nodule Detection scheme based on an Anatomical Model and a Fuzzy Neural Network. He used anatomical model of lung to segment lung parenchyma from the CT image and fuzzy Neural Network to classify by using the fuzzy set of features. Adrien *et al* [2] in 2007 have

done a scheme of Lung Tissue Classification Using Wavelet Frames which has achieved multiclass accuracy of 92.5%. In 008 Liu lu *et al* have proposed to use SVM to detect the pulmonary nodules with the sensitivity of 90% and specificity of 70%. During 2009 jaffar *et al* put forward Fuzzy morphology and FCM classification to identify nodules from CT images. In 2010, lafari *et al* has proposed Automated Detection of Pulmonary Nodules in 3D Thoracic CT Images with the FP of 10.3 and sensitivity of 88%.

In same era Zhang *et al* [] have published the diagnosis system of lung cancer by using different classes of features by the sensitivity of 99% and specificity of 86%.those classes of features included 1st, 2nd and higher order features along with the structural features. Tong *et al* have proposed nodule detection scheme based on Rule Based Classification that achieved Overall detection rate 85%. In 2011, Hua *et al* [3] have developed Graph-Search Algorithm to segment the lung parenchyma and KNN classifier for lung tissue classification which have given Sensitivity of 98.6% and Specificity of 99.8%. In 2011, ying *et al* has proposed Autonomous Detection of Solitary Pulmonary Nodules in which Quantification of feature parameter is done using SVM with the 94.6% of sensitivity. In same year, Aravind Kumar *et al* have developed Robust and Automated Lung Nodule Diagnosis from CT Images based on fuzzy inference system (FIS). This achieved Classification accuracy of 90% and sensitivity of 86%.

In this proposed approach, CADx consisting of four phases as follows. The lung parenchyma is segmented from background by using convexity based Active Contour segmentation. ROIs are then identified and extracted using spatially constrained Kernel FCM clustering. Followed by shape based and histogram

features are extracted from ROIs. Transform domain features are extracted by applying daubechies wavelets which are known as standard wavelets and then finally neural network is trained by combined features to identify the cancer nodules.

Fig.1: The proposed architecture

II. LUNG SEGMENTATION

Segmentation is considered to be the essential first step in Computer analysis. Lung is a container of air in human body and hence it visible as darker in CT scan. This helps in segmenting out the lung region from its background. There are various image segmentation techniques available such as Threshold based, region based and edge based methods. Here region based segmentation method 'snakes' which are termed as Active Contour Model is used to segment both the lungs.

A. Localized Active Contour segmentation

The traditional snakes are described as the structure that enlarge or shrinks so that they can exactly overlap the object contour by minimizing its energy. The snake reshapes according to the minimization of global energy which works well when the object is of homogeneous intensity. Here in lung CT, as lung will have heterogeneous intensity, standard global energy method will not produce good segmentation results.

In Localized Active Contour Segmentation [4], foreground and background are described in terms of local regions. These regions are analyzed to construct local energies. In order to achieve localized region segmentation global energies can be localized such as mean separation energy, and histogram separation energy. In our work mean separation energy is included to the energy terms of region.

$$E_{MS} = \int_{\Omega_y} (u - v)^2 \quad (1)$$

u, v are the means of foreground and background. Respective F is calculated as

$$F_{MS} = (u_x - v_x)^2 \quad (2)$$

In some cases where the lung nodules are at the chest wall, as the nodules and chest wall regions are having nearly same intensity values, snakes are intended to segment the lung with elimination of nodules located near chest so that those nodules will be left out in ROI extraction phase. This drawback can be overcome by using the convexity of lung region as follows.

B. Convexity Based segmentation

Hence the human's left and right lungs are symmetric in shape they will have almost same convexity. Normally snakes are shrinking or extending until the specified iteration threshold. Where as in convexity based segmentation, after segmented the normal size lung and then the area based convexity is calculated using the equation

$$convexity = 1 - \frac{Area_{diff}}{Area_{lung}} \quad (3)$$

$Area_{diff}$ is the area of region that the lung differs from its equivalent convex hull.

From the analysis of various normal lung CT slices, it confirmed that the maximum difference in area between left and right lung is 20%.when this difference exceeds this limit, then it is assumed that the smaller lung must have nodules that are attached to the chest wall. Then the snakes are implemented on these small lungs until the convexity of the region covered by active contour match the convexity of larger one.

Fig. 2 : Difference between traditional and convexity based Contour Segmentation

Figure 1 shows how the convexity based Contour differently segment the small lungs when they considered as their nodules have been attached to the chest walls.

III. SPATIALLY CONSTRAINED KFCM

ROI detection has to be carried out to reduce the participation of irrelevant data for nodule identification and classification. Fuzzy Clustering method is mostly preferred in ROI detection since its capability to process multi dimensional information. FCM technique is also posses low sensitive to noise. It is a centroid based clustering in

which the image pixels having intensities same or close to the centroid of one cluster will be associated with higher membership level than that to the other cluster.

A. Kernelized FCM

KFCM [5] is the enhancement of classical Fuzzy C-Means clustering by the incurring input dataset. This method maps the data from data or feature space $\Xi \subseteq R^p$ to the much higher dimensional space H (Hilbert or Kernel space). this mapping is done by the transform function $\square: \Xi \rightarrow H$.

The kernel function here is used to simulate the distances that would be obtained by transferring the points to a higher dimensional space. As the data in higher dimensional space exhibiting clear and simple structures it will be resulted in effortless clustering by FCM. The well known kernel function is Gaussian radial basis function.

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma}\right) \quad (4)$$

Kernel functions convert non linear problems in input domain to linear problems in frequency domain.

B. Spatial FCM

This approach [6] utilizes the spatial coefficient information which is a sum of neighborhood information of each pixel. Considering the spatial function beneficial in improving the noise sensitivity and even more homogeneous regions can be figured out. At first step, initial membership function for each pixel is computed as follows

$$h_{ij} = \sum_{k \in NB(x_j)} u_{jk} \quad (5)$$

Where $NB(x_j)$ represents the resizable neighborhood window of pixel x_j . After that the new membership function is formed by using this spatial function.

C. Advanced SKFCM

This advanced technique [7] combines both the kernel method and spatial function. This method applies the spatial functions to the data in higher dimensional space for identifying new centroid of cluster. Initially the Kernel membership function is computed as follows

$$u_{jk} = \frac{(1/d^2(x_j, V_k))^{\frac{1}{m-1}}}{\sum_{j=1}^c (1/d^2(x_j, V_k))^{\frac{1}{m-1}}} \quad (6)$$

Where

$$d^2(x_j, V_k) = K(x_j, x_j) - 2K(x_j, V_k) + K(V_k, V_k) \quad (7)$$

m is the fuzzy index which decides the fuzziness of the clusters. In next step spatial function is applied to find the membership factor.

$$K(x_j, \hat{V}_k) = \frac{\sum_{i=1}^N (u_{ik})^m K(x_i, x_j)}{\sum_{i=1}^N (u_{ik})^m} \quad (8)$$

The iteration process will be continued until the error lies above the determined threshold. Or else a new iteration will be started.

Fig. 3 : ROI Extraction using SKFCM

IV. FEATURE EXTRACTION

The working performance of computer aided diagnosis system is mainly depends on the feature set used in it. In this proposed approach, both the spatial and frequency domain features are extracted for each region of interest.

A. Spatial Features

These are the features computed by the pixel values in the image. In many cases the spatial feature set is enough to identify the required object. There are various classes of features in spatial domain as follows.

1. *Geometric Features*: - These features are describing the shape and structure of the regions directly. From this category, area, perimeter, equivalent diameter, major axis length, minor axis length, eccentricity such features are calculated for each of the region.
2. *Histogram Feature*: - These features are derived from the histogram information that presents in image regions. There are various features involved with histogram information. Haralick features such as mean, variance, standard deviation, skewness, energy and kurtosis are computed by using the histogram information.

B. Wavelet Features

Wavelets transform has obtained a huge importance in image processing methodology for its multi-resolution representation of image data and its

transformation of data in both time and frequency. When the image is subjected to the wavelet transform, it produces multi resolution versions of input image. In practical wavelet transform decomposes the image into Approximation and detailed coefficients.

Fig. 4 : Wavelet Decomposition

Approximation coefficients will include low frequency components and detailed coefficients will include high frequency components. Wavelet transform gets the input image and transform into multiple frequency bands [10]. The standard wavelet mostly used in image feature extraction is daubechies wavelet. From db2 to db20 are the mostly used from daubechies wavelet family.

By mentioning the level of decomposition the approximation coefficients can be further decomposed into LL, LH, HL, HH frequency components. Here db4 decomposition coefficients have been given.

Table 1 : DB4 Coefficients

NO	LOW PASS	HIGH PASS
1	-0.129409	-0.48296
2	0.22414	0.83651
3	0.83651	-0.22414
4	0.48296	-0.12940

Thus the wavelet transform develops four matrices of various frequency bands. From that low and high frequency components the features such as vertical mean, horizontal mean and energy are computed and they are combine with the spatial features to train the neural network.

V. LUNG NODULE CLASSIFICATION

The main objective of CADx is to recognize the nodules by their features and it cannot be done efficiently without prior knowledge about the distinct features of the cancerous nodules from the linear tissue structures. There are so many nodule detection schemes and classification methods have been developed.

There are two Major categories in classification as supervised, unsupervised and regression classification techniques. Regression method is the most advanced and have better efficiency. Regression technique includes linear regression, Gaussian regression and neural networks.

A. Artificial Neural Network

Neural networks resemble the circuit of biological neurons which are composed of artificial neurons. ANN is composed of interconnecting neurons. Both networks are similar in their tasks. The biological neurons have the aim of solving particular tasks, while the ANN aims to build mathematical models of biological neural systems. A simple neural network will be constituted of three parts as input layer, output layer and hidden layer.

Back propagation is the most widely used technique to train neural networks. It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks which doesn't have connection that loop. Here in this project all importance have been given to the feature extraction which is sufficient for nodule recognition the basic back propagation algorithm [8] is used to train our neural network.

Back Propagation Algorithm:

Initialize random weights to the networks

Do

For each val in the training set

Out=nn_output (network, val);

T=teacher output for val;

Find error (T-out) at output units;

Compute del_wh for all weights from hidden layer to output layer; (bwd pass)

Compute del_wh for all weights from input layer to hidden layer ;(bwd pass)

Update weights in n/w;

Until all values classified or criterion have met
Return the network

Back propagation computes the gradient of the error of the network regarding the network's random weights. This gradient is almost always used in a simple stochastic gradient descent algorithm to find weights that minimize the error. Often the term back propagation is used in a more general sense, to refer to the entire procedure encompassing both the calculation of the gradient and its use in stochastic gradient descent. After the input pattern was presented to the network and processed by all layers, we have errors.

Our testing set contains 400 samples in which 200 samples include nodules in it and the remaining are

normal. The output of this network is 1 if the input image contains nodule, otherwise 0.

Fig. 5 : Nodule classification

VI. RESULTS AND DISCUSSION

The efficiency of CADx system is described by its sensitivity and specificity measures. The well designed system should identify the true positive regions more than analogous to the expert radiologists. True positive means recognize the affected region and mark it as affected. And whereas the false positive is known as mark wrongly the normal region as the affected one. The FP rate should be very less to provide better diagnosis results.

In our project, our analysis is based on high resolution lung CT images taken from Lung Image Database Consortium (LIDC), which are accessed through the National Biomedical Imaging Archive (NBIA). Each image was annotated by four experts, at first during a blind revision, then by communicating the discrepancies that were found and asking for their correction. The studies are stored in the DICOM standard, with size 512x512 pixels and a grey scale of 12 bits in Hounsfield Units (HU). Each case is associated with the XML document which contains the details about nodules present.

A neural network with 500 neurons and input layer of 25 neurons which accommodate 25 values of combined feature set and corresponding two output layers (to classify nodules either affected or normal) has been used at training stage.

Fig. 6 : ROC plot for classification nodules

The above figure 6 shows that the SKFCM technique gives the low index of false positive. In graph, slice 13 got high FP with low TP rate. It should be because of the main part of nodule might be removed during preprocessing.

The proposed scheme has been experimented by randomly selected 15 slices from LIDC dataset (nodule's information known). The result shown that proposed system detecting nodules with the sensitivity of 94%. Micro nodules of less than 3 mm have been challenge and are not marked as nodules. The regions which are with the diameter of less than 5 mm will possess very less possibility of malignancy [9] and hence they can be avoided as non-nodules.

VII. CONCLUSION

In this paper, we have proposed multi stage Computer Aided Diagnosis scheme which segment not even the lung but as well as the nodules that attached to the chest wall by using convexity information of lung region into the Active contour segmentation. ROI's have been extracted by advanced spatially constrained Kernel FCM clustering. This technique processes the spatial neighborhood information in higher dimensional space. The combined feature set was extracted and used in neural network training by back propagation algorithm. This feature set included geometrical, histogram based and wavelet features. The ROI detection scheme in this work works well when we are using three clusters to separate Region of our interest. As 20 cases selected from LIDC collection were used to test the system and it achieved sensitivity of 94% with high TP rate. In future more advanced cancer diagnosing systems can be designed to recognize nodules with much more accuracy with the help of optimal feature set contains both spatial and frequency domain features.

REFERENCES

- [1] M. Antonelli, G. Frosini, B. Lazzerini, "A CAD system for Lung Nodule Detection based on Anatomical model and Fuzzy Neural Network," Fuzzy Information Processing Society, NAFIPS 2006, pp. 448-453, June 2006.
- [2] Depursinge A, Saga D, Hidki, "Lung Tissue Classification Using Wavelet Frames", 29th Annual International Conference on Engineering in Medicine and Biology Society, pp. 6259 – 6262, Aug 2007
- [3] Panfang Hua, Qi Song, Sonka M, Hoffma E, Reinhardt M, "Segmentation Of Pathological And Diseased Lung Tissue In CT Images Using A Graph-Search Algorithm", International Symposium on Biomedical Imaging: From Nano to Macro, pp. 2072 – 2075, April 2011.
- [4] Shawn Lankton, Allen Tannenbaum, "Localizing Region-Based Active Contours", IEEE Transactions On Image Processing, Vol. 17, NO. 11, pp. 2029-2039, November 2008.
- [5] Zhong-dong Wu, Wei-xin Xie, Jian-ping Yu, "Fuzzy C-means clustering algorithm based on kernel method", International Conference on Computational Intelligence and Multimedia Applications, pp. 49-54, September 2003.
- [6] Keh-Shih Chuang, Hong-Long Tzeng, Sharon Chen, "Fuzzy c-means clustering with spatial information for image

- segmentation", Computerized Medical Imaging and Graphics, pp. 9-15, September 2005.
- [7] Alberto Rey, Bernardino Arcay, Alfonso Castro, "Study on Various Defuzzification Methods for Fuzzy Clustering Algorithms to Improve ROIs Detection in Lung CTs", pp. 2123-2130, June 2011.
- [8] <http://en.wikipedia.org/wiki/Backpropagation>
- [9] Jin Mo Goo, "Computer-Aided Detection of Lung Nodules on Chest CT: Issues to be Solved before Clinical Use", Korean J Radiol, pp. 62-63, June 2005.
- [10] Domingues Barros, Gonçalves da Silva, "Characterization of Bone Tissue by Microwaves Using Wavelets and KNN", Journal of Microwaves, Optoelectronics and Electromagnetic Applications, Vol. 10, No. 1, pp. 217-231, June 2011.