International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Segmentation and Feature Extraction of Lung Region for the Early Detection of Lung Tumor

Jaspinder Kaur[1], Nidhi Garg[2], Daljeet Kaur[3]

[1]M.E. (ECE), UIET, Punjab University Chandigarh, India

[2]Assistant Professor, ECE, UIET, Punjab University Chandigarh, India

[3]Assistant Professor, ECE, UIET, Punjab University Chandigarh India

***Abstract:** Among all types of cancer the lung cancer is most acute type of cancer having very less survival rate. It is very difficult to identify the cancer at its early stage with the pathological tests. From the past few decades many CAD systems have been designed to detect the cancer at its early stage. The most of the CAD systems are implemented on CT scan images. The method proposed in this paper is also tested on CT scan images. In this a CAD system has been designed to segment the lung tumor region and to extract the several features of the extracted region. On the basis of these features, the tumor region is differentiated from the surrounding lung structure. The detection process is divided into two parts; 1.Lung segmentation and 2. Feature extraction. In the lung segmentation median filter is used to remove the noise content then region growing method for the lung region extraction and edge detection is done to extract the ROI. Then in the second stage, different statistical and grey level co-occurrence matrix (GLCM) based features are calculated. This system is implemented in MATLAB.*

Keywords: Computer Aided System, Computer Tomography, Lung Segmentation, Optimal Thresholding, Edge Detection.

1. Introduction

A cancer which starts from the lungs is called as a lung cancer. The human body checks and maintains the growth of the cells, when an unbalance occurs; the division of the cells becomes uncontrollable. When the human body fails to produce the cells when they are needed, then a mass is formed known as tumor. With the increase in tumor size, reduces the ability of lungs to provide the blood stream combination with oxygen. When cancer is identified at its early stage then it is termed as tumor, when it reaches to the uncontrollable stage or the secondary stage, then it is termed as cancer. A type of tumor which can be removed with the treatment and can be stopped spreading into the other part of the body is known as benign tumor. A type of dangerous tumor which grows aggressively and cannot be stopped spreading into the other parts of the body is known as malignant tumor. In the single slice of computer tomography, the nodule part i.e tumor and cancer is located on the lung area .This lung area is not even the half of the lung region. So in order to detect only the affected area in the lung region, the segmentation algorithms are used, to save the processing time .If the segmentation is not done, then the false positive detection per scan will be large. Manual segmentation of the CT images is very difficult. In the medical field, lung diseases are very common [1].The major causes of the lung disease are cigarette smoking, second hand smoking, infection, genetics. The lung diseases affect mainly the airways, alveoli, interstitial, blood vessels, pleura, chest wall. Due to these affected parts, the most occurred diseases are: asthma, tuberculosis, cystic fibrosis, and pneumonia. Medical imaging plays a very important role in the lung cancer diagnosis. There are different types of modalities such as chest x-ray, computer tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI) are used in the diagnosis [2] .For the image analysis the CT images are preferred because of less noisy content present in them. These images are more sensitive to find the size of the tumor and presence of lymph node. In the CT images, it is a difficult task to distinguish the tumor in the lungs denser area. Advancement in CT technologies increases its use in diagnosis of different diseases [3]. For assisting the radiologists various CAD system are designed [4]. By investigating the lung tissues ‘pattern in CT images many lung diseases are diagnosed, using the segmentation analysis of the region of interest [5,6].

2. The Proposed Method

The CAD system is fully implemented on MATLAB. The system is tested on total 10 patients, total 650 images are tested. The images are collected from the database provided by LIDC lung image database consortium and from CT scan center. The images acquired are of 512*512 resolutions and are according to the DICOM format.

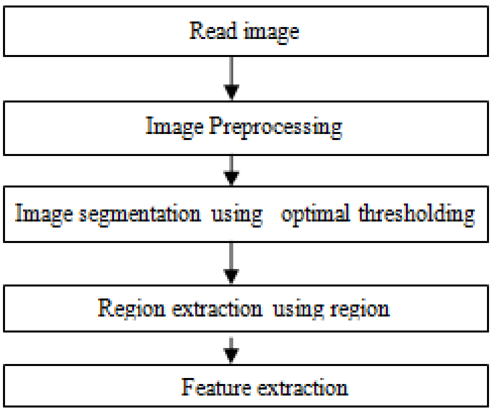

Figure 1: Block diagram of the proposed system

2.1 Preprocessing of CT image

1) In the preprocessing stage, if the image is in RGB scale then it is converted into the gray scale because the gray image reduces the processing time and produces the fast algorithms. The gray scale image is shown in the figure 2.

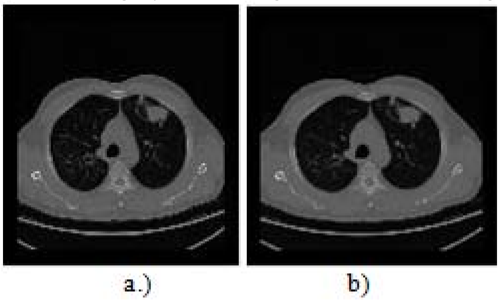

Figure 2: a) Gray scale image b) Denoised image

2) There are several factors which produces the noise, while acquiring the image. To remove the noise content present in the image, median filter is used. Median filter is an order statistic filter, in which each pixel is assigned a rank. This filter produces the less blurring of edges of the image. The filtered image is shown above.

2.2 Segmentation of the lung region in CT image:-

1) The lung and surrounding regions have low value of intensity pixels and body pixels have high values of intensity pixels. The air between dark background and lungs is considered as having non body pixels. The segmentation of the lung regions from the surrounding anatomical part is performed. For the segmentation the optimal thresholding proposed by Shying et al [7], is applied on the denoised image to separate the lung region and non –body pixels using an iterative process.

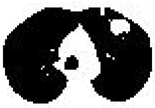

Figure 3: Threshold image

The initial threshold value T_i is selected as the mean of the gray level of foreground and background pixels. To select a new threshold value, to make distinction between two different intensity regions segment the image with the current value T_i. Now T_{i+1} is new threshold level which is the average of body and non body pixels.

$$T_{i+1} = (\mu_a + \mu_b)/2$$

Where μ_a is the average of gray levels of the body pixels and is the average of gray levels of the non-body pixels. The pixels values greater than the threshold value is set to 1 and values less than threshold value are considered as background.

2.3 Extraction of Region of interest:-

After the lung region segmentation using optimal thresholding, the undesirable parts in the lung image which contains air, and low intensity regions are removed using the morphological operations. Morphological operations applied to the segmented image-

1) Morphological opening:-In order to remove the small objects inside and outside the lung region, the opening is applied to the threshold image. The morphological opening is applied with a structure element of the require shape. In this work, structure element of "periodic line" is used with a value of P is 2 and a vector of[2 2].
2) Edge Detection:-Boundary operations are applied to detect the borders of the lung. For this purpose the edge detection using Sobel method is most suitable [8] because of its accuracy and it gives the gradient operator values in both horizontal and vertical directions. An edge detector finds the extent to which the neighborhood of each pixel is partitioned with an arc. The values at both sides of the arc are different.
3) Morphological closing: - To fill the gaps between the borders and to enhance the borders of lung region closing is performed. The closing is done with a structuring element of disk shape having size of 1.
4) Region Growing Algorithm: A pixel is considered as seed pixel, and this pixel continuously determines the similar neighborhood pixel. If the pixel has the same criteria as that of initial seed, then it is added to the region. The region is grown iteratively until no new pixel can be added to the region. The pixel is added to according the similarity index ,if the difference between the region and new pixel is less than the threshold ,then the new pixel value is added to the region, if its value is greater than the threshold value then the region growing process finished. Segmentation process mainly removes the dark area that does not belong to the lungs. This approach selects a minimum pixel value from the boundary of the image, which may be dark pixel. The pixels which follows the 4-neighbourhood connectivity criteria, added to the region. with this algorithm all the black pixels which are connected to the boundary of the lungs are deleted.

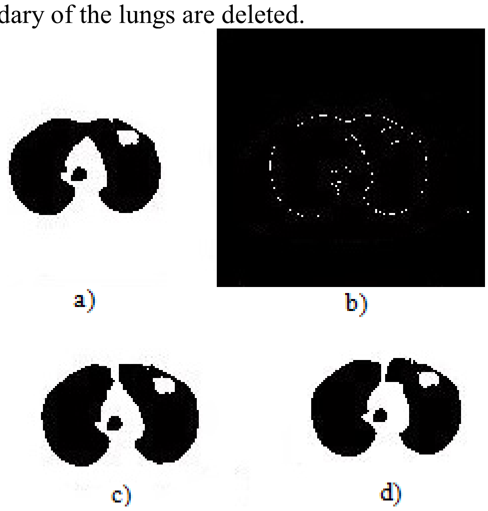

Figure 3:. a) Morphological opening b.) Edge detection c.) Morphological closing d.) ROI

2.4 Feature extraction

After the lung region segmentation and ROI extraction, the most difficult step is to differentiate between the nodule and the structure of the blood vessels and surrounding air.So in order to identify the nodule candidate in the lung region, some geometry based and some texture based features are calculated. The features calculated are best suited to classify between the cancerous/non-cancerous tumors. The features are formed a vector and given to the classifier.

The texture based features are obtained from the gray level co-occurrence matrix (GLCM).Gray level co-occurrence matrix is the matrix that contained all the relative frequencies p(i,j) with intensity value i which occurs in spatial relationship with value j. The total sum of the pixel values in GLCM is 1.

Texture features based on GLCM such as area, entropy, energy, contrast, correlation and homogeneity are considered for classification in the proposed work.

1. Area: It calculates the total number of white pixels in the ROI.

2. Energy: It gives the uniformity of the pixels. It is also known as second order moment. It is calculated from the sum of the square of all the elements in GLCM. The value of the energy lies between 0 and 1.If the image is constant then it has value of 1.

$$Energy = \sum_{i,j} p(i,j)^2$$

3. Correlation: It measures how a pixel is correlated to its neighborhood pixels. Its value lies between -1 and +1.Its value is -1 for perfectly negatively correlated image and +1 for positively correlated image.

$$Correlation = \frac{\sum\sum p(i,j) - \mu_{row}\mu_{col}}{\mu_{row}\mu_{col}}$$

4. Homogeneity: It gives the distribution value of the closeness of elements of the GLCM to the diagonal of GLCM.It gives the value between the range of 0 and 1.

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

5. Entropy:- It is a statistical measure which describes the randomness of an image. Means if the value of the entropy is low it means there is less pixel contrast variation, and high value indicates that the there is greater difference between two pixel values.

$$Entropy = -\sum (p.*\log2\,(p))$$

Where p contains the histogram counts returned from imhist.

6. Contrast: - It gives the local variations in the GLCM.It determines the intensity difference between a pixel and its neighborhood. The value of contrast is 0 for a constant image.

$$Contrast = \sum_{i}^{N_g} \sum_{j}^{N_g} (i-j)^2\, p(i,j)$$

Statistical parameters:

7. Mean: It calculates the mean of all the values in ROI of image.

$$Mean,\ \mu = \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{p(i,j)}{MN}$$

where $p(i, j)$ is the intensity value of the pixel at the point (i, j). The image is of M by N size.

8. Standard Deviation: It calculates the average distance between pixels values and the mean. Low value of standard deviation indicates that there is less deviation of the pixels from the mean where the large value indicates the high variation.

$$\sigma = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} \frac{(p(i,j)-\mu)^2}{MN}}$$

9. Skewness: It gives the asymmetric value of the probability distribution about its mean. The skewness of an image can be positive, negative.

$$= \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{[p(i,j)-\mu]^3}{(MN)\ \sigma^3}$$

10. Kurtosis:-It measures the flatness of the distribution with respect to the normal distribution.

$$= \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{[p(i,j)-\mu]^4}{(MN)\ \sigma^4}$$

11. Fifth and sixth central moment :-These give the deviation from its mean.

Fifth central moment,

$$= \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{[p(i,j)-\mu]^5}{(MN)\ \sigma^5}$$

Sixth central moment

$$= \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{[p(i,j)-\mu]^6}{(MN)\ \sigma^6}$$

These 11 features describe the characteristics of the white mass region called tumor. On the basis of these features, it can be clearly examine that white mass present inside the region is an affected area and can be considered as tumor region. The results of the system are shown in table.

Images	*Area*	*Energy*	*Correlation*	*Homogeneity*	*Entropy*	*Contrast*
1	312	0.7156	09309	0.9908	0.6358	0.0183
2	357	0.6334	0.9400	0.9896	0.7694	0.0207
3	629	0.6685	0.9368	0.9901	0.7137	0.0197
4	4	0.8490	0.9140	0.9940	0.3949	0.0119
5	189	0.6007	0.9521	0.9908	0.8246	0.0182
6	80	0.7740	0.9188	0.9915	0.5332	0.0169
7	164	0.6425	0.9550	0.9923	0.7628	0.0153
8	264	0.6582	0.9553	0.9926	0.7387	0.0146

Figure 4: Table showing the texture feature values

Images	*Mean*	*S.D*	*Skewness*	*Kurtosis*	*FCM*	*SCM*
1	292	70.15	-1.177	2.906	-12.4	23.7685
2	223	59.10	-0.1897	1.010	-4.39	6.797
3	261.04	62.66	-0.38945	1.848	-6.82	11.51826
4	291.43	29.79	-2.57789	8.606	-96.1	27.5831
5	329.83	30.90	-0.58344	4.809	-2.78	3.928539
6	341.58	125.5	-0.73054	1.713	-28.2	63.26745
7	328.76	123.7	-0.38193	1.2853	-4.63	7.2373
8	350.47	104.0	-0.80135	2.0451	-5.60	9.1049

Figure 5: Table showing the statistical feature values

This region can be classified by a classifier. The most commonly used classifiers are support vector machine (SVM) and back propagation network.

3. Results and Discussion

The tables shown above describe the system for total 8 samples. The texture based features are calculated on the basis of GLCM at different offset values. Among all the features calculated above, energy, homogeneity, contrast, correlation describes the values for a non –constant image as described in each feature and in statistical parameters skewness describes the variation in the ROI. All the features calculated are applied to the classifier in the form of vector. On the basis of the output of the classifier it is decided the ROI is cancerous or non-cancerous.

4. Conclusion and Future work

In this paper, the purposed method has been discussed in detail to detect the lung nodule. Firstly nodule part has been detected and several features are calculated. The future scope includes the classification of region of interest. There are different types of classifiers which can be used for the Classification.

References

[1] R. Mason, F. Murray, J. Nadel, “Mason, R. Murray & Nadel's Textbook of Respiratory Medicine 4th Edition”, Elsevier, 2005.

[2] P. Stark, "Use of imaging in the staging of non-small cell lung cancer",UpToDate, 2008. Available at: http://www.uptodate.com/patients/content/topic.do?topicKey=~WskhsDNj2e/3Xq&selectedTitle=4~150&source=search result

[3] I. C. Sluimer, P. F. van Waes, M. A. Viergever, et al., "Computer-aided diagnosis in high resolution CT of the lungs", *Medical Physics*, Vol. 30,pp. 3081-3090, 2003.

[4] H. Abe, H. McMahon, J. Shiraishi, et al., "Computer-aided diagnosis in Chest radiology", Semin. Ultrasound CT MRI, Vol. 25, pp. 432-437 2004.

[5] Y. Uchiyama, S. Katsuragawa, H. Abe, J. Shiraishi, F. Li, Q. Li, C.Zhang, K. Suzuki, K. Doi, "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography", *Medical Physics*, Vol. 30, No. 9, pp. 2440-2454, September 2003.

[6] R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W.Hunninghake, G. McLennan, "Computer recognition of regional lung disease patterns", *American Journal of Respiratory and Critical Care Medicine*, Vol. 160, No. 2, pp. 648-654, August 1999.

[7] Shiying Hu, EricA Huffman, and Jospe h M. Reinhard t," Automatic lung segmentation for accurate quantitiation of volumetric X-Ra CT images", IEEE Transactions on Medical Imaging, vol. 20 , no. 6 pp. 490 -498, June 2001.

[8] Anam Tariq, M.Usman Akram and M.Younus Javed,” Lung Nodule Detection in CT Images using Neuro Fuzzy Classifier,” ©IEEE 2013.

[9] P. Soille, "Morphological Image Analysis - Principles and Applications", Springer-Verlag, 1999.

Author Profile

Jaspinder Kaur is pursuing M.E. (Electronics & Communication) from UIET, Punjab University, and Chandigarh, India. She had received the B. Tech degree in Electronics and Communication from SBBSIET, Jalandhar in 2011. Her area of interest includes Image Processing and Optical fiber communication

Mrs Nidhi Garg is working as an Assistant Professor in UIET, Punjab University Chandigarh, India. She had M.E (Electronics and Communication) in 2006. She is pursuing PhD from PTU, Jalandhar. Her area of interest includes image processing.

Mrs Daljeet Kaur is working as an Assistant Professor in UIET, Punjab University, and Chandigarh, India. She had M.E (Electronics and Communication) from NITTTR, Punjab University, Chandigarh in 2008.She is pursuing PhD from Punjab University, and Chandigarh. Her area of interest includes image processing.