Lung Tumor Segmentation Based On Combination of Concave Hull Region Growing Algorithm

K.Vijila Rani
Research Scholar
Faculty of Information & Communication Engineering,
Anna University, Chennai, India
E-mail: vijilarani.k@gmail.com

Abstract

In this Paper, the lung tumor segmentation and classification from CT images is done. Image processing is used in the medical field for detection of tumor. Image segmentation is a vital part of image processing. Segmentation is the process of partitioning an image into distinct regions. The proposed algorithm has six steps. They are image acquisition, preprocessing, lung boundary correction, tumor part segmentation, feature extraction and classification. The image is preprocessed using Adaptive median filtering. The lung lobe is extracted usingcanny edge detection. The lung boundary correction is performed using Adaptive Concave Hull algorithm. Segmentation is performed using Region growing based technique. Then for the segmented tumor region, the features are extracted using the GLCM (Gray Level Co-occurrence Matrix) algorithm. From the features extracted, the image is classified as the benign or malignantlung cancer by using the SVM with BOVW (Bag of Visual Word) classifier.

Index Terms: *Segmentation, Region growing, Feature extraction, GLCM, Classification, BOVW.*

INTRODUCTION

Segmentation is an important process in digital image processing that has found extensive applications in several areas. Image segmentation plays a major role in the field of biomedical applications. Lung cancer is the fourth most common cancer among male & female globally and the second most common cancer among Indian people. The lung cancers have the ability to spread to other parts of the body that leads to death.The segmentation technique is widely used by the radiologists to segment the input medical image into meaningful regions. The specific application of this technique is to detect the tumor region by segmenting the abnormal CT input image.

The medical images can be segmented manually or automatically. But the accuracy of image segmentation using the segmentation algorithm is more when compared with the manual segmentation. Large amount of time was spent by radiologist and doctors for identification of tumor and segmenting it from other lung tissues. However, exact labeling lung tumors is a time-consuming task, and considerablevariation is observed between doctors. Subsequently, over the last decade, from various research results it is being observed that it is very time consuming method but it will get faster if we use image processing techniques. Primary lung tumors do not spread to other body parts and can be malignant or benign and secondary lung tumors are always malignant. Malignant tumor is more dangerous and life threatening than benign tumor. The benign tumor is easier to identify than the malignant tumor. Also the first stage tumor may be malignant or benign but after first stage it will change to

dangerous malignant tumor which is life threatening.

A tumor is an abnormal growth of cells that serves no purpose. A benign tumor is not a malignant tumor. It does not invade nearby tissue or spread to other parts of the body the way cancer can. In most cases, the outlook with benign tumors is very good. But benign tumors can be serious if they press on vital structures such as blood vessels or nerves. Therefore, sometimes they require treatment and other times they do not.

LITERATURE SURVEY

Armato explains Automated lung segmentation for thoracic CT: impact on computer aided diagnosis1 [1]. Here the information is reliable. But the segmentation is inaccurate 5% - 17% lung nodules are missed. Shiying Hu et al."[7] explains fully automatic method for identifying the lungs in 3D pulmonary X-ray images.It is a simple method does not require the user interaction and time complexity is reduced. But the optimal size selection is difficult because the algorithm depend on the size and shape of the selected morphological structuring elements. JiantaoPu et al. [5] propose Adaptive border marching algorithm: automatic lung segmentation on chest CT images. This method is more efficient and straight forward to implement. But the juxtapleural nodule is missed while minimizing over segmentation of adjacent areas. Thangaraj et al. [8] explains different types of c- means algorithms. The conventional fuzzy c- mean algorithm is good for noise free image. Modified fuzzy c means clustering technique is used for image segmentation. The PCM is more robust in the presence of noise and the performance is good. But fail to segment the images corrupted by noise. Hua et al. propose [2] an automatic algorithm for pathological lung CT image segmentation that uses a graph search algorithm. This

method is more efficient and mostly used in 3D images. But the time complexity is high.Ying et al. [12] explains the fully automatic lung segmentation and repairing based on improved chain code and Brenham algorithm. The segmentation accuracy and implementation speed have been improved greatly than the rolling ball method. It requires Low computational cost. The performance is good. It reduces the missing of juxtapleural nodules. But the complexity is high.Wang et al. [9] explains three phases skin boundary detection, rough segmentation of lung contour and pulmonary parenchyma refinement. The algorithm accurately finds the nodules. But the over segmentation problem occurred.Sarah et al. explains "Robust Lung Segmentation Combining Adaptive Concave Hulls with Active Contours" [6]. This method is a robust segmentation method. But this method doesnot decrease over and under segmentation problem. The novelty of this work includes adaptive median filtering for preprocessing, making the concave hull algorithm "Adaptive" and finally adding Active Contours into the overall method. In this proposed method the image is preprocessed using Adaptive median filtering and Fuzzy thresholding. The lung lobes are extracted using canny edge detection and border tracing algorithm. The lung boundary correction is performed using Adaptive Concave Hull algorithm. Segmentation is performed using Region growing based technique. Then for the segmented tumor region, the features are extracted using the GLCM algorithm. From the features extracted, the image is classified as the benign or malignant tumor by using the BOVW classifier.

METHODOLOGY

The main purpose of this paper is to identify the region of tumor and to do the detailed diagnosis of that tumor; this will be helpful for treating the cancer patient. The detailed about the proposed system is

shown in figure 1.The lung segmentation algorithm is introduced to segment the cervix and correct the segmentation defects caused by different types of nodules.

Pre-processing

Before lung segmentation, some pre-processing techniques are applied on the CT images to produce appropriate images for the next steps. First, image enhancement and noise removal needs to be performed by adaptive median filtering of the CT images. The Adaptive Median Filter detects noisy pixels by comparing them to their neighborhood. Both the size of the neighborhood and the threshold for the comparison are adjustable. A pixel which exceeds this threshold and is also structurally misaligned with similar pixels in its neighbourhood is classified as noise. In that case, the noisy pixel is replaced by the median pixel value of the non-noise pixels in its neighbourhood. After noise removal, a fuzzy thresholding method is used for CT image binarization this is a robust binarization method and helps to remove all the unrelated parts in the tumor areas the goal being keeping only the important areas that are helpful for lung tumor segmentation.

Lung lobes extraction

Following the pre-processing step the canny edge detector is used to specify the borders of the connected components. Then, the well-known border tracing algorithm is used to compute the lung border sequential points; forming a collection of directed closed contours. The extra contours can be removed based on the number of pixels in each contour. In the next step, a novel adaptive concave hull algorithm is applied to obtain accurate initial tumor masks for an active contour model. Angles and curvature are probably the most widely used features for concavity calculation but both of them are vulnerable to noise. In our method, an adaptive concavity degree computation of each boundary pixel is applied for each point on the tumor boundary, its corresponding boundary points by distance "h" are considered as the two endpoints of its line segment. The concavity degree is measured as the ratio of the outside points of the line segment to the total number of the line segment points, as shown in Equation (1):

$$C_i^h = \frac{\sum L_i \cap R'}{(2h-1)} \qquad (1)$$

Where, L_i is the region of the line segment, C_i is the concavity degree corresponding to the I_{th} boundary pixel, R is the region of the tumor image and R' is its complementary region. Furthermore operator $\sum$ shows the number of non-zero values in the corresponding operand. Thus the total number of the line segment points is equal to $(2h-1)$ because two end points are not considered as the line segment points during computation. In addition the other innovation in the adaptive concavity degree computation is that instead of choosing a single value for parameter h we choose multiple values for this parameter. Tumor nodules can be classified into three groups based on their diameter as small (diameter $\leq$ 5), medium ($5 \leq$ diameter $\leq$ 8) and large ($8 \geq$ diameter). Thus three values are considered for 'h' corresponding to these three categories that help us have a robust algorithm that can include different kind of tumor or nodules in our lung segmentation method. In the concave image there is lower intensity on the border points that are in non-convex parts while for the other parts of the contours the intensity is high.

Lung boundary correction

In this step the boundary points of the concave image are traced and a novel method is used to smooth the lung borders by closing the holes on the borders that are caused by tumor or nodules in order to include them in the segmented tumor.

Each of these tumor or lesion contours in the concave image can be characterized with a histogram of the measured concavity degree of the sequence of points that is referred to as concavity histogram. Since this histogram is noisy a Gaussian filter is used for denoising to remove false local maxima and produce a robust segmentation method. Finally find the pair of strong local maximums in the concavity histogram that have strong deep valley between them. These points indicate holes on the lung border because small perturbations in shape might cause a rapid change of curvature value in the histogram. If their Euclidean distance between these points is lower than a pre-defined threshold these two points indicate the two side of the tumor or nodule otherwise this curvature is a natural border along the tumor like the hole on the lung boundary that is due to the space between cervixes. Thus the nodule curvature parts should be closed but the main borders should be kept.

Segmentation
In the last step the result of the previous step is applied as an initial mask of a region growing. This iterative process will deform the model by finding the lung tumor. The region growing is useful for segmenting the borders of tumor accurately and to exclude some tiny white parts that exist in the tumor areas following the result of previous step and decreasing the effects of over segmentation. Thus our approach generally produces robust lung tumor segmentation.
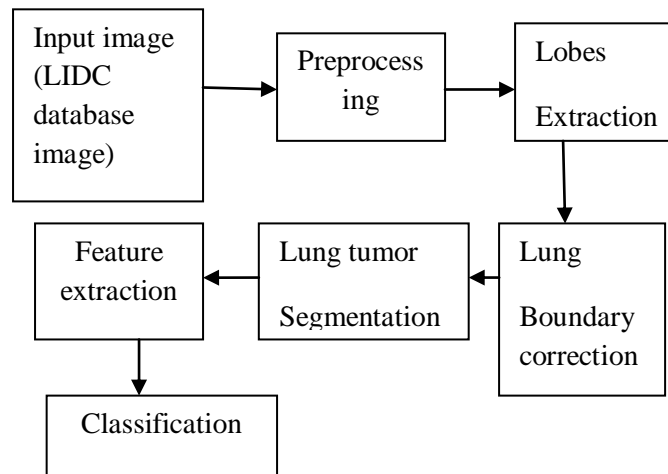

Fig 1. Block diagram of proposed method

Feature extraction
Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power.

Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Feature extraction is done by using Gray Level Co-occurrence Matrix. The gray level co-occurrence matrices represent spatial distribution and gray levels dependency within a local area. There are many texture features available, but here only four features are used. They are Energy, Contrast, Correlation, and Homogeneity.

Energy

Energy $= \sum_{i=0}^{N_p-1} \sum_{j=0}^{N_p-1} p^2(i,j)$ (2)

Contrast

Contrast $= \sum_{i=0}^{N_p-1} \sum_{j=0}^{N_p-1} (i-j)^2 p(i,j)$ (3)

Correlation

Correlation $= \sum_{i=0}^{N_p-1} \sum_{j=0}^{N_p-1} \frac{(i-\mu)(j-\mu)p(i,j)}{\sigma_i \sigma_j}$ (4)

Homogeneity

Homogeneity $= \sum_{i=0}^{N_p-1} \sum_{j=0}^{N_p-1} \frac{p(i,j)}{1+|i-j|}$ (5)

Classification

SVM (Supervised learning) Classifier is a highly efficient classifier. The Bag of Visual Words technique is also called as bag of words. Visual image categorization is a process of assigning a category label to an image under test. Categories of image for this paper will be benign or malignant images. This method contains following steps.

Step 1: Load Image Sets: Cancer image dataset contains two categories: benign and malignant, which can be loaded for classification.

Step 2: Prepare training and validation image Sets: to equalize the images in each and every category.

Step 3: Create a visual vocabulary and train an image category classifier:

- Extracts SURF (Speed up Robust Features) features from all images in all image categories.
- Constructs the visual vocabulary by reducing the number of features through quantization of feature space using K-means Clustering.

Step 4: Evaluate Classifier Performance: First test it with the training set, which should produce near perfect confusion matrix.

Step 5: Try the newly trained classifier (SVM) on test images: then apply the newly trained classifier to categorize new images.

Step 6: Accuracy and confusion matrix will be displayed.

RESULTS AND DISCUSSION

In the experimental setup input images are extracted. During testing normal and abnormal lung images are taken for analysis. The images are taken that of CT scan images. The test image can be of any size. The input images are preprocessed. Then the enhanced images are segmented. The dense tissue regions within the tumors are segmented. Finally the image is classified as normal or an abnormal lung image.

Database

The CT image is collected from International Cancer Center (ICC) and Lung Image database Consortium (LIDC). The image consists of both benign and malignant lung image. The images are shown in figure 2, figure 3.
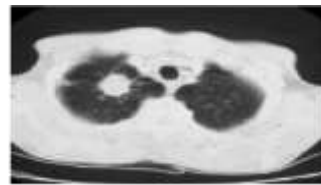

Fig 2: Benign lung images

Fig 3: Malignantlung images

Input Image

The benign and malignant input images are taken from the figure 2, 3, which are shown in figure 4.
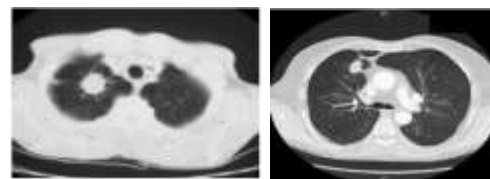

Fig 4: Input Images

Preprocessing

The input images are filtered to remove noise and filtering is done using the adaptive median filter and fuzzy thresholding. Preprocessing output for the corresponding input image is shown in figure5.
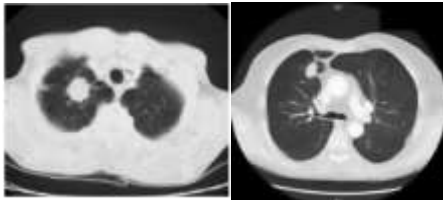

Fig 5: *Preprocessing output*

Lung lobes extraction and lung lobeboundary correction

Lobes extraction and Lung border correction also done for identifying the region of interest, the outputs are shown in figure 6.
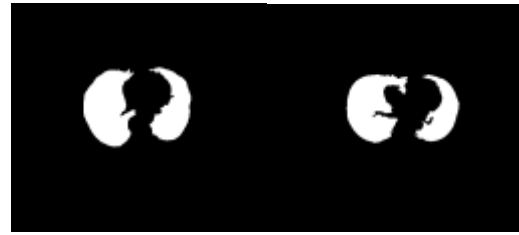

Fig 6: *Lobes Extraction And Lobe Border Correction Output*

Segmentation

The preprocessed output is segmented to detect the tumor region. Segmentation is done using the region growing algorithm. The output of segmentation is shown in figure 7.
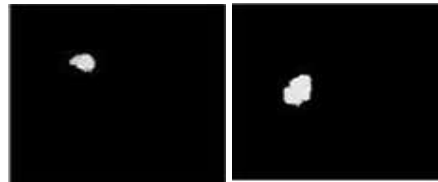

Fig 7: *Segmentation output*

Feature extraction

From the segmented region the features are extracted. The extracted features are shown in table 1.

Table 1 *Feature Extraction Output*

GLCM Features	Benign			Malignant		
	1	2	3	1	2	3
Contrast	0.040	0.036	0.025	0.023	0.041	0.031
Correlation	0.961	0.970	0.971	0.966	0.967	0.972
Energy	0.964	0.9409	0.9637	0.9642	0.9476	0.9547
Homogeneity	0.996	0.9944	0.9962	0.9960	0.9944	0.9959

Classification

Features are extracted for the segmented region and based on this feature values, the input image is classified as benign or malignant lung by using the BOVW classifier.
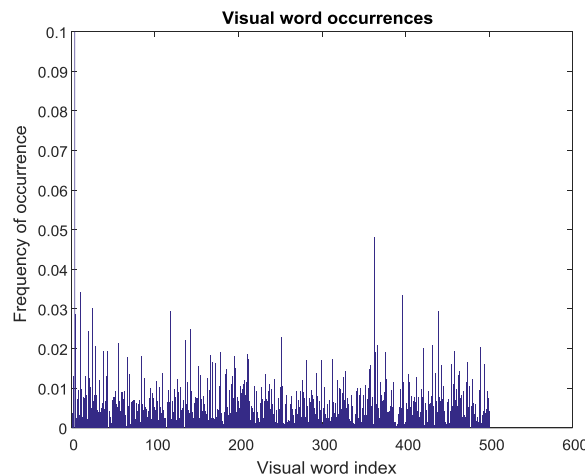
Fig. 8: Visual Vocabulary BOVW Classifier

This is the extension of NLP algorithm. BOVW is the supervised learning model. Use the computer vision system toolbox functions for image category classification by creating a bag of visual words. The process generates a histogram of visual word occurrences that represent an image. This histogram is used to train an image category classifier.Figure 8 shows that the bag of visual words classification results in which the accuracy will be 92.8%. Before do classification the image bag of features are extracted and then form the confusion matrix for this method.

CONCLUSION

In this work, a system that can be used for segmentation of lung CT images for detection and identification of lung tumor. Lung tumor detection is done by preprocessing, which is the first step in that noise is reduced by the adaptive median filter, the image is enhanced and lobes extraction, lung border correction also done for identifying the region of interest. After this process fatty tissues and other unwanted details get smoothened. Preprocessed image is segmented with the region growing. After that texture features are extracted using the GLCM algorithm and BOVW is used for classification. Then the area of tumor and its type of tumor is found.

REFERENCES

1. SG Armato, WF Sensakovic, 'Automated lung segmentation for thoracic CT: impact on computer-aided diagnosis1 SG' - Academic Radiology, 2004.
2. Hua, Panfang, et al., 'Segmentation of pathological and diseased lung tissue in CT images using a graph-search algorithm', in: IEEE International Symposium on Biomedical Imaging: from Nano to Macro, 2011.
3. Hwang H. and R.A. Haddad, 1995. 'Adaptive Median Filters: Algorithms and Results': IEEE Transactions on Image Processing, 4(4): 499-502
4. Jiahui Wang, Feng Li, Qiang Li, 'Automated segmentation of lungs with severe interstitial lung disease in CT', Med. Phys. 36 (2009) 4592.
5. JiantaoPu, et al., 'Adaptive border marching algorithm: automatic lung segmentation on chest CT images', Comput. Med. Imag. Graph. 32 (6) (2008)
6. Sarah Soltaninejad, Farshad Tajeripour, ' Robust Lung segmentation combining Adaptive concave hull with active contours', 11th Intelligent Systems Conference (ICIS), (2016).
7. Shiying Hu, Eric A. Hoffman, Joseph M. Reinhardt, 'Automatic lung segmentationfor accurate quantitation

of volumetric X-ray CT images', IEEE Trans. Med. Imag. 20 (6) (2001) 490498.

8. Thangaraj P, Gomathi, 'A New Approach to Lung Image Segmentation using Fuzzy Possibility C-Means Algorithm;, International Journal of Computer Science and Information Security, Vol. 7, No. 3, pp. 222-228, 2010.

9. Wang L.A, Bray F, Siegel R.L, Ferlay J, Lortet-Tieulent J, Jemal A. CA. 'Automatic Approach for Lung Segmentation with Juxta-Pleural Nodules from Thoracic CT Based on Contour Tracing and Correction' 2016 Mar;65(2):87-108

10. Withey D.J. and J.H. koles (2007), 'Three Generations of Medical Image Segmentation : Methods and available Software'.

11. Xu T, Mandal M, Long R, Cheng I, Basu A. 'An edge-region force guided active shape approach for automatic lung field detection in chest radiographs'. Comput Med Imaging Graph. 2012;36(6):45263

12. Ying Wei, Guo Shen, Juan-Juan Li, 'A fully automatic method for lung parenchyma segmentation and repairing', J. Dig. Imag. 26 (3) (2013).