International Journal of Innovative Computing, Information and Control
Volume **7**, Number **4**, April **2011**
ICIC International ©2011 ISSN 1349-4198
pp. 1875–1890

GENETIC FUZZY BASED AUTOMATIC LUNGS SEGMENTATION FROM CT SCANS IMAGES

M. Arfan Jaffar, Amjad Iqbal, Ayyaz Hussain, Rauf Baig and Anwar M. Mirza

Department of Computer Science
FAST National University of Computer and Emerging Sciences
A. K. Brohi Road, H-11/4, Islamabad, Pakistan
{ arfan.jaffar; amjad.iqbal; ayyaz.hussain; rauf.baig; anwar.m.mirza }@nu.edu.pk

Received August 2009; revised April 2010

Abstract. *Segmentation of images has become an important and effective tool for many technological applications like lungs segmentation from CT scan images, medical imaging and many other post-processing techniques. Lung cancer is one of the leading causes of death in the world. In this paper, a fully automatic un-supervised strategy has been developed for the segmentation of lungs. No prior assumption is made about features, types, contents, stochastic models, etc. of the images. A fuzzy histogram based image filtering technique has been used to remove the noise, which preserves the image details for low as well as highly corrupted images. The proposed technique finds out optimal and dynamic threshold by using genetic algorithms. For edge detection, we have used morphological operators. The proposed system is capable to perform fully automatic segmentation of CT scanned lung images. It can be used as a fundamental building block for a computer aided diagnosis systems. We have tested our technique against the datasets of different patients received from Aga Khan Medical University, Pakistan.*
Keywords: Genetic algorithm, Computer aided diagnosis, Mathematical morphology, Thresholding and segmentation

1. **Introduction.** Lung cancer is one of the leading causes of death in the world. It is very difficult for patients to detect lung cancer symptoms until the cancer is in an advanced stage [1]. The early detection of abnormal lung nodules via computer aided diagnosis of lung CT scan images has been, however, a significant step towards this end. The abnormality nodules can be that of tuberculosis, cancer, pneumoconiosis, infectious and non-infectious granulomas, mucous plugs and hypersensitivity pneumonia. These nodules are situated within the lung parts of the CT scan image that is usually less than half of the area of the CT slice. If nodules have to search in the whole slice, it will take a long time. To reduce time to search nodules in the CT slice, we have to search only in the area where the nodules exist in the CT slice. Therefore a mechanism is needed to segment that part of lung. For this purpose, we have developed a fully automatic method based upon Genetic algorithms and Morphology based image processing techniques to segment that part. After segmenting that part, nodules can be searched only in that segment part of lung. Thus segmentation can be used as a preprocessing step of a CAD.

Segmentation refers to the process of partitioning a digital image into sets of pixels or multiple regions [2,21]. The goal of segmentation is to change the representation of an image into something that is more meaningful and easier to analyze [3,22]. The result of image segmentation is a set of regions that collectively cover the entire image. In a region, each of the pixels is similar with respect to some characteristic such as intensity, color, or texture. Adjacent regions are significantly different with respect to the same characteristics.

Computer-aided diagnosis (CAD) can be defined as a diagnosis made by a radiologist/clinician whose decision-making process incorporates output from computerized analyses of the medical image. The CAD systems assist the radiologist in their final decisions [4,23]. For pulmonary imaging, high-resolution X-Ray computed tomography (CT) is the standard that can be used for applications such as lung parenchyma nodule discovery, airways analysis and density analysis for diagnosing early lung cancer [5,24].

In this paper, we propose a fully automatic system that is capable to perform segmentation of lungs images. Genetic algorithm has been used to calculate optimal threshold that extract the lung part from the original image. Morphology has been used for edge detection and Susan algorithm is used for thinning the edges.

Main contributions of the proposed technique are:

- Fuzzy based noise detection method and filtering technique have been used to remove the noise, while preserving the image details for low as well as highly corrupted images.
- Histogram based background removal operator have been used that removes background fully in an automatic way.
- Proposed technique finds out optimal and dynamic threshold by using genetic algorithm.
- Morphological image processing techniques have been used for edge detection.

The paper is organized as follows. Section 2 contains a survey on previous research that is most closely related to the present work and finds out problems. This is followed by s detailed description of the proposed system follows (Section 3). Data characteristics, implementation and relevant results are presented in Section 4. Finally, Section 5 ends the paper with several conclusions drawn from the design and the work with the proposed system.

2. **Related Work.** Brown et al. [6] have used precise anatomical knowledge that is used to generate an anatomical model. They showed that 86% of the lung segmentation was correct while 14% required manual corrections while considering 104 data series comprising 1313 images. Shiying et al. [7] have developed an automatic method for identifying lungs in 3D pulmonary X-Ray CT images. They have divided their work into three main stages: 1) Grey-level thresholding have been used to extract lung from CT-Scan image 2) left and right lungs are separated through identification of the anterior and posterior junctions by using dynamic programming and 3) to smooth the irregular boundary along the mediastinum, a sequence of morphological operations for acquiring results. It can be observed that that this technique has the following very serious short comings: (i) Fixed ball size (ii) addition of unnecessary areas as lung regions (iii) Processing time overhead. It can also be observed that the ball algorithm includes the unnecessary areas as the lung region, (not actually the part of the lung).

Hu et al. [8] have used an iterative search method to calculate an optimal threshold for each CT case. They have used conditional morphological operations to segment lung regions. An automatic CAD system for lung cancer screening was developed by Ayman El-Baz et al. [9]. For this purpose, they have used chest spiral CT scans. This paper presents the first phase of an image analysis system for 3-D reconstruction of the lungs and trachea, detection of the lung abnormalities, identification or classification of these abnormalities with respect to specific diagnosis. Optimal gray-level thresholding is applied by Ayman El-Baz et al. for the extraction of thorax area. After the selection and application of threshold, region growing and connectivity analysis are used for extraction of the exact cavity region with pre determined accuracy. Some outstanding results have been observed using this scheme for all CT-sections having distinctive intensities of lung parenchyma. But these techniques do not perform well, when there is overlapping of intensities in lung parenchyma and surrounding chest wall.

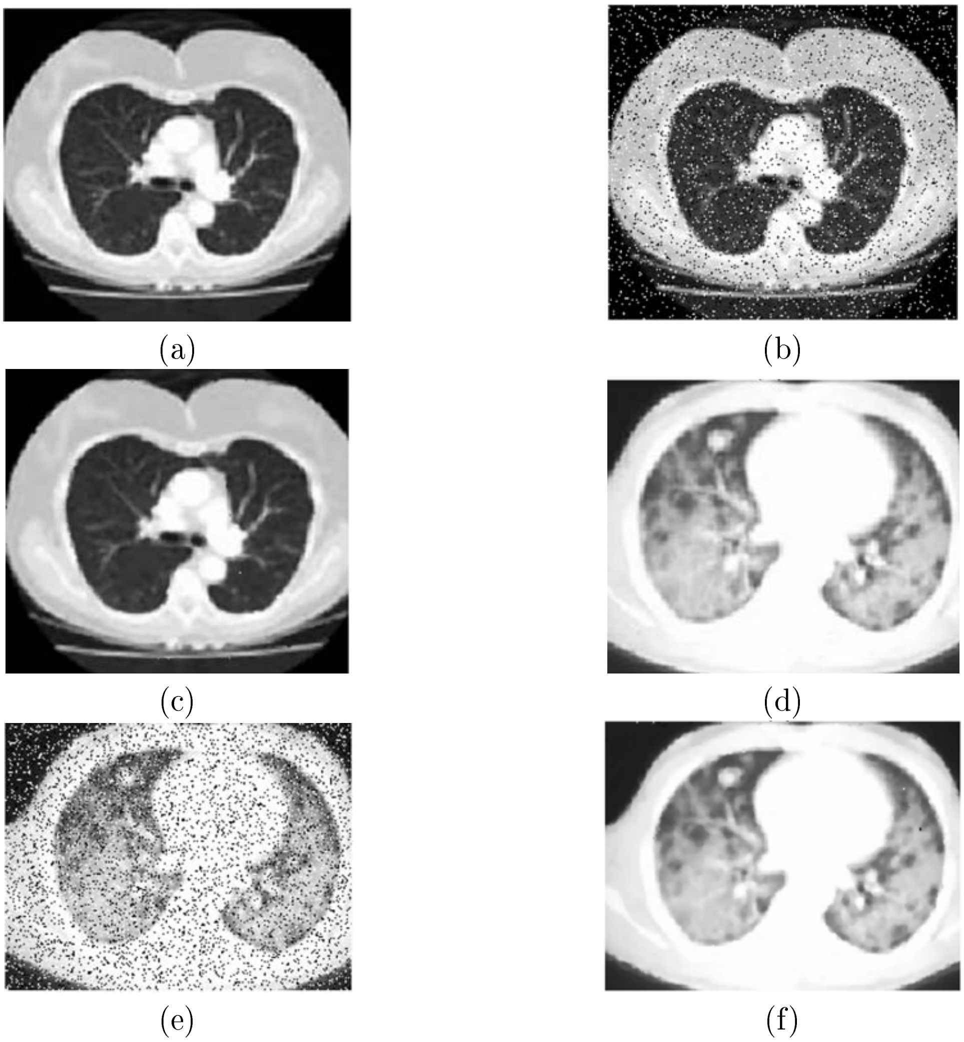


FIGURE 1. (a) Original lung image (b) Image corrupted with 10% impulse noise (c) After preprocessing step (PSNR = 37.848 (d) Original lung image (e) Image corrupted with 20% impulse noise (f) After preprocessing step (PSNR = 34.738)

Zhang and Valentino [10] suggested using artificial neural networks to classify each pixel in the CT slice into different anatomical structure. Kuhnigk et al. [11] have used anatomy guided 3D watershed transform for lung lobe segmentation. Wei [12] proposed a DCT descriptor for each pixel of the image. An adaptive k-means clustering algorithm was used for assignment of labels to each pixel. K-means algorithm performs well for large sets of data. Binsheng et al. [1] have used histogram for threshold selection. Then this threshold is used to separate the lung parenchyma from the other anatomical structures on the CT images in the first stage. There is always a difference in the apparent density of voxels and bronchial walls in the lungs of CT scanned images. Thus it can be possible that structures with higher densities including some higher density nodules could be grouped into soft tissues and bones. As a result, it produces an incomplete extraction of lung mask. In order to obtain complete hollow free lung mask, morphological closing is applied. Morphological operators help to choose spherical shape of the structural element so that filter size can be approximately determined. With the help of 3D mask, the lungs can be readily extracted from the original chest CT images.

Antonelli, et al. [13] have used a background removal operator and iterative grey level thresholding methodology for lung segmentation. Background cannot be eliminated completely when there is noise at the corners. It also fails when the lung portion is connected to the boundary of the image. However, the presence of a distinct outer black region is a pre-requisite for their accurate and precise performance. This region should enclose the whole lung parenchyma and the adjoining features.

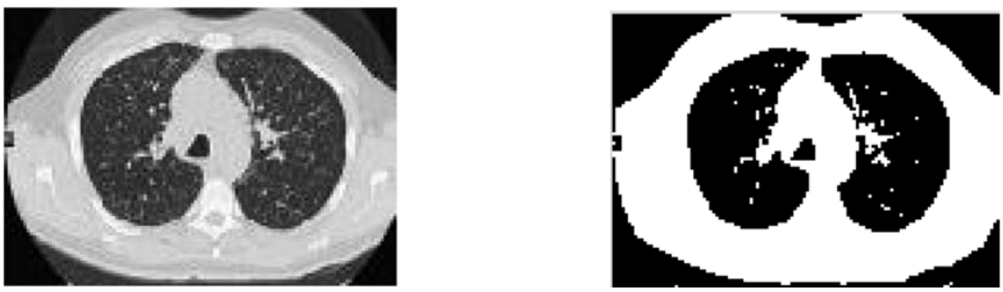
Samuel et al. [14] has used Ball-Algorithm for the segmentation of lungs. Grey level thresh-holding process have been used and applied at each CT image. The major purpose of this process is to segment the thorax from background and then the lungs from the thorax area. After then, rolling ball algorithm is used for lung segmentation contours, acquired in the first stage. The major advantage of application of this algorithm is to avoid the loss of juxtapleural nodules. The main drawback of this technique is the size of the ball selected for morphological operation. It has been observed that selected ball size for morphological closing did not work for the whole database of a single patient. Some times a ball of specified size worked for one patient while it was not enough for another patient. Thus we have to vary the ball size patient by patient which makes the algorithm semi-automatic.

Thus if such a situation arises, these approaches are most likely to yield poor results. Thus we can say that these methodologies only work partially.

Prewitt et al. [16] proposed the mode method to choose thresholds at the valleys on the histogram. Some smoothing of the histogram data is required for the automatic selection scheme, searching for modes, and placing thresholds at the minima between them. Their method relied heavily on the structure of the gray level histogram, which contained peaks and valleys corresponding to gray level subpopulation of the image. Heuristic search method fails to find the two peaks. It is also difficult to find the exact threshold if the valley is flat. However, the bottom of the valley is some thing difficult to locate.

3. **Proposed Method.**

We propose a system that consists of combination of fuzzy logic and genetic algorithm. It performs automatic and robust segmentation of lungs images by calculating optimal threshold of the image. A block diagram of the proposed algorithm is depicted in Figure 2. Our proposed method is based on the following steps:



Original image Threshold image

FIGURE 2. Thresholding step

- Fuzzy based noise detection method has been used to detect corrupted image. In case of noisy image, fuzzy based filter have been used to remove noise, while preserving the image details for low as well as highly corrupted images.

- Genetic Algorithms have been used based to calculate optimal and dynamic threshold.
- The background (i.e., the pixels outside the lungs) is removed from the image by using histogram based technique.
- Morphological operations have been incorporated to find edges; in particular, the opening operator is adopted to eliminate the small objects inside the lung and to separate some regions that should be separate but are joined by a thin area of foreground pixels, whereas the closing operator aims to get rid of small holes that represent border interruptions.
- Thinning is used to reduce the border size to one pixel by using SUSAN thinning algorithm.
- Reconstruction of pulmonary lobe's border is performed by an operator which, based on the border shape, reinserts erroneously erased zones; in particular, the nodules adjacent to the pleura that have been eliminated by thresholding are recovered.
- Filling operator is applied to the pulmonary lobes chains, thus reintroducing the correct values of grey levels inside the lungs.
- Lung part is separated from original image.

In the following, the various steps of the method are described with more details.

3.1. **Preprocessing block.** Usually medical images are characterized as low contrast images with complex noises. This complex nature of noise is attributed to various acquisitions, transmission storage and display devices. Another reason for this is application of different types of quantization, reconstruction and enhancement algorithms. Irrespective of whichever techniques, devices or algorithm is used; all medical images contain visual noise. As a result, image is transformed into a new mottled, grainy, textured or snowy appearance. This noise in images is contributed by various sources. Noise is much more prevalent in certain types of imaging procedures than in others. Nuclear images are generally the noisiest whereas noise is also significant in Magnetic Resonance Imaging (MRI), Computer Tomographic (CT) and ultrasound imaging. Most of these types of images are extensively used in medical domain. Most significant factor of the noise apart from giving an undesirable look is that it can cover and reduce the visibility of certain features within the image. This is even more significant for low-contrast objects (medical images fall in this class as we had mentioned earlier).

CT scan images of a lung slice of 25 patients were denoised using fuzzy based filter. It is possible that there can be various types of noises like the Random noise, Gaussian noise, Salt, Pepper and speckle noise in the images. We have tested our technique with salt and pepper noise. So it is possible that there can be noise in the medical images. So preprocessing step is necessary.

In order to segment the medical image correctly, we have added a preprocessing step in the proposed method to detect whether the input image is corrupted with some kind of noise or not. In case of a corrupted image, nodule detection after segmentation process can be affected. Therefore, before applying the segmentation process, we have used a noise detection method [17], which gives us a noise map indicating whether the image is corrupted or not. Input image will be considered corrupted if at least one of the entries in noise map is one. In case of noisy image considered through noise map, a noise filtering process [17] will be carried out, so that the noisy pixels should not affect the accuracy of the segmentation process. Since the medical images contain highly sensitive information, the filtering process should be efficient enough to preserve image details such as blood vessels and nodules. Therefore, the filtering method used for noise removal also contains detail preservation process which preserves image details that might have been blurred during the noise removal process.

In order to detect and remove noise from the corrupted CT-scan images, a Detail Preserving Fuzzy Filter (DPFF) [17] is used. DPFF is used to remove salt and pepper impulse noise from digital gray scale 8-bit images. Since images corrupted with salt and pepper noise contains pixels with extreme values such 0 or 255 for 8-bit gray scale images, therefore noise detection becomes quite easily. In the first step, noise detection is performed in DPFF, which gives the set of noisy pixels N_{pixels}. To determine N_{pixels}, we scan the image with window of size 3×3. The central pixel will belong to the set N_{pixels}, if it is minimum, maximum, less than some threshold T or greater than 1-T. Based on histograms of noisy pixels (N_{pixels}) and corrupted image, histogram estimation is performed to find out the parameters of the fuzzy membership functions. Further a window of size 3×3 pixels is used to scan across the entire image to calculate the predicted intensity and fuzzy mean values. Predicted intensity is the average of the non-noisy pixels in the considered window whereas fuzzy mean is calculated using the fuzzy membership function whose parameters will be calculated using Fuzzy Number Construction Process [17]. Fuzzy parameters, fuzzy mean values and predicted intensities obtained are further used in fuzzy control and fuzzy decision making to remove noise and preserve the image details in an efficient manner.

Fuzzy controller is the main contributing factor for impulse noise removal. Following are the two main steps involved in fuzzy control.

Fuzzification Noisy window of size 3×3 pixels is given as input in the first step to fuzzy controller. The component performs fuzzification of these pixels. In fuzzification, degree of membership for each pixel of sub-image is determined against fuzzy sets very dark (vdk), dark (dk), medium (md), bright (br) and very bright (vbr).

Defuzzification In this step, defuzzification is performed on the fuzzified values of the pixels under considered window. All the outputs from the previous step belonging to each membership function are separately used for defuzzification. Resultantly, the outputs of this step will be five crisp values. These five crisp values are then compared with the fuzzy mean and the closest value to fuzzy mean is then chosen as the final output of the fuzzy controller.

Fuzzy decision making includes four computation functions which are f_{diff}, f_{x1}, f_{x2} and f_{plus} and two fuzzy membership function f_{large} and f_{small}. Output of the fuzzy controller and intensity estimation are used in fuzzy decision making. Function f_{diff} is used to compute the absolute difference between input values. Difference computed is then used to decide the weights given to the output of intensity estimation process and fuzzy controller process using the fuzzy membership function f_{large} and f_{small}. For more detail reader should refer to [17].

3.2. **Threshold block.** In an ideal case, for images having two classes, the histogram has a deep and sharp valley between two peaks representing objects and back ground respectively. Thus the threshold can be chosen at the bottom of this valley. However, for most real pictures, it is often difficult to detect the valley precisely, because (i) valley could be flat and broad and (ii) the two peaks could be extremely unequal in height, often producing no traceable valley. In our proposed method we used GA to find out the peaks and valley in bimodal class of images. GAs are used for function optimization process and hence determining the global optimal solutions.

We have considered the images whose histogram has two peaks. Crowding method will help us to detect the two peaks. After getting the two peaks we can use GA to find out the valley bottom between these peaks. Here we have considered both types of image having flat valley as well as sharp valley in the histogram. Our discussion is confined to the elementary case of threshold selection where only the gray-level histogram suffices without other a priori knowledge. Our algorithm does not require any valley sharpening techniques.

The best case for any images that have two classes is that their histograms have a sharp and deep valley between two peaks representing two different objects. This deep valley can be used as a threshold to segment those objects [18]. Original CT scanned lung images have also two different and clear objects. Background and lungs part of the original image is considered as a single object because both have similar grey levels while other object is the middle portion of images. When we draw a histogram and normalize it, then there are two clearly defined peaks and a single deep valley. We have to find out these two peaks and valley and for this purpose, we have applied GA on the normalized histogram to find out these peaks and valley.

Genetic Algorithms can be used to obtain global solution. However, in this problem we need global as well as local solutions. Thus a mechanism is needed to maintain the subpopulation and local solution. We have used a crowding operator to maintain subpopulation at the two peaks. Then we have used GA to find out valley between these two peaks. This method can also work even if the valley is flat. Experimental results show that this method works correctly and efficiently for CT scanned images of lungs. This is because normalized histogram of these images has two distinct peaks.

To implement GA on image segmentation, we initially create chromosomes using input image. For this problem 8 bit length chromosomes are created. Then we draw normalized histogram. To find out two peaks, we implement Genetic Algorithm. To maintain peaks of both classes, we use the crowding operator in GA. After finding peaks, we select specific range to find out valley between two peaks. Genetic Algorithm is applied again to find out valley. The value of valley thus determined is well defined threshold to separate out background and object.

Algorithm

GA consists of determining the two peaks of the histogram distribution and find out the minima in between each pair of peaks [18]. The salient steps of the algorithm are:

Find out min (minimum) and max (maximum) grey level value of the image.

Input: Image after Noise Removal

Output: Thresholds T

Procedure

Step 1: Initialize Population of size N randomly between min-max values and their classes are determined.

Step 2: Calculate fitness of each chromosome (The fitness function is the normalized histogram function p(g).

- *Step 3:* Choose two parents for reproduction (Crossover and mutation) operators.
- *Step 4:* Create offspring and compete with parents.
- *Step 5:* Select best chromosome based on Tournament Selection.
- *Step 6:* Put these selected chromosomes in their respected classes.
- *Step 7:* Steps 2-6 are repeated for all chromosomes in the population and Step 7 is repeated till convergence met (elements of respective classes are equally fit).

Step 8: Peaks can be determined from these converged classes.

Step 9: All Steps 1-7 is repeated with min = peak1 and max = peak2 again.

Step 10: The converged value is the gray value corresponding to the valley between the two peaks. This value is used as a threshold to segment the image.

3.3. **Background removal.** By simply applying threshold to the image, we cannot get whole lungs part from background. There is high degree of similarity between the gray levels of the lungs and the image background as shown in the Figure 3(c). So there is need a mechanism to remove whole background. Most of the techniques remove background by starting at the four corner points and move inward till the grey level of that object changed or rows or columns ends. But it is possible that there is noise at the corners or at any other point or may be corners or any other boundary pixel is connected to lung part. In these cases their technique did not remove whole background. The major problem in these techniques is the seed point. From where the operator has to start to remove background.

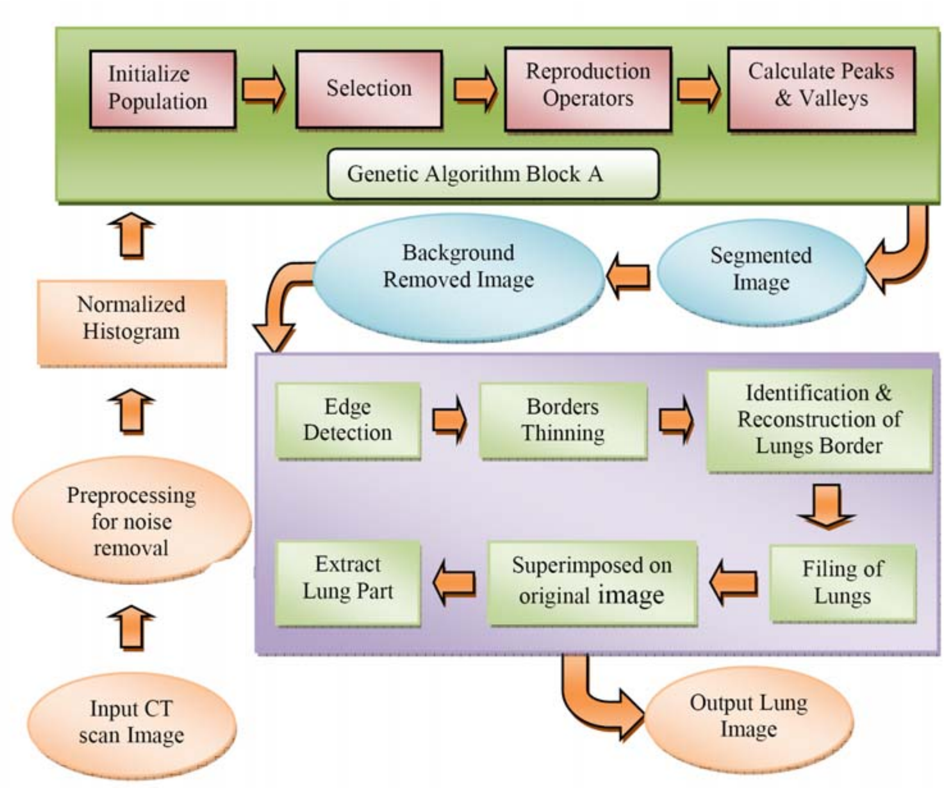


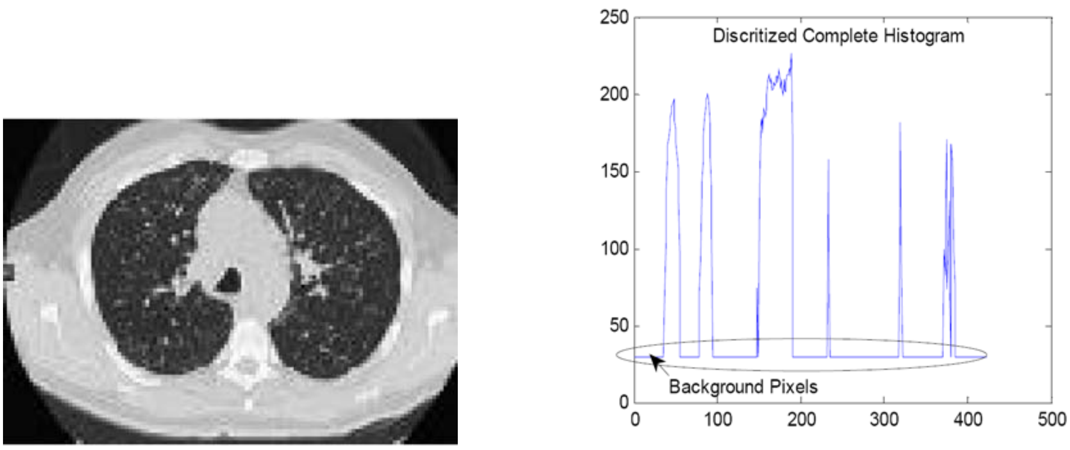
FIGURE 3. Block diagram of proposed system

We have proposed a histogram based method for this purpose. This technique works fully automatically. The technique starts traversing the image from start of the first column and traverses the whole boundary finishing at the same point where it started. While traversing, it keeps track of all the pixel values and plots them in the form of histogram. When applied experimentally, it showed regions of the background attached to the boundary. So the histogram properly showed peaks and valleys as shown in the Figure 3. Once this histogram is obtained, we can start at any point in the grey level region (background area) and start removing it until the background area is finished. This process provides us a seed point to start background removal operator. The advantage we have using this technique is that, with the help of histogram, we now exactly knows where the background area lies in the image and we remove all those areas. This technique also worked in the case of noise or for those images whose corners have changed. Thus we achieve fully automated and complete background removal process as shown in Figure 7(b). The image resulting from the application of this operator consists of just the chest and the lungs, as shown in Figure 8(b).

3.4. **Edge detection and thinning.** Morphological edge detection algorithm selects appropriate structuring element of the processed image and makes use of the basic theory of morphology including erosion, dilation, opening and closing operation and the synthesization operations of them to get clear image edge. In the process, the synthesized modes of the operations and the feature of structuring element decide the result of the processed image. Detailedly saying, the synthesized mode of the operations reflects the relation between the processed image and original image, and the selection of structuring element decides the effect and precision and the result. Therefore, the keys of morphological operations can be generalized for the design of morphological filter structure and the selection of structuring element. In medical image edge detection, we must select appropriate structuring element by texture features of the image. The size, shape and direction of structuring element must been considered roundly. Usually, except for special demand, we select structuring element by 3×3 square.

The basic operators of binary morphology are erosion, dilation, opening and closing. In this paper, a mathematical morphology edge detection algorithm is used. Opening-closing operation is firstly used as pre-processing to filter noise. Then smooth the image by first closing and then dilation. The perfect image edge will be got by performing the difference between the processed image by above process and the image before dilation.

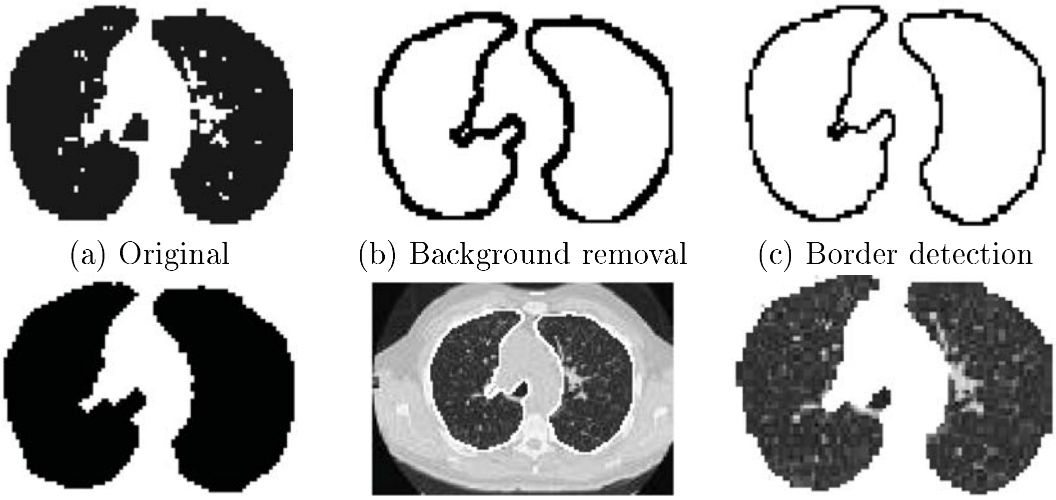
Border detection algorithms detect borders whose width consists of more that one pixel. To reduce the borders width to one pixel, pixels chains are built that characterize borders. We have used the Susan thinning algorithm [20]. In this algorithm, a pixel is measured and consider border pixel if at least one of its neighbours is white. In this way a border chain is build that is a set of connected border pixels. The algorithm starts considering all the pixels as border chains. Then it eliminates the pixels for each chain in such a way that these eliminated pixels does not influence the border connectedness. In this way it produce an image with one-pixel contours (see Figure 3(f)).



Original image

FIGURE 4. Background removal step

3.5. **Lungs border identification and reconstruction.** After thinning the borders, the two pulmonary lobes that represent the region of lungs part are chosen. The two longest border chains are chosen as lungs. But due to thresholding and edge detection, it is possible that border of lungs part is wrongly eliminated. So we have to reconstruct it [13]. Reconstruction algorithm is shown in Figure 2.



(a) Original (b) Background removal (c) Border detection

(d) Lung area filling (e) Restored (f) Segmented

Figure 5. Proposed system steps

3.6. **Filling and lungs part extraction.** We have applied region filling technique for identification of the set of pixels belonging to the internal part of the two pulmonary Lobes and for restoration of their initial gray level. The two thinned chains that represent the lung lobes are superimposed on the original image to show the accuracy of our method in identifying the pulmonary regions and extraction by using that boundary. Extracted part is only lung image that contains all information. There is no loss of nodules within the lungs part. The region to examine for nodule detection has now been highly reduced.

Input: Image after thinning border, threshold T
Output: Reconstruction of Lungs border
Procedure
For each pair of border pixels
Step 1: Compute d_1
$d_1 = (x_2 - x_1)^2$
Step 2: Compute d_2
$d_2 = (y_2 - y_1)^2$
Step 3: Compute ratio r
$r = d_2/d_1$
Step 4:
if $(r \geq T)$
pixel is candidate for reconstruction
else
skip pixel
endif

Figure 6. Algorithm for reconstruction

4. **Experimental Results and Discussion.** The following are the major features of the segmented data set:

1. 512×512 pixels of each image
2. 8-bit grey scale images (255 grey levels)
3. Slice thicknesses of varying data

We have collected data in the form of DICOM. We have converted into $*$.tif for our experiments. The collaboration with Aga Khan University Hospital for this research project, made it possible to get the database from Aga Khan University Hospital Karachi. The CT scanning facilities available at the hospital are the best available in Pakistan. The embedded UNIX-based computing systems of the CT scan equipment are interfaced with windows machines as well which led to an easier exporting of the CT scan data. Complete cases for 25 patients were acquired from AKU which were axial CT scan slices at a slice width of minimum of 5 mm. Table shows the index of the database.

The most common standard for receiving scans from hospital is DICOM. While the analyze format is required for image processing algorithms. Thus there is need of converting DICOM images to analyze format. The DICOM software provides Export utility that can export DICOM format file to different analyze formats. The DICOM software provides export utility that can export DICOM format file to different analyze formats like: JPEG, BMP, TIFF, PCX, PNG, WMF, TGA, EMF, etc. As JPEG format is lossy compression and in case of medical images a loss of single bit may cause to misdiagnosis of the disease, so we have selected TIF format which is lossless compression format. All images are converted into TIF for further processing.

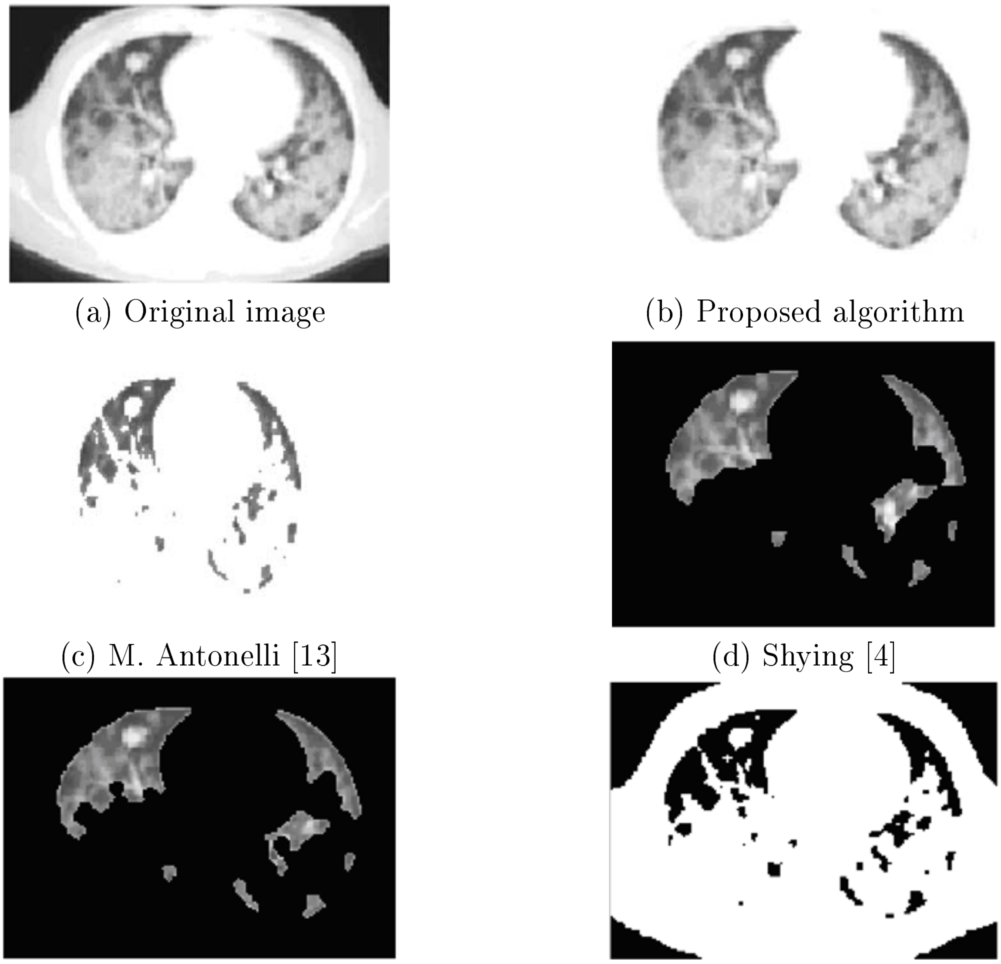
MATLAB environment have been used to implement the proposed system. To test our proposed system, we obtained datasets from Aga Khan Medical University, Pakistan. In this work, we have studied the performance of different segmentation techniques that are used in Computer Aided Diagnosis (CAD) systems using thorax CT Scans. These methods give good results on test databases of reasonable size. CT lung density is inclined by the factors such as subject tissue volume, air volume, image acquisition protocol, physical material properties of lung parenchyma and trans-pulmonary pressure. These factors make the selection of gray-level segmentation threshold difficult, as different thresholds are likely required for different subjects. The selection of optimal threshold by iterative method is one possible solution. But again due to different density of anatomical structures on each slice as discussed above, there is need of computing the threshold for each slice which is costly from computational point of view.

Table 1. Aga Khan university hospital Karachi

Patient Name	Patient-ID	No. of Slices	Slice Thickness (mm)	Age (years)
Anonymous19863	Anon19863	50	7.0	46
Anonymous20122	Anon20122	13	3.0	50
Anonymous21611	Anon21611	50	7.0	66
Anonymous21632	Anon21632	50	7.0	58
Anonymous21634	Anon21634	50	7.0	58
Anonymous21739	Anon21739	50	7.0	38
Anonymous21769	Anon21769	50	7.0	48
Anonymous21865	Anon21865	50	7.0	73
Anonymous21973	Anon21973	50	7.0	30

Thus there is a need of refined algorithm which can calculate a single threshold for whole database of a lone patient. Optimal thresholding performs better when lung volumetric differences are inevitable.

Figures 7(b) and 8(b) gives the result of our proposed technique on the test image. It is evident through observation that the proposed system produces much smoother results than the schemes that have been used earlier. It can be proved that nearly all previous techniques don't work for the images when there is overlapping of intensities in lung



(a) Original image (b) Proposed algorithm

(c) M. Antonelli [13] (d) Shying [4]

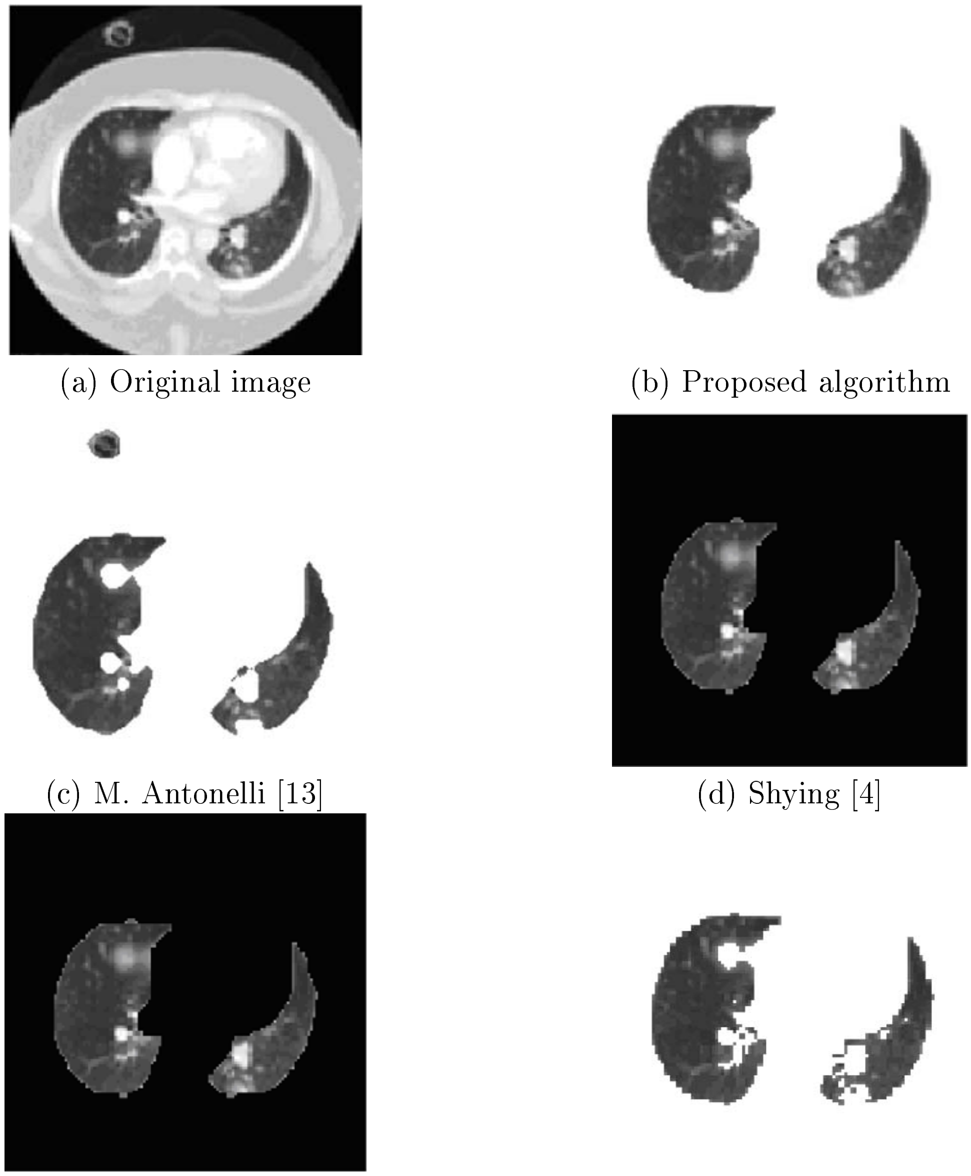
(e) Samuel [7] (f) FCM

Figure 7. Comparison of results of different algorithms with our proposed system

parenchyma and surrounding chest wall. However, out proposed technique has shown promising results on the different test cases as shown in figure. Results of our proposed method that are shown in Figures 7(b) and 8(b) demonstrate significant improvement. The figure shows that using our method, we are able to segment the image almost completely thus extracting lung part from the original image. There is also no loss of lung nodules in our proposed method. This is a good advancement for next stage of CAD system employed in lungs nodule detection and classification system.

Results of our proposed method that are shown in Figures 9(b), 9(d), 9(f) and 9(h) demonstrate significant improvement. The figure shows that using our method, we are able to segment the image almost completely thus extracting lung part from the original image. The essence of our segmentation method lies in its ability to fully automatically segment the lungs part from whole CT scan image. Search for higher density structures including nodules scattered in the lungs through sequentially declining threshold level.

4.1. **Why does GA perform better in this case?** Genetic algorithm (GA) is a search technique used in computing to find exact or approximate solutions for optimization and search problems. Genetic algorithms are categorized as global search heuristics. GA can be used effectively to find out the peaks and valley in bimodal class of images. GA is used for function optimization process and hence determines the global optimal solutions. In our case, input image is also bimodal class image. So like in the case of other images



(a) Original image (b) Proposed algorithm

(c) M. Antonelli [13] (d) Shying [4]

(e) Samuel [7] (f) FCM

FIGURE 8. Comparison of results of different algorithms with our proposed system

of this class, GA provides good results for our input image as well. For thresholding, iterative thresholding has been used by different authors to find out optimal threshold in medical images. They have used mean (average) as a criterion to find out threshold. However, this criterion fails in the case of noisy or corrupted images. In this case GA can perform better to find out optimal threshold. Usage of mean as a mechanism to find threshold is also a reason why other techniques provide poor results.

Genetic algorithms are implemented as a computer simulation in which a population of abstract representations (called chromosomes or the genotype of the genome) of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves toward better solutions. It is an iterative process that finds out better solution.

4.2. **Why other techniques fail?** Other techniques fail due to some reasons. The major problem occurs due to same intensity values of lung tissues available in lung parenchyma. In some cases, some methods fail for those images where there is overlapping of intensities in lung parenchyma and surrounding chest wall.

(a) Original image

(b) Proposed algorithm

(c) Original image

(d) Proposed algorithm

(e) Original image

(f) Proposed algorithm

(g) Original image

(h) Proposed algorithm

Figure 9. Proposed method results on different CT scan images

While testing the database slice by slice it was found that the spherical structuring element selected for morphological opening to get the hollow-free lung mask was almost working for the entire database of the patient. However for few slices it did not give promising results. Thus it was observed that ball size selected for morphological closing did not work for the entire database of all single patients. Some times a ball of specified size worked for one patient while it was not sufficient for another patient. Thus we have to vary the ball size patient by patient which makes the algorithm semi-automatic. Hence it can be concluded that the main limitations of some methods are the size of the ball selected for morphological operation.

While testing the database of each patient slice by slice, we observed that ball algorithm was almost working for each patient's database, but for few slices it did not give satisfactory results. Also applying the rolling ball algorithm for overcoming the loss of juxtapleural nodules is an additional processing time overhead. From Figure 12, we can see that the rolling ball algorithm fills the indentations of left side of the lung but does not completely fill the indentations produced in the right side of the lung. Also from Figure 13, we can observe that the ball algorithm includes the unnecessary areas as the lung region, (not actually the part of the lung). Thus from experiments we can conclude that the some methods has the following very serious short comings: (i) Fixed ball size (ii) Processing time overhead (iii) Inclusion of unnecessary areas as lung regions.

5. **Conclusion and Future Work.** We have described an adaptive method for automatic segmentation of pulmonary parenchyma. The "heart" of the proposed system, Genetic Algorithm that performs the adaptive thresholding process to determine the thresholds This is just the first step of a CAD system which is still under development. The results we obtained are comparable with those of other known methods. In addition, the proposed system has the advantage that it does not require any human expert intervention, nor any a priori information about the input image. The results achieved by applying the method to a database consisting of eight CT scans with a total of more than 240 digital images have been presented.

The next steps we are functioning on are nodule detection and false positive reduction. At the moment, our method is just at an experimental stage and needs to be evaluated through a double blind procedure by a number of radiologists, with comparison with their current method of nodule detection.

Acknowledgment. The authors, Mr. M. Arfan Jaffar would like to acknowledge the Higher Education Commission (HEC) of Pakistan and National University of Computer and Emerging Sciences, FAST for providing the funding and required resources to complete this work. It would have been impossible to complete this effort without their continuous support.

REFERENCES

[1] B. Zhao, G. Gamsu and M. S. Ginsberg, Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm, *Journal of Applied Clinical Medical Physics*, vol.4, no.3, 2003.

[2] N. A. Memon, A. M. Mirza and S. A. M. Gilani, Segmentation of lungs from CT scan imges for early diagnosis of lung cancer, *Proc. of World Academy of Science, Engineering and Technology*, vol.14, 2006.

[3] A. C. Cilva, P. Cezar and M. Gattas, Diagnosis of lung nodule using Gini coefficient and skeletonization in computerized tomography images, *ACM Symposium on Applied Computing*, 2004.

[4] A. El-Baz, A. A. Farag, R. Falk and R. L. Rocca, Detection, visualization and identification of lung abnormalities in chest spiral CT scan: Phase-I, *International Conference on Biomedical Engineering*, Cairo, Egypt, vol.12, no.1, 2002.

[5] M. A. Jaffar, A. Hussain, A. M. Mirza and Asmatullah, Fuzzy entropy and morphology based fully automated segmentation of lungs from CT scan images, *International Symposium on Intelligent Informatics*, Kumamoto, Japan, 2008.

[6] M. S. Brown, M. F. McNitt-Gray, N. J. Mankovich, J. G. GOldin, J. Hiller, L. S. Wilson and D. R. Aberle, Knowledge based segmentation of thoracic computed tomography images for assessment of split lung function, *Medical Physics*, vol.27, pp.592-598, 2000.

[7] S. Hu, E. A. Huffman and J. M. Reinhardt, Automatic lung segementation for accurate quantitiation of volumetric X-ray CT images, *IEEE Trans. on Medical Imaging*, vol.20, no.6, 2001.

[8] S. Hu, E. A. Hoffman and J. M. Reinhardt, Automatic segmentation of accurate quantitation of volumetric X-ray CT images, *IEEE Trans. on Medical Imaging*, vol.20, pp.490-498, 2001.

[9] A. El-Baz, A. A. Farag, R. Falk and R. L. Rocca, A unified approach for detection, visualization and identification of lung abnormalities in chest spiral CT scan, *Proc. of Computer Assisted Radiology and Surgery*, London, 2003.

[10] D. Zhang and D. J. Valentino, Segmentation of anatomical structures in X-ray computed tomography images using artifical neural networks, *Proc. of SPIE*, vol.4684, pp.1640-1652, 2001.

[11] J.-M. Kuhnigk, H. K. Hahn, M. Hindennach, V. Dicken, S. Krass and H.-O. Peitgen, Lung lobe segmentation by anatomy guided 3D watershed transform, *Medical Imaging, Proc. of SPIE*, vol.5032, pp.1482-1490, 2003.

[12] J. Wei, Image segmentation based on situational DCT descriptors, *Pattern Recognition Letters*, vol.23, pp.295-302, 2002.

[13] M. Antonelli, B. Lazzerini and F. Marcelloni, Segmentation and reconstruction of the lung volume in CT images, *ACM Symposium on Applied Computing*, 2005.

[14] S. G. Armato II, M. L. Giger and C. J. Moran, Computerized detection of pulmonary nodules on CT scans, *RadioGraphics*, vol.19, pp.1303-1311, 1999.

[15] N. A. Memon, A. M. Mirza and S. A. M. Gilani, Deficiencies of lung segmentation techniques using CT scan images for CAD, *Proc. of World Academy of Science, Engineering and Technology*, vol.14, 2006.

[16] J. M. S. Prewitt and M. L. Mendelsohn, The analysis of cell images, *Ann. N. Y. Acad. Sci.*, vol.128, pp.1035-1053, 1966.

[17] A. Hussain, M. A. Jaffar, A. M. Mirza and A. Chaudhry, Detail preserving fuzzy filter for impulse noise removal, *International Journal of Innovative Computing, Information and Control*, vol.5, no.10, pp.3583-3591, 2009.

[18] P. Kanungo, P. K. Nanda and A. Ghosh, Classification of objects and background using parallel genetic algorithm based clustering, *Electronic Letters on Computer Vision and Image Analysis*, vol.6, no.3, pp.42-53, 2007.

[19] M. A. Jaffar, A. Hussain, A. M. Mirza and A. Chaudhry, Fuzzy entropy and morphology based fully automated segmentation of lungs from CT scan images, *International Journal of Innovative Computing, Information and Control*, vol.5, no.12, pp.4993-5002, 2009.

[20] S. M. Smith and J. M. Brady, SUSAN – A new approach to low level image, *Proc. of Int. Journal of Computer Vision*, vol.23, no.1, pp.45-78, 1997.

[21] Y. Wang, Fuzzy clustering analysis by using genetic algorithm, *ICIC Express Letters*, vol.2, no.4, pp.331-337, 2008.

[22] R. Sammouda, J. Hassan and M. Sammouda, CT image analysis for early detection of lung cancer, *International Journal of Innovative Computing, Information and Control*, vol.4, no.11, pp.2847-2860, 2008.

[23] W. Liu, K. Yuan, J. Zou, S. Zhou, W. Chen, S. Jia and P. Xiao, Nonnegative tensor factorization for brain CT image retrieval, *International Journal of Innovative Computing, Information and Control*, vol.4, no.11, pp.2911-2918, 2008.

[24] R. Kubota, N. Suetake, E. Uchino, G. Hashimoto, T. Hiro and M. Matsuzaki, Polynomial-based boundary extraction of plaque in intravascular ultrasound image by using its local statistical information, *ICIC Express Letters*, vol.2, no.4, pp.387-393, 2008.