9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 24-25 August 2017, Budapest, Hungary

Lung tumor segmentation algorithm

Selin Uzelaltinbulat[a], Buse Ugur[b]*

[a]*Computer Engineering Department, Near East University, POBOX:99138, Nicosia, North Cyprus, Mersin 10, Turkey*
[b]*Biomedical Engineering Department, Near East University, POBOX:99138, Nicosia, North Cyprus, Mersin 10, Turkey*

Abstract

This paper is a development of an algorithm based medical image processing to segment the lung tumor in CT images due to the lack of such algorithms and approaches used to detect tumor where most of researches involve machine learning to solve such segmentation problem. The work involves different image processing tools which successfully achieved the required goals when combined and successively applied. The segmentation system comprises of different stages to finally reach its target which is to segment the lung tumor. Image pre-processing takes place first where some enhancement techniques are used to enhance and reduce noise in images. The next stage is where the different parts in the images are seperated to be able to segment the tumor in later stages. In this phase threshold was selected automatically which assures the right selection of all images since the tumor have different gray-levels intensities in each image. Another technique was also used here to remove the tumor from the thresholded image. Finally, the lung tumor is accurately segmented by subtracting the thresholded and the other image.

Keywords: Lung tumor; image processing techniques; segmentation; thresholding; image enhancement; accuracy; sensitivity; specificity.

1. Introduction

The lungs are the parts of our body that we use to breathe. They supply oxygen to the organs and tissues of the body. The lungs are divided into areas called lobes. The right lung has three lobes and the left lung has two. Lung cancer is the type of cancer which unchecks the growth of unusual cells either in one or in both the lungs. These anomalous cells do not perform the functions of healthy human cells and do not mature into normal cells. This

* Corresponding author. Tel.: +90-533-862-9324.
E-mail address: buse.ugur@neu.edu.tr

abnormality affects the proper regular functioning of the lung of supplying oxygen to the human body through blood. All though there are many advances in treatment procedures, the lung cancer which is at an advanced stage or late stage is not often easily curable Pataer et al. (2012).

There are two main types of primary lung cancer which one is known as the most common type, Non-small cell lung cancer (NSCLC) and the other one is Small cell lung cancer (SCLC) .The treatment of these two types of lung cancers is quite different as they behave in different ways and respond to treatments differently. The most common known of lung tumor is non-small cell lung cancer (NSCLC), as one of genuine ailments bringing on death for human beings. Computer-aided diagnosis and survival prediction of NSCLC is of great sign of cancer in diagnosis and treatment of people suffering from lung cancer Wang et al. (2014). The prognosis of lung malignancy is poor, in many nations only 10 percent of survival rate for about five-year. Two major types of NSCLC are adenocarcinoma which is about 40% and squamous cell carcinoma about 25 - 30% Wang et al. (2014). This study is a development of an image processing algorithm for accurate segmentation of the lung tumor.

The developed approach is purely image processing where it uses different image preparing tools to reach its target: lung tumor segmentation. The segmentation system can be mainly divided into three main stages. During the first stage, the CT images are smoothed and enhanced using some image enhancement techniques such as median filtering and image erosion. This helps in reducing the noise in an image and clearly smoothes its edges. Secondly, the image is to be partitioned so that each part of it is separated from the other. This is achieved by using thresholding which eliminated the intensities which are lower than an automatically selected threshold. As a result of thresholding, an image of tumor and clavicles is remained. The thresholded image also undergoes a technique which helps in removing all parts in the image keeping only the tumor. The third part of the segmentation system is to extract the tumor by subtracting the thresholded image from the cleaned image which results in an image that contains only a tumor. Experimentally, the system is tested using many tumor and non-tumor lung images it outperforms the related systems listed in the state of art Chen and Chen (2009).

2. The proposed system

In this study, a Lung Tumor Segmentation System is designed. The system is a model which found in image processing that combines different image processing techniques to accomplish its goal that is the segmentation of the lung tumors placed into the lung and segment them on the original image using threshold.

Lung Tumor Segmentation System is based upon different image processing techniques used for segmenting the lung tumor into a lung. The lung image database is an on-line CT image dataset available for the researchers in the field of digital image processing. Firstly, removing the noise from images, erosion and median filter are applied to the system respectively. Afterwards, thresholding is applied to the filtered image which is one of the most common techniques used in CT images. Threshold value is a specific value that turns a grayscale image into a binary image due to thresholding method. Then, binary areas are opened to remove small objects from threshold images. In last method, in order to create tumor cells alone, difference between filtered image and small objects removed from the images are taken into account.
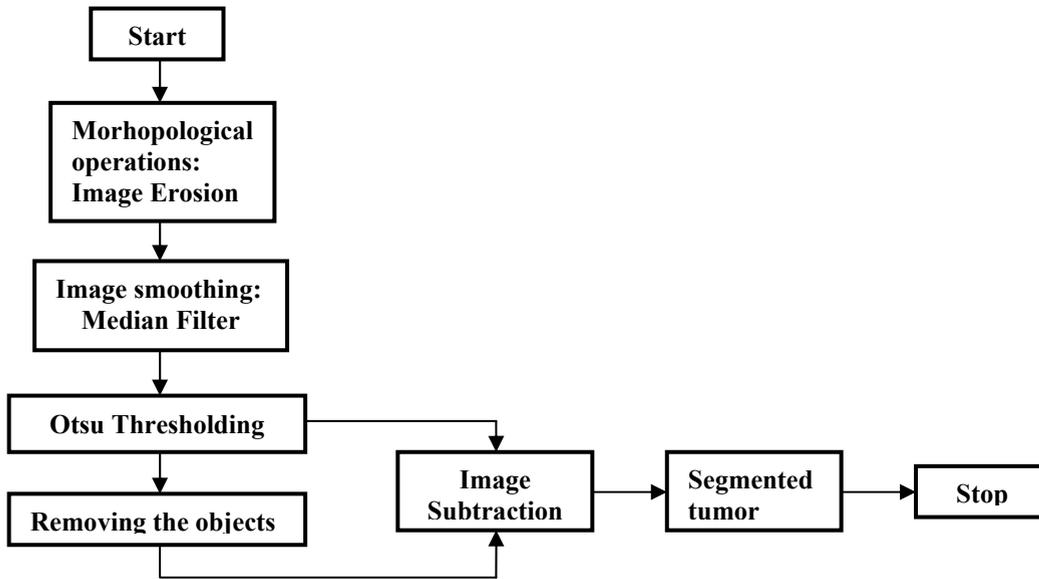
Fig. 1. Flowchart of the proposed system.

2.1. Dataset description

Computed tomography (CT) scan images are preferred to be used in this research, due to low noise and better clarity compared to X-Ray and MRI images. The lung image database has been chosen from NIH/NCI Lung Image Database Consortium (LIDC) which is an on-line CT image dataset available in the "Cancer Imaging Archieve" (Cancer Imaging Archieve, 2014) for the researchers in the field of digital image processing. The images used within this research are in .jpg format. Figure 4.2 shows samples of original CT images. It converts original DICOM images which are in grey color to JPEG format by removing the tint and saturation information while maintaining the luminance. The original size of the images was 512*512 and then it converted to 256*256 to get better quality. Then database was created by own system. There is 70 images which 50 of them has tumor and 20 of them has non tumor in total. This number is enough since the designed system is an image analysis and processing system.

Fig. 2. Samples of original CT images

2.2. Morphological operations

Morphology operations can be characterized as a gathering of image processing procedures that process images taking into account shapes. These morphological operations are taking into account applying an organizing component to an input image keeping in mind the end goal to make a yield image of the same size. In such operation, the estimation of every pixel in the yield image is taking into account an examination of the comparing pixel in the information picture with its neighbours. This is carried out by picking the size and state of the area. At

that point, we can build up a morphological operation that is delicate to particular shapes in the input image Assefa et al. (2013). The structure element is a matrix consists of 0's and 1's, where the 1's are called the neighbours. The value of each pixel in the output image is set according to a comparison of the corresponding pixel in the input image with its neighbours. It has many shapes according to its application. The most common and basic morphological operations are dilation and erosion. Dilation is to add pixels to the boundaries of objects in an image, while erosion is to remove pixels on object boundaries. The number of pixels that are added or even removed from the structure in an image depends on the size and shape of the structuring element that is used to process that image. In these morphological operations, the condition of any given pixel in the output image can be determined by applying a rule to the studied pixel and its neighbours in the input image Gonzalez & Woods (2004).
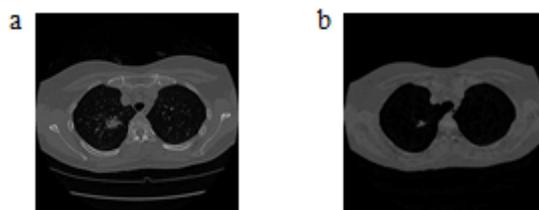

Fig. 3. (a) original image; (b) eroded image

The figure above shows the erosion of the input original grayscale image which contains tumor. It shows that the erosion helped in margining the tumor and also shrinking it to be separated from other tissues which help in later processing of the tumor. In this study, only erosion was used to shrink the images. In our case, or system, the "diamond" structure element with a "radius" of 2 is used to extract the background of the image Rao and Rao (2014).

2.3. Image filtering

Smoothing is an image processing technique used in order to reduce the noise in an image to produce less pixelated and clearer image. Most smoothing techniques are based on low pass linear filters. It is mostly based on the averaging technique of the input image or the middle (median) value technique. To perform a smoothing operation we will apply a filter to our image. The most common type of filters is the linear filters such as median filter which is used in our proposed system. This filter is used to reduce impulsive noise or the salt-and pepper in an image with preserving the useful features and image edges. Median filtering is a linear process in which the output of the being processed pixel is found by calculating the median of a window of pixels that surrounds that studied pixel Unnikrishnan et al. (2007). In other words, the median filter goes through each element of the image and replaces each pixel with the median of its neighboring pixels which are located in a square neighborhood (kernel) of 5*5 pixels around the evaluated pixel.
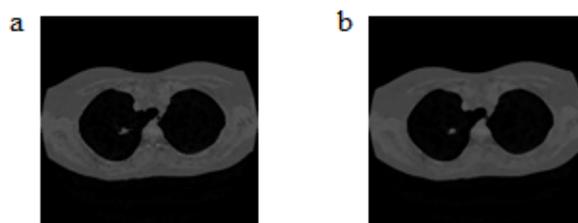

Fig. 4. (a) eroded image; (b) median filtered image

It can be seen that the median filter helped in setting and smoothing the edges of the tumor which makes it easy to be segmented in later techniques and also this filter helped to reduce impulsive noise or the salt-and pepper in the image which is clear in the Fig. 4 (a) with preserving the useful features and image edges by going through each

element of the image and replace each pixel with the median of its neighboring pixels which are located in a square neighborhood (kernel) around the evaluated pixel.

2.4. Threshold segmentation

Image segmentation by thresholding is a simple but powerful approach for segmenting images having light objects on dark background. Thresholding technique is based on image space regions i.e. on characteristics of image. Thresholding operation convert a multilevel image into a binary image i.e., it choose a proper threshold T, to divide image pixels into several regions and separate objects from background. The gray-level histogram of an image is usually considered as efficient tools for development of image thresholding algorithms. Thresholding creates binary images from grey-level ones by turning all pixels below some threshold to zero and all pixels above that threshold to one. If g(x, y) is a threshold version of f(x, y) at some global threshold T by Otsu (1979).
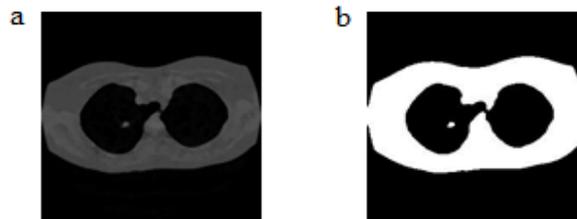

Fig. 5. (a) median filtered image; (b) thresholded image

Fig. 5 shows the thresholding process of a tumor images. It is noticeable that the tumor is correctly separated from other parts such as lungs and the other structure in the image. This technique allows the detection of the tumor as a separate object in the image. After this technique the tumor is easier to be segmented since it is a standalone object in an image of three structures or parts: lungs, cage, and tumor. For this study, Otsu thresholding is used due to its simplicity and the difference in graylevels intensities which requires different threshold different images. Therefore, the threshold is selected automatically for different images using Ostu thresholding. The figure above shows the thresholding process of a median filtered non-tumor image. It can be seen that the image edges were correctly detected after this technique which allows the separation of the lungs and the other parts in the image.

2.5. Small object removal

The images then undergo an operation called as area opening. This technique is used to remove the un-wanted components into an image. This technique is to remove form a binary image all the connected components (objects) that have pixels lower than a set value Helwan and Abiyev (2016). After applying this technique we could get a result image showing the segmented region of interest (tumor region) for abnormal images and a full black image representing the normal image.
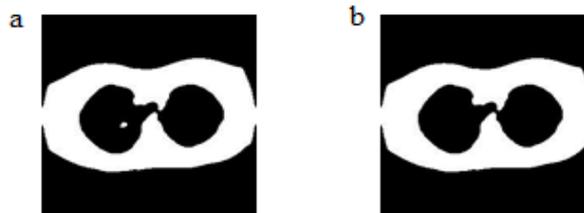

Fig. 6. (a) thresholded image; (b) object removed image

Fig. 6 shows the results of applying a technique or function the removes the objects that are connected to a tumor in order to make it easier to segment it. Usually this technique is used to remove the components connected to a tumor which are not related to it. However, in this study this technique is used to remove tumor to keep only the

other parts in the image. In later techniques this image will be subtracted from the one which has tumor in addition to other parts. As results the tumor will be obtained as a segmented image.

2.6. Image subtraction

The pixel subtraction operator takes two images as input and produces as output a third image whose pixel values are simply those of the first image minus the corresponding pixel values from the second image. It is also often possible to just use a single image as input and subtract a constant value from all the pixels. Some versions of the operator will output the absolute difference between pixel values rather than the signed output. The subtraction of two images is performed straightforwardly in a single pass. The output pixel values are given by

$$R(m,n)=P(m,n) - Q(m,n) \tag{1}$$

Or if the operator computes absolute differences between the two input images, then:

$$R(m,n) = | P(m,n) - Q(m,n) | \tag{2}$$

Or if it is simply desired to subtract a constant value *C* from a single image, then:

$$R(m,n)=P(m,n) - C \tag{3}$$

If each pixel value corresponds to an n-tuple instead of a number (for instance, the color image would have each pixel as a vector of three values corresponding to red, blue, and green components), then such pixels are operated on by subtraction at each corresponding vector element to produce the output value Dimililer et al. (2016).
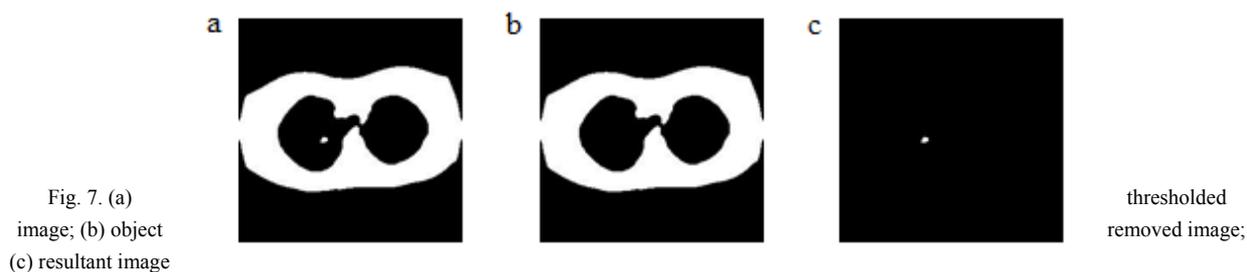

Fig. 7. (a) image; (b) object (c) resultant image thresholded removed image;

The fig. 7 shows the subtraction of the object removed image (b) from the thresholded image (a). As a result of this subtraction operation, the tumor is obtained and segmented in (c).

3. System performance

The developed system will automatically segment any lung tumor in a lung image which contains many other components, or structures, such as clavicles and lungs. Thus, the images were pre-processed first before starting the segmentation process. This allows the tumors to be separated from other components in images. This then makes it easier to segment tumors using thresholding, and some other image processing operations. To test the system we had to apply some images containing tumors and others with no tumor. 50 lung tumor images and 20 with no tumor images were used to evaluate the effectiveness of the system. The measurements that were used for evaluating the developed system are the accuracy, sensitivity, and specificity.

Accuracy is the total number of correctly segmented tumors over the total number of images

$$ACC = CS \tag{4}$$

Where CS is the number of corrected segmented images, and T is the total number of images. ACC is the accuracy of the system.

Sensitivity was also used as a measure to evaluate the segmentation system. It can be defined as a measure of true positive which are the correctly identified areas.

It can be defined as the following:
$$Sensitivity = [TP/TP + FN] * 1 \qquad\qquad (5)$$

Where TP is true positive this represents the images that they were segmented and that they have tumor.
FN represents the images that were segmented and they have no tumor.
Another parameter was used for system performance evaluation is the specificity. It can be defined as:
$$Specificity = [TP/TP + FN] * 1 \qquad\qquad (6)$$

TN are the images that have no tumor and they were segmented correctly.
FP are the images that were not segmented and they have tumors.

Table 1 & 2 show the system performance, where the accuracy achieved is 97.14% as a total accuracy obtained after testing the system by tumor and non-tumor images. The sensitivity and specificity of the system were 100% and 96% respectively.

Table 1. The segmentation ratio of the system

	Number of Images	Correct segmented images	Segmentation ratio
Non Tumor	20	20/20	100%
Tumor	50	48/50	96%
Total	70	68/70	97%

Table 2. The performance ratio of the system

	Number of images	Accuracy	Sensitivity	Specificity
Total	70	97.14%	96%	100%

3.1. Result discussion

This study is to develop a new image processing approach for the segmentation of the lung tumor compared to other studies. In this study different image processing techniques are used. Firstly, images where are enhanced by using image smoothing techniques like erosion and median filtering. Then, the tumor is segmented by using Otsu thresholding. This techniques was used because of the difference in Gray-levels intensities of images which require different thresholds for different images. The useless connected pixels were removed from threshold image to make it ready for the last technique. The last step is the extraction of the tumor which is the main part of the system. The threshold images, which contain tumor, and other lung structures and the other images which only contain lung structures without tumor. This non tumor image which is clean was subtracted from the threshold images in order to extract tumor.

Many challenges were encountered when developing this system. One of them is the selection of the right technique for the final tumor extraction. Thresholding was first selected to be used in which different thresholds were used to segment different areas of the image. This technique had some advantages, however some tumors were not segmented using it. In addition some images were detected to have tumor but actually they don't which decreases the accuracy and sensitivity of the system. The reason for this incorrect segmentation was the difference in

intensities of tumor and other parts in each image. Thus, Otsu's thresholding was used where the threshold is set automatically depending on each image intensity. This technique was definitely good in segmenting an image, by keeping the tumor and clavicle in an image since they have almost the same intensities. Therefore, another technique was used to remove the tumor from the images keeping only the clavicle. After that the two images were subtracted and the tumor was segmented.

Experimentally, the developed approach showed a great efficiency in segmenting tumors in a CT image using the above discussed technique. This developed lung tumor segmentation approach outperforms many of the researches listed in the state of art in terms of accuracy, sensitivity, and specificity.

4. Conclusion

The experimental results show that the techniques used in this system were efficient enough to segment the tumor in many images since they consider the image's intensities variance separately for each image It found that the developed system was capable of segmenting tumors of different tumor and non-tumor images of the used database by achieving high accuracy of 97.14%

Acknowledgements

References

1. Pataer A., Kalhor N., Correa A. M., Raso M. G. Erasmus, J. J. Kim, E. S. & Vaporciyan, A., 2012. A Histopathologic response criteria predict survival of patients with resected lung cancer after neoadjuvant chemotherapy. *J Thorac Oncol*; 7, 825-832.

2. Wang H., Xing F., Su H., Stromberg A., & Yang L., 2014. Novel image markers for non- small cell lung cancer classification and survival prediction. *BMC bioinformatics* 15, 4-15.

3. Chen Y. B. and Chen O. T. C., 2009. Image Segmentation Method Using Thresholds Automatically Determined from Picture Contents *Eurasip journal on image and video processing*.

4. Helwan A.& Abiyev R., 2016. Shape and Texture Features for the Identification of Breast Cancer. In *Proceedings of the World Congress on Engineering and Computer Science*.

5. Assefa, M., Faye, I., Malik, A. S., & Shoaib, M., 2013. Lung nodule detection using multi-resolution analysis. *In Proceedings of the International Conference on Complex Medical Engineering,* pp. 457-461.

6. Gonzalez and Woods. Digital Image Processing, New York: CRC Press, 2004.

7. Rao, K.M.M., Rao, V.D.P., 2014. Medical Image Processing. *In Proceedings of Image Analysis and Interpretation* , pp. 6-8.

8. Dimililer K., Ever Y. K., & Ugur B., 2016. ILTDS: Intelligent Lung Tumor Detection System on CT Images. *In Proceedings of The International Symposium on Intelligent Systems Technologies and Applications,* pp. 225-235.

9. Unnikrishnan R., Pantofaru C.,Hebert M., 2007. Towards objectiveevaluation of image segmentation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell ,* 29, 929–944.

10. Otsu N., 1979.Thresholds selection method form grey-level histograms. *IEEE Trans Syst Man Cybern* , *9*