South Asian Journal of Engineering and Technology Vol.2, No.17 (2016) 10–17

ISSN No: 2454-9614

K-Means Clustering using Fuzzy C-Means Based Image Segmentation for Lung Cancer

K. Kaviarasu , V. Sakthivel*

Department Of Computer Science and Engineering K.S.Rangasamy College of Technology, Namakkal, Tamilnadu, India.

*Corresponding Author: K. kaviarasu

E-mail: kaviarasu91.k@gmail.com

Received: 09/11/2015, Revised: 12/12/2015 and Accepted: 04/03/2016

Abstract

Lung lesion segmentation refers to the process of partitioning an image into mutually exclusive regions. This study gives a new approach to K-means clustering technique (K-CT) integrated with Fuzzy C-means algorithm for lung segmentation. In the study, large number of images with various types of segmentation was selected and examined. It is followed by thresholding and level set segmentation stages to provide an accurate region growing detection. The method starts with lung segmentation based on region growing and standard image processing techniques. K-means clustering technique Segmentation is an important process to cluster information from complex lung lesion. Image Segmentations refers to the process of fuzzy c means an image into groups of pixels which are standardized with some criteria. Fuzzy C-means algorithms are area oriented instead of pixel oriented. The result of lung segmentation is the splitting up of the image into connected region growing. Thus segmentation is concerned with dividing an image in to meaningful regions. The proposed technique can get benefits of the K-means clustering for lung lesion segmentation in the aspects of minimal computation time. In addition, it can get advantages of the Fuzzy C-means in the aspects of accuracy. The method starts with lung segmentation based on region growing and standard image processing techniques.

Reviewed by* **ICETSET'16 *organizing committee*

Keywords: K-means clustering technique, Fuzzy C-means Algorithm, Image Segmentation, lung cancer.

1. Introduction

Image segmentation occurs as a set of regions that collectively covers the entire image. Therefore, medical image segmentation plays a significant role in clinical diagnosis. It can be considered as a difficult problem because medical images commonly have poor contrasts, different types of noise, and missing or diffusive boundaries. The anatomy of the brain can be scanned by lung cancer image scan or computed tomography scan. The extraction is more comfortable than k-means cluster for diagnosis. It is not affect the human body because it does not use any radiation. It is based on the magnetic field and radio waves. On the other hand, lung cancer is one of the leading causes of death among people.

It is evidence that the chance of survival can be increased if the cancer is detected correctly at its early stage. In most cases, the physician gives the treatment for the strokes rather than the treatment for the cancer. Therefore, detection of the cancer is essential for the treatment. The lifetime of the person who affected by the lung cancer will increase if it is detected early. Thus, there is a need for an efficient medical image segmentation method with some preferred properties such as minimum user interaction, fast computation, accurate, and robust segmentation results.

We used image segmentation techniques based on clustering to detect the lung lesion and calculating the cancer area. We developed a novel image segmentation approach, called K-means integrated with Fuzzy C-means, for abnormal images.

We integrated K-means clustering algorithm with the Fuzzy C-means algorithm to overcome the limitations and get benefits of them. After clustering stage, the extraction of the segmentation is done automatically without user interaction by using thresholding and level set methods to contour the segmentation area.

It describes the image datasets used in this work. It also shows the proposed image segmentation system based on clustering. K-means algorithm based segmentation, local standard deviation guided grid based coarse grain localization, and local standard deviation guided grid based fine grain localization. The extraction of the lung cancer segmentation region from the processed image requires the segmentation of the brain MRI images to two segments.

The last stage of our proposed technique is calculating the cancer area in the processed image. K-means algorithm can detect segmentation faster than Fuzzy C-means. However, Fuzzy C-means predicted cancer cells that are not predicted by K-means algorithm.

The proposed technique gives an accurate result as compared to the K-means algorithm. Even though, original Fuzzy C-means algorithm yields good results for segmenting noise free images, it fails to segment noisy images. Therefore, we get benefits from integrating these two algorithms to reduce the number of iterations, which affects execution time and give an accurate result in cancer detection.

2. Related works

A seed point is selected automatically in the lung lesion regions we obtained. The multi-constraints are proposed to control the lesion segmentation. As the intensity of vessels and visceral pleura is close to that of the lung lesion, they are sometimes considered to be part of the adjacent lesions. These tissues are giant obstacles for lesion segmentation. Lung lesion refinement, a lung lesion refining method is used to get rid of the incorrect vascularised regions and other tissues **[1, 2].**

Medical image processing has experienced dramatic expansion, and has been an interdisciplinary research field attracting expertise from applied mathematics, computer sciences, engineering, statistics, physics, biology and medicine **[3].**

Computer-aided diagnostic processing has already become an important part of clinical routine. Accompanied by a rush of new development of high technology and use of various imaging modalities, more challenges arise; for example, how to process and analyses a significant volume of images so that high quality information can be produced for disease diagnosis and treatment.

This classifier training involves collection of a large Number of image data sets and then extraction of a large number of features from each data set **[4, 5].**In an imaging research setting, there are typically many variables being investigated, for example, variables in lung CT image acquisition are collimation, tube current, reconstruction algorithm, and breathing state **[6]**.

For each different imaging protocol, there are also many different quantitative features being extracted to search for the optimal combination of imaging parameters and features to characterize the disease process or clinical question to be answered. This requires a complex data model and queries **[7].**

Once the meaningful variables are selected for use by system to perform a particular diagnostic, the queries become less complex since only those variables need be retrieved. Therefore, an accurate image segmentation method, other than the conventional region of interest analysis, is often needed for diagnostic or prognostic assessment **[8, 9].**

This functional characterization has a higher potential for proper assessment due to recent advances imaging. Indeed, this higher potential has renewed interest in developing much more accurate segmentation methods to turn hybrid imaging systems into diagnostic tools. Specifically, after the adoption of multi-modal imaging systems, optimal approaches for precise segmentation and quantification of metabolic activities were crucial **[10].**

3. Proposed System

Lung segmentation refers to the development of subdividing a digital image into multiple sections. The objective of segmentation is to alteration the illustration of an image to be more expressive and easier to examine. It is used in order to locate objects and boundaries in images. There are some lung segmentation systems which use K-means algorithm for detecting mass region growing based segmentation in image. The K-means algorithm is fast and simple to run on large datasets, but it suffers from incomplete detection of lung segmentation, mainly if it is a malignant segmentation. On the other hand, other systems use Fuzzy C-means algorithm because it retains the more information of the original image to detect malignant region based image segmentation cells accurately compared to the K-means. These systems are sensitive to noise and outliers and they take long execution time.

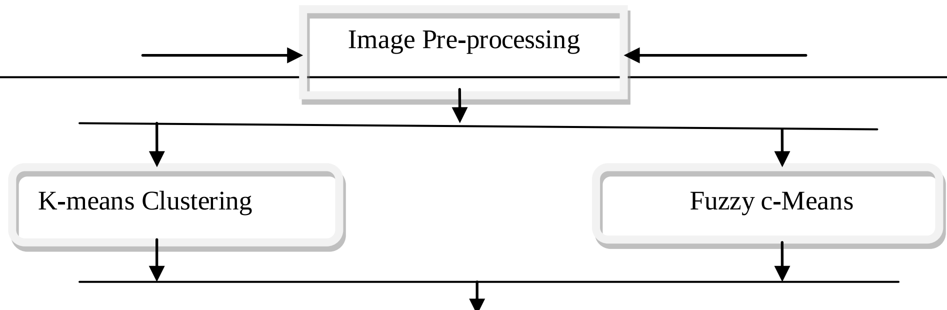

Figure 1. Proposed system Architecture

3.1 K Means Clustering For Lung Lesion Based Segmentation

Automated system (detection) of lung cancer through k-ct is basically called computer-aided diagnosis system. The k-ct scan provides highly accurate reconstruction of the original image, the valuable outlook and accuracy of earlier lung cancer detection. It consists of two or more stage. In the initial stage pre-processing has mandatory after that steps post-processing, segmentation are required. Then detection strategies and other information, feature extraction, feature selection, classification, and performance analysis are compared and studied. Pre-processing techniques are used to improvement of image quality based segmentation and remove small artefacts and cancer for the accurate detection of the undesired regions. Post-processing issued to segment with different strategy the lung lesion segmentation from the k-ct of lung segmentation images.

Algorithm:

Step1: read the input image or grayscale lung cancer image.

Step2: converts input color image in to grayscale image which is done by forming a weighted sum of each three (rgb) component, eliminating the saturation and hue information while retaining the luminance and the image returns a grayscale color map.

Step3: extraction to the multidimensional array with the multidimensional filter. Each element of the output an integer or in array, then output elements that exceed the certain range of the integer type is shortened, and fractional values are rounded.

Step4: add step2, step4 image and an integer value 45 and pass it in to a median filter to get the resultant enhanced image.

Step5: computes a global threshold that can be used to convert an intensity image (step5) to a binary image with a normalized intensity value which lies in between range 0 and 1.

Step6: compute watershed segmentation by mat lab command watershed (step6 image).

Step7: compute the morphological operation by two mat lab command improve and imitate and strel with arbitrary shape.

Step8: store the size of the step 8 image into var1 and var2 i.e. No. Of rows and column in pixels by [var1 var2] =size (step8 image)

```
for i=1:1:var1 do
    for j=1:1:var2 do
        if step8 image (i,j) == 1 do
            step2 image (i,j) = 255
        else do
            step2 image (i,j) = step2 image (i,j) * 0.3
        end if
    end for
end for
```

Step9: convert in to binary image and traces the exterior boundaries of objects, as well as boundaries of holes inside these objects, in the binary image and into an RGB color image for the purpose of visualizing labelled regions.

Step10: show only cancer portion of the image by remove the small object area.

Step11: compute edge detection using sobel edge detection technique.

3.2 Fuzzy C-Means Based Region Growing

Region growing based image enhancement is the development of digital image excellence short of any knowledge about the unique source image degradation. The fuzzy means enhancement methods mainly divide into fuzzy c-means segmentation, direct and indirect methods. In direct method is to show the contrast of the image and then improve the agreement but in the indirect method difference of the image is not important. Under-enhanced when some regions of the image may be over-enhanced are the great difficulty of the contrast improvement methods.

Algorithm:

Step1: original lung image

```
if image name==0
    return
end
```

Step2: preprocessing of an image

Step3: histogram of an image

j = histeq (image);

Step4: threshold segmentation with fuzzy c-means algorithm

m = zeros (size (i, 1), size (i, 2)); (zeros-create an array for all zeros)

m (125:150,145:160) =1;

i = imresize (i, .5); --->make image smaller

m = imresize (m, .5); --->fast computation

seg = region_seg(i, m, 290);

i2 = image – background portion;

i3 = imadjust (i2);

level = graythresh (i3);

bw = im2bw(i3,level); (im2bw-convert image to binary image based on threshold)

bw = bwareaopen(bw, 50); (bwareaopen- remove small objects)

Step5: extracted the segmented image

Step6: calculate an area of extracted tumor

Area of the lung cancer = a*total

4. Results and discussion

We tested our various lung lesion image segmentation algorithm based on the proposed k-means clustering technique (k-ct) integrated with fuzzy c-means image segmentation methods. To prove the robustness of our method, the new feature has been tested on a various level in different lung cancer images. To show how the segmentation of region growing works to classify the results, we give example images that are classified differently by single segmentation methods and by the multi-scale fuzzy c means image segmentation. By the results enhance classification lung images due to blur, smoke and other dry particles reduce visibility for distant regions by causing a distinctive gray hue in the captured images. However, our technique has been successfully tested as well for a slightly different case improve the classification technically it appears as a dense cloudy regions based image extracting by various conditions are matched to the relevant images, and absorbed by various segmentations level. We assume that the input lung images are colour images and the images may contain achromatic objects. With more complex techniques. Compared with the techniques of and our techniques' able to better preserve the fine transitions in the classification regions without introducing unpleasing artifacts. Moreover, the technique of classifier produces results with over-saturated colours.

4.1 approximation result

Half of the images in the dataset were used for training and the other half of the images were used for testing. for each segmentations of each image, the sampling result shows extraction result e1,e2and the dense sampling are used respectively, the k-means clustering features were extracted as greater enhancement. The table shows that, overall the detected salient regions sampled images correspond to the foreground and in the final detection and extraction image much of the background is filtered out saliency filtered result.

methods of image segmentation for lung cancer	**Detection rate (%)**
K-means cluster, fuzzy c-means(k-ct)	**>93.45**
Back-off mechanism	**>90.50**
Toboggan based image segmenting	**>85.75**

Table 1: image segmentation approximation results

In this approximation, we empirically use three representative segmentations: = {s1, s2, ss}, where ss is the maximum scale at which the diffusion process converges, s1 ands2 is a mid-image segmentation which is set to into (s/3).

• When the image is larger than s, there is almost no change in the diffused image. At scale si, foreground/background segmentation is completed.

• The inclusion of the original image corresponding to lung cancers1 in s2 can provide a correction if the foreground is incorrectly filtered out and using the image at scale ts alone is sufficient to obtain a correct classification result.

• The mid-scale tm is a compromise between smoothing the background and preserving the foreground. Although there are no clear cut criteria to pick the mid-scale tm, the experiments show that the use of ss improves the classification.

5. Conclusion:

To conclude, we have proposed k-means clustering technique (k-ct) integrated with fuzzy c-means image segmentation, to enhance the classification of images using lung cancer diffusion, and determining the distribution parameters using the scanner detection results. We have further applied this new method to image classification. The k-means clustering technique driven nonlinear multi-scale space preserves and even enhances important image local structures, such as lines and edges, at large scales. Multi-scale information has been fused using a weighted function of the distances between images at different scales. The k-ct representation can include information about the background in order to improve image classification results.

References

[1] H. B. Mitchell, Image Fusion: Theories, Techniques and Applications. 2010.

[2] N. Hautiere, J.-P. Tarel, D. Aubert, and E. Dumont, “Blind contrast enhancement assessment by gradient ratioing at visible edges,” 2008.
[3] K. Nishino, L. Kratz, and S. Lombardi, “Bayesian defogging2012.
[4] Vision Based Hand Gesture Recognition for Human Computer Interaction: A Review .IADIS International Conference Interfacesand Human Computer Interaction.2008
[5] Tirthankar Dasgupta ,Sambit Shukla, Sandeep Kumar A Multilingual Multimedia Indian Signlanguage Dictionary Tool..2008
[6] R.M. Haralick, and L.G. Shapiro (1985) Survey: Image Segmentation Techniques. Computer Vision,Graphics, and Image Processing, Vol. 29, pp. 100-132.
[7] D. Betal, N. Roberts and GH. Whitehouse, "Segmentation and numerical analysis of micro calcifications on mammograms were using mathematical morphology", 1997.
[8] E. Scheirer and M. Slaney, “Construction and Evaluation ofa Robust Multi feature Music/Speech Discriminator”, 1997.
[9] K. El-Maleh, M. Klein, G. Petrucci and P. Kabal,“Speech/Music Discrimination for Multimedia Application”, 2000.
[10] Diego Cast´an, Alfonso Ortega, Eduardo Lleida, “Speech/Music classification by using the C4.5 decision tree algorithm”, 2014