

# A New Clustering-based Algorithm for Image Segmentation

Shermin Sarrafha

Islamic Azad University  
South Tehran Branch,  
Tehran, Iran

e-mail: Shermin\_shs {at} yahoo.com

Maryam Khademi

Islamic Azad University  
South Tehran Branch,  
Tehran, Iran

Mehdi Heidari

School of Electrical & Computer  
Engineering, College of Engineering,  
University of Tehran,  
Tehran, Iran

**Abstract**— Segmentation is a crucial task in medical image processing. One of the greatest significant usages of image processing in medicine is recognizing cancerous or leisured tissues in MRI images. In this paper a new Clustering-Based algorithm is proposed for finding leisured segments in cancerous breast MRI images. The algorithm recursively detect the similar pixels to the current pixel and place them in the desired cluster. A threshold is defined due to the train data and if two adjacent pixels' colors differ less than the threshold, they are considered as similar. Experimental evaluations show that the algorithm is able to recognize the leisured areas with 94.3% accuracy and high speed.

**Keywords**- image processing; segmentation; breast MRI; clustering algorithm

## I. INTRODUCTION

Significance of the visual aspect to medical science is beyond doubt. Using medical image processing applications and tools, enable specialists easily share research data, remote analyses and as a result, diagnose, monitor, and treat medical disorders more accurate than before [1]. Today, one of the most disastrous diseases which human beings come across is cancer. This misadventure is a great stimulus for scientists to investigate on it more than before. Utilizing image processing methods for recognizing leisured tissues in different organs is one of the important objectives. Significance of this process obliges investigators to work hard on it for gaining highest possible accuracy and speed. An appropriate algorithm with acceptable accuracy is a great help for physicians to recognize borders in areas that are not easily recognizable with naked eye.

Breast MRI is a test used to detect breast cancer and other abnormalities in the breast. Breast MRI usually is performed when the doctor needs more information than a mammogram, ultrasound or clinical breast exam can provide. In certain situations, such as when a woman has a very high risk of breast cancer, breast MRI may be used; so accurate analysis of these images is a critical requirement [2].

In this paper we propose a new clustering based algorithm for image segmentation for separating leisured tissues in breast MRI images.

The rest of the paper is organized as follows. Section 2 describes previous works of investigators. Section 3 introduces a precise definition of a new clustering based algorithm. The evaluation algorithm is provided in Section 4, followed by the experimental results in Section 5. Finally the paper will be ended in section 6 by conclusions and future works.

## II. LITERATURE REVIEW

A magnetic resonance imaging instrument or MRI scanner uses powerful magnets to polarize and excite hydrogen nuclei in water molecules in human tissue. By this way it radiates the detectable signal which produces body image. While the proton of the hydrogen atom is so ubiquitous, it returns a large signal and allows the excellent soft-tissue contrast achievable with MRI. An accurate image processing technique is able to distinguish these contrasts and separate considering tissues in the image [3].

Hoyer et al [4] in 1979 were the first who proposed a method for separating tissue of the breast from background in MRIs. Their method was based on specifying threshold on histograms. In 1995 Beyk [5] et al segmented mammograms by considering in surface diagram. Another method for image segmentation is active path. Ojela [6] et al used this method in 1999. In 2000 Raba [7] and his colleagues used this method for extracting breast border and at last in 2004 Ferari [8] et al proposed a final algorithm for segmentation.

In 2005, Shadroo et al [9] proposed an algorithm which was based on multi-fractal estimation and fuzzy clustering. Her algorithm had more than 90% accuracy. After that Tavakolkhah [10] et al proposed a method which was able to separate benignant and malignant glands in mammograms with 85% accuracy. The method was based on fuzzy thresholds achieving from neural networks. Shahedi [11] proposed another algorithm which was working by best threshold finding and bind extraction in each loop. This algorithm had a high accuracy too.

### III. OUR APPROACH

Mathematical models are the foundation of biomedical computing. Basing those models on data extracted from images continues to be a fundamental technique for achieving scientific progress in experimental and biomedical research and as a result in modern medical imaging, we need to integrate these models in order to generate new useful algorithms [12].

On the other hand, the image processing algorithms which are used in recognizing cancerous and leisedured tissues are being used more and more so leading efforts to improve their accuracy and reliability are highly required.

In this paper we propose a new clustering based algorithm for image segmentation which is used in separating leisedured tissues in breast MRI images. The algorithm is able to recognize leisedured segments through a process of pixel labeling. By sampling from clusters, the leisedured areas can be distinguished and introduced to physicians. By this way the physician is able to recognize all troubled parts even the ones which may be unnoticeable with naked eye.

The special usage that we have considered in this study like many other usages, needs data preprocessing. To do so, at first we should make sure that data is scaled to the range 0 ~ 1.0. To avoid problems with data types, especially when working with predefined image processing functions, it is advisory to normalize data in this scale. For achieving this goal we have to convert the image array to double precision.

After that, we have to convert the true color image RGB to the gray scale intensity image in order to have just one index for each pixel. In gray scale images the value of each pixel is a single sample that carries only intensity information.

The proposed algorithm compares each pixel to its neighbors in order to find out whether they are cluster mates or not; so it is probable that we come across this scenario: consider a sequence of pixels  $\{i_1, i_2, \dots, i_n\}$  which  $i_1$  is adjacent with  $i_2$ ,  $i_2$  is adjacent with pixel  $i_3$  and so on, on the other hand  $i_1$  is not cluster mate with  $i_n$  but difference of each two neighbors is less than the considered threshold, it means the algorithm decides  $i_2$  is cluster mate with  $i_1$  and  $i_3$  is cluster mate with  $i_2$  and as a result  $i_n$  is cluster mate with  $i_1$ . The spectrum property of the nodes is the reason of wrong decision of the algorithm. So we have to put down the spectrum property of pixels and this goal is reached by discretization. Here discretization refers to the process of converting continuous numbers (which express intensity) to predefined discrete numbers. This process can be carried out by rounding each number to the nearest target number (target numbers are 5, 15, 25...). For example 22 should be round to 25. Although the results seem to be less satisfactory after this step, it should be pointed out that manipulating the digits in such scales are ignorable.

It should be mentioned that the threshold is tuned due to the train data. We have totally 600 samples where 400 of them are used as train data and the rest for test. In training, we calculate the average of f-measure of train data with different thresholds

and finally consider threshold as 0.05. “Fig. 1” shows the tuning diagram.

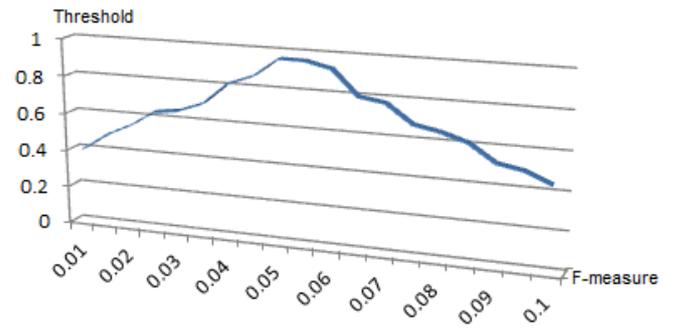


Figure 1. The tuning diagram

Now the image is ready for the main algorithm. The algorithm works with recursive logic. It loops through all of the unlabeled pixels by starting from the upper-right pixel and ends up reaching the last one in the image. In each iteration, a recursive function with two input arguments should find all cluster mates of the input. The current pixel and a new cluster number are input arguments. This function assigns the input cluster number to the input pixel. At the next step it compares the input pixel with its neighbors and if any neighbor color is near the color of input pixel within the threshold, the function calls itself recursively with these input arguments: that similar neighbor as first input argument and the same input cluster as second input argument. It should be pointed out that the neighbors for a given pixel are those that border it: up, down, left, right, upper-left, upper-right etc. and are maximum 8 pixels. The algorithm as “(1)” The flowchart of the algorithm is shown in “Fig. 2”.

### IV. EVALUATION METHOD

The performance of our algorithm was evaluated using f-measure. Let the correct clustering be  $C_1 \dots C_k$  and the nodes of the clustering found by the algorithm be  $G_1 \dots G_k$ . Then for each correct cluster  $C_i$ , the f-measure of that cluster is as “(1)”.

$$F(i) = \max_{j=1}^k ((2P_j R_j) / (P_j + R_j)) \quad (1)$$

While  $P_j$  stands for precision of cluster  $G_j$  and is the fraction of retrieved pixels that are relevant and is shown in “(2)” and  $R_j$  stands for recall of cluster  $G_j$  and is the fraction of relevant pixels that are retrieved and is shown in “(3)” [13].

$$P_j = |C_i \cap G_j| / |G_j| \quad (2)$$

$$R_j = |C_i \cap G_j| / |C_i| \quad (3)$$

### V. EXPERIMENTAL RESULTS

To evaluate the accuracy of the proposed algorithm, we developed an application using MATLAB. In order to evaluate the method, 200 breast MRI images from Cancer Imaging

Archive were input to the application. Since it was probable that some suspected regions stay out of sight, the images were labeled by 3 different experts and the union of all 3 was considered as gold standard. The overall average of 94.3% for f-measure has been gained for the sample pack. “Table 1” shows the results of first 50 samples.

“Fig. 3” shows one of this test images and “Fig. 4” indicates the final result that was gained by the proposed algorithm.

Table 1. The Results of First 50 Samples

Image	Precision	Recall	F-measure
img1	0.961	0.952	0.956478829
img2	0.895	0.921	0.907813877
img3	0.92	0.929	0.924478096
img4	0.942	0.951	0.946478605
img5	0.935	0.943	0.938982961
img6	0.955	0.966	0.960468506
img7	0.902	0.92	0.910911087
img8	0.95	0.961	0.955468341
img9	0.943	0.901	0.921521692
img10	0.939	0.941	0.939998936
img11	0.879	0.949	0.912659737
img12	0.97	0.921	0.944865151
img13	0.901	0.911	0.905972406
img14	0.925	0.944	0.934403424
img15	0.91	0.891	0.900399778
img16	0.889	0.952	0.919422053
img17	0.931	0.933	0.931998927
img18	0.941	0.949	0.944983069
img19	0.911	0.93	0.920401955
img20	0.989	0.97	0.979407861
img21	0.899	0.882	0.890418866
img22	0.955	0.92	0.937173333
img23	0.943	0.942	0.942499735
img24	0.937	0.89	0.912895457
img25	0.946	0.938	0.941983015
img26	0.961	0.943	0.951914916
img27	0.861	0.9	0.880068143
img28	0.96	0.961	0.96049974
img29	0.912	0.892	0.901889135
img30	0.977	0.981	0.978995914
img31	0.931	0.954	0.942359682
img32	0.951	0.961	0.955973849
img33	0.88	0.882	0.880998865
img34	0.981	0.965	0.972934224
img35	0.899	0.91	0.904466556
img36	0.944	0.942	0.94299894
img37	0.921	0.889	0.904717127
img38	0.91	0.901	0.905477637
img39	0.95	0.957	0.953487153
img40	0.921	0.934	0.927454447
img41	0.882	0.869	0.875451742
img42	0.933	0.954	0.943383148
img43	0.989	0.971	0.979917347
img44	0.943	0.945	0.943998941
img45	0.976	0.987	0.98146918
img46	0.965	0.966	0.965499741
img47	0.934	0.945	0.939467802
img48	0.961	0.969	0.96498342
img49	0.987	0.966	0.976387097
img50	0.977	0.979	0.977998978
<b>AVERAGE</b>	<b>0.93548</b>	<b>0.9366</b>	<b>0.935898</b>

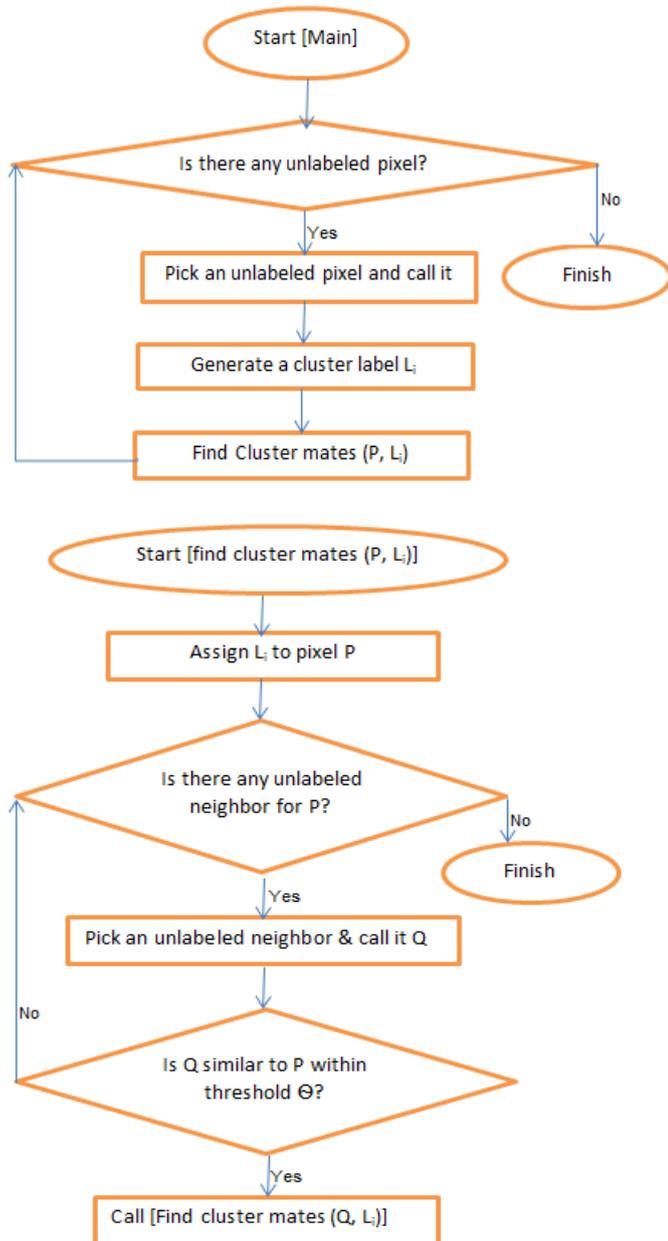


Figure 2. The Flowchart of the Proposed Algorithm

## VI. CONCLUSION AND FUTURE WORKS

As mentioned earlier, there are many different algorithms that eventuate to acceptable results and beside all, our proposed algorithm with its high speed and 94.3% accuracy is an efficient.

In the future, efficient utilization of image processing capabilities in the service of different medical usages will be achieved by the emerging capabilities for many complicated procedures. So as a future work, we can focus on improving the final result of the proposed algorithm by combining different algorithms in order to obtain the most efficient algorithms for image processing specially in medical usages. The results gained here can also be utilized in the next step for recognition of the whole MRI images.

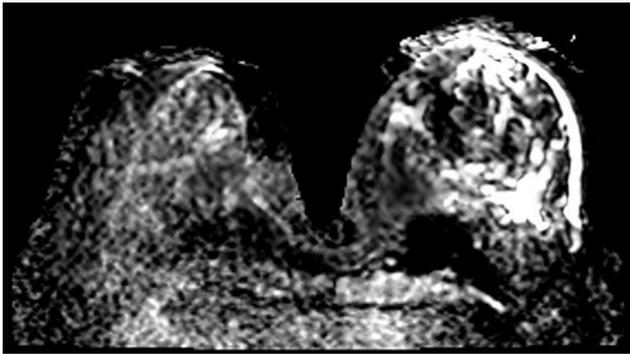


Figure 3. A Test Image

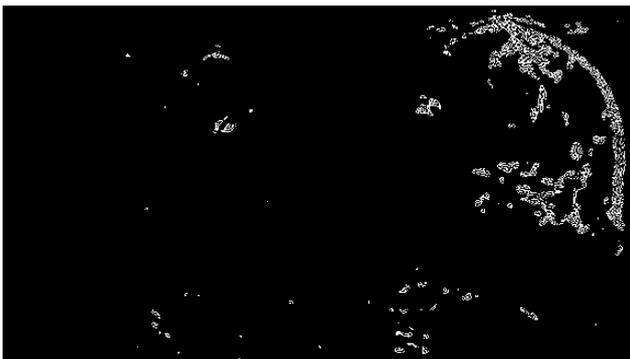


Figure 4. The result of the algorithm on test image

## REFERENCES

- [1] Kalisman A. Image processing--principles and its future applications in reconstructive and aesthetic plastic surgery. US national library of medicine national institute of health,1986.
- [2] [Http://www.mayoclinic.com/health/breast-mri/MY00300](http://www.mayoclinic.com/health/breast-mri/MY00300).
- [3] I.Christoyianni , A. Koutras, E. Dermatas, G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms" – Computerized Medical Imaging and Graphics, vol. 26, pp. 309-319, 2002.
- [4] Piesberg, A. Hoyer and W. Computerized Mammograms Processing. Phillips Technical Review. Vol. 38, pp. 347-355, 1979.
- [5] Giger, U. Bick and M. Automated Segmentation of digitized mammograms. Academic Radiology. Vol. 2, 1995.
- [6] Liang, T. Ojala and J. Interactive Segmentation of the Breast Region from Digitized Mammograms with United Snake. Turku Centre for Computer Science. 1999.
- [7] D. Raba, A. Oliver, J. Marti, M. Peracaula and J.Espunya. Breast segmentation with pectoral muscle suppression on digital mammograms. Iberian Conference on Pattern Recognition and Image Analysis. 2005.
- [8] Rangayyan, R. Ferrari and R. Segmentation of Mammograms:Identification of the Skin Boundary and the Pectoral Muscle. International Workshop on Digital Mammography. 2000.
- [9] Shadroo, Maarefdoost, Yaghoobi, Poorreza. Image segmentation by multi-fractal estimation and fuzzy clustering, A.I. group of Ferdowsi university,2005.
- [10] Tavakolkhah, Safabakhsh. A new method for classifying and segmenting cancerous glands in breast.Amirkabir University, 2005.
- [11] Shahedi, AmirFattahi, TorkamaniAzar, Sadri.Digital mammogram segmentation for separating breast tissues, 2007.
- [12] Sigurd Angenent, Eric Pichon, Allen Annenbaum, Mathematical methods in medical image processing, American mathematical society,2007.
- [13] David Cheng, Ravi Kannan, Santosh Vempala, On a Recursive Spectral Algorithm for Clustering from Pairwise Similarities, MIT university, 2003.