I J M I

International Journal of Machine Intelligence
ISSN: 0975–2927 & E-ISSN: 0975–9166, Volume 3, Issue 3, 2011, pp-130-133
Available online at http://www.bioinfo.in/contents.php?id=31

# COMPARATIVE PERFORMANCE EVALUATION OF SEGMENTATION METHODS IN BREAST CANCER IMAGES

**BHAGWATI CHARAN PATEL AND SINHA G.R.***
Shri Shankaracharya College of Engg. & Technology, Bhilai, Chhattisgarh, India
*Corresponding author. E-mail: bhagawaticharanpatel@gmail.com, drgrsinha@ieee.org

Received: September 29, 2011; Accepted: November 03, 2011

**Abstract-**Breast cancer is one of the major causes of death among women. Segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries in images. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic. We have used various segmentation algorithm methods. In this paper the comparison of the segmented images is done by taking the entropy and SNR information measures and it has been found that the lesion segmentation algorithm closely matches radiologists' outlines of these lesions.
**Keywords-** Breast Cancer, Image Segmentation, Image analysis, Contour, Entropy and SNR

## Introduction

One in eight deaths worldwide is due to cancer. Cancer is the second leading cause of death in developed countries and the third leading cause of death in developing countries. In 2009, about 562,340 Americans died of cancer, more than 1,500 people a day. Approximately 1,479,350 new cancer cases were diagnosed in 2009. In the United Sates, cancer is the second most common cause of death, and accounts for nearly 1 of every 4 deaths [1]. The chance of developing invasive breast cancer at some time in a woman's life is about 1 in 8 (12%) [2]. Xray mammography is the most common investigation technique used by radiologists in the screening, and diagnosis of breast cancer they could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis. In order to perform a semi-automated tracking of the breast cancer, it is necessary to detect the presence or absence of lesions from the mammograms [3, 4]. These lesions can be benign or malignant, according to their contour (sharp or blurred)- Stellar opacities (malignant tumors); micro calcifications: small calcified structures that appear as clear points on a mammogram. The work we have done is to propose a segmentation process which identifies on a mammogram the opaque areas, suspect or not, present in the image [5, 6]. Segmentation of an image is the division or separation of the image into regions of similar attribute. The most basic attribute for segmentation is image luminance for monochrome image and color components for the color image. Segmentation is required to distinguish objects from background.

## Breast cancer detection methods

Breast cancer screening is vital to detecting breast cancer. The most common screening methods are mammography and sonography. Compared to mammography, breast ultrasound examinations have several advantages [7]. Breast ultrasound examinations can obtain any section image of breast, and observe the breast tissues in real-time and dynamically. Ultrasound imaging can depict small, early-stage malignancies of dense breasts, which is difficult for mammography to achieve. Several statistical studies on the accuracy rate of breast disease diagnosis using ultrasonic examination have been carried out [8, 9]. The ultrasound examination has a high detection rate of tumors, in particular of malignant tumors. Accuracy rate of breast disease diagnosis using ultrasonic examination depends segmentation of images. Images are composed by a set of pixels whose values encode different colors or gray levels. Image segmentation methods have been used to find regions of interest (e.g. objects) in images. The image segmentation can be illustrated in diverse practical applications, such as in medical imaging (e.g. diagnosis [10]), satellite images [11], face recognition [12], traffic control system [13] and machine vision [14]. Different algorithms have been proposed for image segmentation such as those founded on image thresholding [15] ; clustering methods (e.g. neural networks [16]); region growing methods [17]; graph partitioning methods [18]; multi-scale segmentation [19], and semi-automated segmentation [20]. Methods related to physics concepts have also been more and more applied for image segmentation, such as those based on Markov random fields [21] and entropy [22].

### Image segmentation algorithms

Image segmentation is the process of assigning a level to every pixel in an image such that pixels with the same level share certain visual characteristics.

#### *A. K-MEANS CLUSTERING*

The K-means algorithm is an iterative technique that is used to partition an image into $K$ clusters. In statistics and machine learning, ***k*-means clustering** is a method of cluster analysis which aims to partition $n$ observations into $k$ clusters in which each observation belongs to the cluster with the nearest mean [23]. The basic algorithm is:

- Pick $K$ cluster centers, either randomly or based on some heuristic;
- Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center;
- Re-compute the cluster centers by averaging all of the pixels in the cluster

Repeat last two steps until convergence is attained (e.g. no pixels change clusters. Given a set of observations ($\mathbf{x}_1$, $\mathbf{x}_2$, …, $\mathbf{x}_n$), where each observation is a $d$-dimensional real vector, $k$-means clustering aims to partition the $n$ observations into $k$ sets ($k < n$) $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares :

$$\arg\min \sum_{i=1}^{k} \sum_{x_j \in s_i} \left\| x_j - \mu_i \right\|^2 \quad (1)$$

where $\boldsymbol{\mu}_i$ is the mean of points in $S_i$.

#### *B. SELF SIMILAR FRACTAL*

Fractals are of rough or fragmented geometric shape that can be subdivided in parts, each of which is a reduced similar of the whole. A fractal dataset is known by its characteristic of being self-similar. The dataset has roughly the same properties for a wide variation in scale or size i.e., parts of any size of the fractal are almost similar to the whole fractal [24]. Intuitively, a set of points which exhibit self similarity over all scales fractals are creased objects that defy conventional measures, such as length and area, and are most often characterized by their fractional dimension [25-26].

##### *a) **Fractal Dimension measurement***

The fractal dimension is found to be a measure of roughness and hence is used to model the texture.

$$D = \frac{\log N(r)}{\log(1/r)} \quad (4)$$

where D; the fractal dimension, N; number of copies of a self similar set, which has been scaled down by ratio 'r', r; scaled ratio of the self similar set. Instead, log N*(r)* versus log *r* is usually plotted for better results.

##### *b) **Methodology***

For the image analysis of mammograms, the measure of a region is defined as a function of the gray levels of the points belonging to the region. With the fractal approach, instead of one quantity or measure. Describing the phenomenon in all scales ω (as in case of fractals), a set of measures, weight factors depicting statistically the same phenomenon in different scales, has to be used for characterizing such structures. At the first step, the quantity called roughness exponent is derived as:

$$\delta = \frac{\log \omega(box)}{\log_{\varepsilon}} \quad (5)$$

Where $\delta$ quantifies the strength of the singularities of the measure, describing the local regularity of the object, with the determined measure of the box w(box) and size of the box ϵ. A single roughness exponent denotes number of self similar fractal, while in the self similar fractal case the different parts of the structure are characterized by different values of $\delta$, leading to the existence of the spectrum [27].

$$f(\delta) = \frac{-\log N_{\varepsilon}(\delta)}{\log \epsilon} \quad (6)$$

where $N_\varepsilon$ is the number of boxes of size having the common roughness exponent equal to $\delta$. The value of $f(\delta)$ may be seen as the fractal dimension of the image region that corresponds to a singularity $\delta$.

#### *C. REGION GROWING*

Region growing is a procedure that groups pixels or sub regions into larger regions. The simplest of these approaches is pixel aggregation, which starts with a set of "seed" points and from these grows regions by appending to each seed points those neighboring pixels that have similar properties (such as gray level, texture, color, shape) [28-29].

1. 1. The other procedure is Similarity Measures, in which individual pixel intensities are compared.
2. 2. The other method is comparing to neighbor in region. By this way, each pixel that is already in the region can bring in neighbors that are alike.
3. 3. The other method is merging, in which adjacent similar pixels and similar regions are merged. Eventually, this method will converge when no further such merging are possible

#### *D. WATERSHED ALGORITHM*

The watershed algorithms have been developed and tested on several mammogram breast cancer images [30]. It has been found that the results of segmentation gave very good clue to a radiologist /physician to further investigate on the presence of micro calcifications in the breast tissue.

**Steps**

- Read the mammogram image.
- Adjust intensity distribution by using suitable shareholding methods.
- Group individual cells under different colors.
- Extract detected small structures.
- Characterize no uniform background using morphology.
- Remove background by image subtraction.
- Segment after background removal.
- Extract new areas and update distribution.
- Compare distributions.
- Remove partial segments from cutoff segments.

- Compare the first order statistics of the segmented image.
- Finally indicate the detected micro calcifications.

### Comparision parameter

In this paper the comparison of the segmented images are done by taking the Entropy and SNR information measures. Entropy is a measure of disorder, or more precisely unpredictability. The entropy of an image can be defined as a measure of the uncertainty associated with a random variable and it quantifies, in the sense of an expected value, the information contained in a message. Here we take the concept of entropy in the sense of information theory where entropy is used to quantify the minimum descriptive complexity of a random variable [31]. Entropy of a discrete random distribution p(x) is defined as The entropy H of a discrete random variable *X* with possible values $\{x_1, .., x_n\}$ is

$$H(X)=E(I(X)) \tag{10}$$

Here E is the expected value, and *I* is the information content of *X*. *I*(*X*) is random variable. If *p* denotes the probability mass function of *X* then the entropy can explicitly be written as

$$H(X) = \sum_{i=1}^{n} p(x_i)I(x_i) = -\sum_{i=1}^{n} p(x_i)\log_b p(x_i)$$

(11)

The performance of noise removing algorithms is measured using quantitative performance measures by SNR as well as in term of visual quality of the images. Performance of all algorithms is tested with breast cancer Images. The Statistical Measurement for SNR [32] are given below

$$SNR = 10\ \log_{10}\frac{\sum_{i=1}^{k} S_i^2}{\sum_{i=1}^{k}(S'_i - S_i)^2}$$

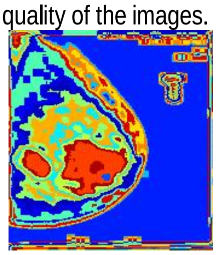
(12)

Where S'=noise image, S=original image, k= image size

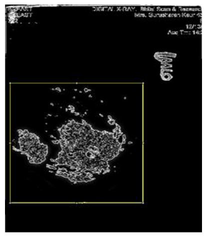
### Result and discussions

All of the real time breast images were collected from a reputed cancer diagnostic and research center to have a image database. 50 images were subjected to segmentation process using MATLAB 7.3 and P-IV. Fig.1 (a) shows the original image of affected breast cancer image. In K- means clustering segmentation method Fig.1 (b) shows the results for constant value of number of classes and number of bins. This can be observed that the affeced regions are more accurately located i.e. the identification of affected area with malignant effects gets more prominent. In Fig.1(c), image segmentation consists in finding the characteristic entities of an image, either by their contours or by the region they lie in. In the classical methods for edge detection, edges are usually considered to correspond to local extrema of the gradient of the gray levels in the image. The acquired 8-bit gray images were cropped and various regions of each mammogram were analyzed. The box size range varies between 8 and 64, for each 256 × 256 pixel region. From the self similar fractal standpoint characterization is done by both high $\delta$ and low $f(\delta)$ values, because they represent sharp local changes of contrast and rare events in global sense. In Fig. 1 (d) seed point or starting point must be found from where the particle can start its move. Where the seed point is located is irrelevant. All object pixels will be visited anyway. The movement of the particle starts from the seed point and the particle is jumped to a random position in its neighbor based on the condition that the random gray level is less than the gray level of the randomized position. In Fig. 1(e) is a result of application of the watershed algorithms on sample Images. Fig. 1(a) shows the various stages in image segmentation. The area of dense tissue is designated as shown in Fig. 1(e). The results are in accordance with the images diagnosed by an expert. Although the mammography images are textured and exhibit not homogeneous grey level dynamic, this approach provides promising segmentation results. The performance of Segmentation algorithms is measured with the help of quantitative measures such as Entropy and SNR (see Table 1) as well as in term of visual quality of the images.


(a)Original Image


(b) K-means Cluster


(c) Self Similar Fractal


(d) Region Growing

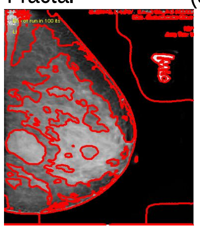

(e) Watershed algorithm

Fig. 1 Different types of image segmentation result.

*Table I- statistical measurements*

| S.N | Methods | Entropy | SNR |
|---|---|---|---|
| 1 | *K*-means clustering | **1.5768** | **28.65** |
| 2 | Self Similar Fractal | **1.8235** | **29.32** |
| 3 | *Region Growing* | **1.5012** | **23.65** |
| 4 | Watershed algorithm | **1.4372** | **31.74** |

The entropy can provide a good level of information to describe a given image. In this case, if all pixels in an image have the same gray level or the same intensity of color components, this image will present the minimal entropy value. On the other hand, when each pixel of an image presents a specific gray level or color intensity, it will exhibit maximum entropy. Thus, the pixel intensities are related to texture, because different textures tend to result in different distribution of gray level or color intensity.

## Conclusion

The goal of image segmentation process is to identify the segments of the image according to the image characteristic e.g., image color, objects shape etc. The simplified working of the image segmentation system is stated here. The work will lead to several experiments based on the algorithms introduced, which improve the quality based on the segment analysis on given images Performance of all algorithms is tested with breast cancer images. The computational result showed that K-means clustering and self similar fractal have the better results in terms of entropy values but watershed algorithm has highest SNR. Due to its nonlinear nature the self similar fractal has excellent both visual quality and perception and detail preserving properties. This system can be very helpful for the segmentation of the images which are used in different fields of life. The research content of this system was segmentation and image enhancement. Our future work will address several important problems such as the migration from region-based image matching to case-based image matching, better for processing poor quality radiographs, and retrieval from the dental Image repository. Results of segmentation methods are found very much consistent with the opinion and diagnosis of radiologists.

## Acknowledgment

The authors extend their sincere thanks to Dr. Dilip Soni, a senior radiologist at APPOLO BSR centre for cancer research and diagnosis; for providing necessary support and guidance throughout the research work.

## References

[1] Garcia M., Jemal A., Ward E., Center M., Hao, Y., Siegel R., and Thun M. (2007) *Global Cancer Facts & Figures 2007, American Cancer Society.*
[2] *Network of Strength. Breast Cancer Statistics. http://www.networkofstrength.org/information/bcnews/stats.php.Jan. 2010*.
[3] Sterns E.E. (1996) *Can. J. Surg.,* vol. 39, pp. 128-132.
[4] Highnam R. and Brady M. (1999) *Mammographic Image Analysis, Kluwer Academic Publishers, ISBN: 0-7923- 5620-9*.
[5] Matthew A. Kupinski and Maryellen L. Giger (1998) *IEEE transactions on medical imaging*, Vol. 17, No. 4.
[6] Wirth M.A., Stapinski A. (2003) *Visual Communications and Image Processing, Lugano*,Vol. 5150, pp. 234-240.
[7] Laine H.,Rainio J., Arko H. and Tukeva T. (1995) *European Journal of Ultrasound* 2, Vol. 2, p. 107-115.
[8] Hieken T., Harrison J. Herreros J. and Velasco J. (2001) *American Journal of Surgery* 182, Vol. 4, p. 351-354.
[9] Saarenmaa I., S., Isola J., Kataja V., Kokko M., Kokko R. and Kumpulainen E. (2001) *BCRT* 67, Vol. 2, p. 117-123.
[10] Noble J., Boukerroui D. (2006) *IEEE Transactions on Medical Imaging* Vol.25 (8) 987–1010.
[11] Masson P., Pieczynski W. (1993) *IEEE Transactions on Geoscience and Remote Sensing* Vol. 31 (3) 618–633.
[12] Zhao W., Chellappa R., Phillips P., Rosenfeld A. (2003) *ACM Computing Surveys* Vol. 35 (4) 399–458.
[13] Cucchiara R., Piccardi M., Mello P. (2000) *IEEE Transactions on Intelligent Transportation Systems* Vol. 1 (2), 119–130.
[14] Haralick R., Shapiro L. (1992) *Computer and Robot Vision, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA*.
[15] Navon E., Miller O., Averbuch A. (2005) *Image and Vision Computing* Vol. 23 (1) 69–85.
[16] Kuntimad G., Ranganath H., Div R., American B., Huntsville A. (1999) *IEEE Transactions on Neural Networks* Vol. 10 (3) ,591–598.
[17] Haralick R.M., Shapiro L.G. (1985) *Computer Vision, Graphics, and Image Processing* Vol. 29, 100–132.
[18] Shi J., Malik J. (2005) *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 22 (8) , 888–905.
[19] Kuijper A., Florack L., Viergever M. (2003) *Journal of Mathematical Imaging and Vision* Vol.8 (2), 169–189.
[20] Mortensen E., Barrett W. (1998) *Graphical Models and Image Processing* Vol. 60 , 349–384.
[21] Celeux G., Forbes F., Peyrard N. (2003) *Pattern Recognition* Vol. 36 (1) ,131–144.
[22] Portes de Albuquerque M., Esquef I., Gesualdi Mello A., Portes de Albuquerque M. (2004) *Pattern Recognition Letters*, Vol. 25 (9), 1059–1065.
[23] Kanungo T. (2002) *IEEE Trans. Pattern Analysis and Machine Intelligence* 24, p. 881–892.
[24] Pentland A. and Choudhury T. (2000) *Computer* 33(2): 50-55.
[25] Daugman and Downing C. (2001) *Proceedings of the Royal Society, Biological Sciences* 268: 1737-1740.
[26] Kronfeld P. (2002) *The Eye Academic Press: London*.
[27] Mandelbrot B. (1983) *The fractal geometry of nature, Freeman, San Francisco*.
[28] Kasmir Raja S.V., Shaik Abdul Khadir A., Riaz Ahamed S.S. (2009) *Journal of Theoretical and Applied Information Technology*, Vol. 3, No. 2, pp. 209-213.
[29] Hojjatoleslami S. A. and Kittler J. (1998) *IEEE Transactions On Image Processing*, Vol. 7, 7.
[30] Vincent L. and Soiled P. (1996) *IEEE Trans. On Machine Intel,* Vol. 18, No 12, pp. 1163-1173.
[31] Cover T., Thomas J., Wiley J. (1991) *Elements of Information Theory, Published in Inter Science*.
[32] Lopes A., Nezry E., Touzi R. and Laur H. (2004) *(IGARSS-2004),* pp. 345-352.