DOI: 10.21917/ijivp.2011.0040

MAMMOGRAM IMAGE SEGMENTATION USING AUTO ADAPTIVE FUZZY INDEX MEASURE

I. Laurence Aroquiaraj[1] and K. Thangavel[2]
Department of Computer Science, Periyar University, Tamil Nadu, India
E-mail: [1]laurence.raj@gmail.com, [2]drktvelu@yahoo.com

Abstract
Breast Cancer involves the uncontrolled growth of abnormal cells that have mutated from normal tissues. A radiologist looks for certain signs and characteristics indicative of cancer when evaluating a mammogram. The main task is to obtain the locations of suspicious regions to assist radiologists in diagnosis. Image segmentation has been approached from a wide variety of perspectives: region-based approach, morphological operation, multi-scale analysis, fuzzy approaches and stochastic approaches have been used for mammogram image segmentation but with some limitations. In spite of the several methods available in the literature, image segmentation still a challenging problem in most of image processing applications. The challenge comes from the fuzziness of image objects and the overlapping of the different regions. In this paper we propose fast auto adaptive image segmentation algorithm for finding the optimal thresholds for segmenting gray scale images. The proposed method is based on fuzzy index which decreases the similarity between pixels increases. The system uses initial estimation of the parameters. The fuzzy subsets derived from the image histogram using weighted fuzzy entropywill shows the similar cost measure as in pixels of the same subset. Experimental results demonstrate the effectiveness of the proposed approach.

Keywords:
X-ray Mammography, Fuzzy Entropy, Ostu multi-level Method, Segmentation

1. INTRODUCTION

Breast cancer has been one of the major causes of death among women since the last decades and it has become an emergency for the healthcare systems of industrialized countries. This disease became a commonest cancer among women. If the cancer can be detected early, the options of treatment and the chances of total recovery will increase. Intra-operative diagnosis of the disease has steadily become more important with respect to the recent introduction of sentinel lymph node biopsy.

The term benign refers to a condition, tumor or growth that is not cancerous. This means that it does not spread to other parts of the body or invade and destroy nearby tissue. Benign tumors usually grow slowly. In general, benign tumor or condition is not harmful. Breast cancer, also known as carcinoma, is a malignant growth that begins in the tissues of the breast.

Image preprocessing and enhancement methods help to improve the visual appearance of mammogram medical images such as removal of film artifacts and labels, filtering the image, normalization and removal of pectoral muscle region.

After image acquisition, the first one aims to segment the background and annotations from the whole breast area, while the second one involves separating the pectoral muscle (when present) from the rest of the breast area. For segmenting the breast from the pectoral muscle a new histogram of this biggest region is used. This histogram contains two zones: the pectoral muscle and the breast tissue. A region growing algorithm [14] is used to extract the pectoral muscle region from the breast. The seed of this region growing is placed inside the pectoral with value between the brightness maximum and the minimum between the two zones of the histogram.

The last step is the use of morphological operations in order to smooth the boundary of the breast. This biggest region can be extracted using a Connected Component Labeling algorithm. A good survey of both breast and pectoral segmentation types can be founded in [4].

Image segmentation is referred to as the procedure in which the input image is divided into meaningful regions in such a way that the output image will consist of a set of labeled region describing the input image. The output image will contain a set of non-overlapping objects representing pixels of similar gray values [7]. Image segmentation is a crucial step in a wide range of medical image processing systems. It is useful in visualization of the different objects present in the image. Numerous segmentation algorithms have been proposed and surveys of these techniques can be found in [12]. Image segmentation techniques can be categorized into three approaches. The first category uses clustering techniques such as adaptive fuzzy C means and K-means [6, 10, 18]. In the clustering techniques each pixels in the image is assigned as a class according to its features. The second category uses algorithms based on histogram thresholding [5, 18]. Histogram based methods work well for images which are can be clearly separated into two regions but fail there is no significant contrast between the objects and background. The third Category uses iterative approaches to achieve pixels separation [13]. Fuzzy entropy has been used for image segmentation [7, 9]. Most of the image segmentation algorithms produce binary image, or "foreground and background". While these results are acceptable in some image processing applications such as document processing and Optical Character Recognition systems, they are not satisfactory in medical images where several features, which are present in the image, need to be detected. In modem orthodontic practice, a great reliance is placed on objective and systematic methods of characterizing craniofacial forms, using measurements based on a set of agreed upon points know as craniofacial landmarks. When the X-ray images have been acquired, certain points (anatomical landmarks) on the X-ray mammography image have to be located in order to determine the proper breast treatment or the effect of previous treatment. Distance and angles among these landmarks are compared with normative values to diagnose patient's deviations from ideal form, evaluate the craniofacial growth and measure the effect of treatment. Without accuracy in land marking, it is impossible to determine craniofacial parameters correctly. This process is carried out manually and consisted of two steps: producing cephalometric tracing then they try to locate the anatomical landmarks based on their distance from the comers or based on the shape of soft and bony tissue of the skull. However, this procedure is inevitably affected by many factors such as human error. It is difficult to place markers exactly on the X-ray comers and furthermore, repeatability cannot be assured. Panjabi *et.al.* [13] discussed in detail errors that arise when manually marking X-ray mammography images.

In this paper, we propose a system capable of performing multi-level segmentation of X- ray mammography images in an automatic way. The proposed system will divide the image into three segments based on the use of fuzzy entropy and fuzzy set theory. The image is divided into three parts, namely, dark (background of X-ray which is air), gray (soft tissues or the skin) and white part (bony tissues). The proposed system can be used as a building block by a more advanced image processing systems such localization of the X-ray mammogram or as a standalone system for image segmentation.

This paper is organized as follows: in section 2 we provide as basic concepts of fuzzy sets theory and fuzzy entropy, section 3 outlines the steps of the proposed algorithm, section 4 is results and discussion section 5 is for conclusions.

2. BACKGROUND

2.1 FUZZY SET THEORY

The crisp set is defined in such a way as to dichotomize the individuals in some given universe of discourse into two groups: members (those that certainly belong in the set) and non-members (those that certainly do not). A sharp, unambiguous distinction exists between the members and non-members of the class or category represented by the crisp set. Many of the collections and categories we commonly employ, however (for instance, in natural language), such as the classes of tall people, expensive cars, highly contagious diseases, numbers much greater than 1, or sunny days, do not exhibit this characteristic. Instead, their boundaries seem vague, and the transition from member to non-member appears gradual rather than abrupt. Thus, the fuzzy set introduces vagueness (with the aim of reducing complexity) by eliminating the sharp boundary dividing members of the class from non-members [4].

A fuzzy set A is a subset of the universe of discourse X that admits partial memberships. The fuzzy set A is defined as an ordered pair A = $\{x, \mu(x)\}$, where, $x \in X$ and $0 \le \mu_A(x) \le 1$. The membership function $\mu_A(x)$ describes the degree to which the object x belongs to the set A, $\mu_A(x) = 0$ represents no membership and $\mu_A(x) = 1$ represents full membership. There exist several types of membership functions that characterize A. In this research we use Generalized Bell Membership function (GBMF) defined as follows,

$$\mu_A(x) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{1}$$

where, c is the center of the membership function, as is the width of the set at the cross-over point and b is the slope of the curve.

2.2 BASIC IMAGE MODEL

This image model based on the theory of fuzzy sets, it is possible to consider images as fuzzy subsets of a plane. With gray level resolution of k bits per pixel, a gray level of $2^k - 1$ corresponds to the upper bound of the membership function, then the gray level at a given pixel position can be interpreted as the degree of membership of that pixel to the object. An image X with size of M x N having L gray-levels ranging from 0 to L = $2^k - 1$ can be defined as an array of fuzzy memberships. Each membership function value denotes its degree of brightness relative to some gray level L.

The image can be represented by,

$$X = \{\mu_x(x(i,j));\ i = 1, 2,\ldots, M;\ j = 1, 2,\ldots,N\} \tag{2}$$

where, $\mu_x(x(i, j))$ denotes the fuzzy grade of brightness of pixel located at (i, j).

2.3 FUZZY ENTROPY

It is a measure of fuzziness that becomes smaller when the similarity of its argument is increased. It measures degree of fuzziness of a subset. There are several types of such measures proposed in the literature [2]. The most common types are Shannon's Entropy [16] and the distance measure which is based on the distance between fuzzy subset A and its complement subset A^C.

The complement subset of A is defined as follows,

$$\mu_{A^C}\left(x(i,j)\right) = 1-\mu_A\left(x(i,j)\right) \tag{3}$$

The index of fuzziness based on the distance between A and its complement set is defined in [8].

$$\nu(A) = \frac{2}{(N.M)^{1/k}} d_k(A, A^C) \tag{4}$$

where, d, (A, A') denotes the distance between the two sets, N.M is the size of A, and k is the order of the distance used.

$$d(A, A^C) = \left(\sum_{i=1}^{N}\sum_{j=1}^{M}\mu_A(x(i,j) - \mu_{A^C}(x(i,j))\right)^{1/k} \tag{5}$$

If Hamming distance is used (k=l). In this research we use the Euclidian distance (k=2) and the fuzzy index will become

$$\nu(A) = \frac{2}{\sqrt{N.M}}\sqrt{\left(\sum_{i=1}^{N}\sum_{j=1}^{M}\left|2\mu_A(x(i,j))-1\right|\right)} \tag{6}$$

The fuzzy index measures the ambiguity of the fuzzy subset or the homogeneity between pixels of the set. The fuzzy index of a fuzzy set reflects the degree of ambiguity present in it. That is, a fuzzy set having a low fuzzy index indicates that its elements are similar.

2.4 X-RAY MAMMOGRAPHY

X-Ray Mammography is currently performed using a conventional phosphor screen-film combination as the image receptor. Properly exposed film mammograms reveal fine detail in the breast, with the capability of detecting contrast levels as low as 2 to 5%. Mammography provides high sensitivity on fatty breast and excellent demonstration of micro calcifications; it is highly indicative of an early malignancy. Due to its low cost, it is suitable for mass screening program. Mammography has its limitations. It is less reliable on dense breast of young women or women underwent a surgical intervention in the breast because glandular and scar tissues are as radiopaque as abnormalities. Furthermore, there is low dose X-Ray radiation [14].

3. PROPOSED METHOD

Since fuzzy index is a measure of the amount of confusion between pixels in an image, it can be used as a cost term in a minimization process in order to reduce the confusion between pixels in subsets. The subsets are constructed by forcing the pixels to be in the subset that minimizes the fuzzy index given in Eq. (6). Our objective is to segment the gray-level image by splitting the image into three crisp subsets, background (black), skin subset(gray) and bones subset(white) using the measure of fuzzy entropy. We will define three linguistic variables (background, skin, bone} modeled by three fuzzy subsets denoted by G, S and B, respectively. The fuzzy subsets are associated with the normalized image histogram intervals defined by the following,

$$\left[x_{G_{\min}}, x_{G_{\max}}\right], \left[x_{S_{\min}}, x_{S_{\max}}\right] \text{ and } \left[x_{B_{\min}}, x_{B_{\max}}\right]$$

where, $x_{S_{\min}}$, $x_{B_{\min}}$, $x_{G_{\min}}$ and $x_{S_{\max}}$, $x_{B_{\max}}$ $x_{G_{\max}}$ are the initial and final gray-level limits for these subsets and are defined as follows,

$$x_{S_{\min}} = \mu_S - \sigma_S,\ x_{S_{\max}} = \mu_S + \sigma_S,$$
$$x_{G_{\min}} = 0,\ x_{G_{\max}} = \mu_G + \sigma_G,$$
$$x_{B_{\min}} = \mu_B - \sigma_B,\ x_{B_{\max}} = N$$

where, μ_G, μ_S, μ_B and σ_G, σ_S, σ_B are the averages and the variances of the gray values of G, S and B subsets defined as follows,

$$\mu_G = \frac{\int_0^{T_1} xp(x)dx}{\int_0^{T_1} p(x)dx} = \text{mean of the dark pixels} \tag{7}$$

$$\mu_S = \frac{\int_{T_1}^{T_2} xp(x)dx}{\int_{T_1}^{T_2} p(x)dx} = \text{mean of the gray pixels} \tag{8}$$

$$\mu_B = \frac{\int_{T_2}^{T_3} xp(x)dx}{\int_{T_2}^{T_3} p(x)dx} = \text{mean of the bright pixels} \tag{9}$$

where, p(x) the Probability density functions of gray-level and x is variable representing the gray-level. The initial threshold values for each subset are decided based on the iterative approach presented in [17] the threshold values are defined as follows,

$$T_{i,k+1} = \frac{1}{2}\left[\frac{\int_{T_{i-1}}^{T_{i,k}} xp(x)dx}{\int_{T_{i-1}}^{T_{i,k}} p(x)dx} + \frac{\int_{T_i}^{T_{i+1}} xp(x)dx}{\int_{T_{i,k}}^{T_{i+1}} p(x)dx}\right] \tag{10}$$

where, $T_{i,k+1}$ is the threshold value at iteration k + l, for i=0,…,3 and T_0=0 and T_3 = N, where N is the number of gray levels in the image. The values T_0 and 0,N and ∞ are the same and they will be used interchangeably.

σ is the variance of pixels in each subset defined as follows,

$$\sigma_G = \sqrt{\int_0^{T_1} (i-\mu_G)^2 p(x_i)dx} \tag{11}$$

$$\sigma_S = \sqrt{\int_0^{T_3} (i-\mu_S)^2 p(x_i)dx} \tag{12}$$

$$\sigma_B = \sqrt{\int_{T_3}^{T_4} (i-\mu_B)^2 p(x_i)dx} \tag{13}$$

The initial values of subsets, means, variances and slop are shown in Fig.2. The slop of each of the fuzzy subsets is decided based on an iterative approach of minimizing the fuzzy index of the fuzzy regions. For example the slope of the fuzzy subset G is decided as follows.

Starting with a value of b as $b = \frac{255}{(\mu_G + \sigma_G) - \mu_S}$ and incrementing or decrementing the value of b toward the reduction of subset's fuzzy index. Similar approaches are used for deciding the other subset slopes.

4. RESULTS AND DISCUSSION

Obtaining real mammogram images (322 images) for carrying out research is highly difficult due to privacy issues, legal issues and technical hurdles. Hence the Mammography Image Analysis Society (MIAS) database (ftp://peipa.essex.ac.uk) is used in this paper to study the efficiency of the proposed image segmentation and evaluated using mammography images. After computing the histogram we obtain initial threshold using Eq.(10). This is a fast algorithm and most of times it will converge after 4-7 iteration. These 322 mammogram images were processed in such a way that after removal of pectoral muscle region, only the reduced image is used for segmentation. Average results were compared with Ostu's 3-levels method proposed in Multi-level thresholding [17].

Table.1. Fuzzy Entropy Values for Normal, Ostu and Proposed Method

Image Name	**Normal Fuzzy Entropy Method**	**Ostu method (Ground)**	**Ostu method (Skin)**	**Ostu method (Bone)**	**Proposed Method (Ground)**	**Proposed Method (Skin)**	**Proposed Method (Bone)**
mdb001	461.8593	217.3345	495.4028	488.3123	217.6558	492.1632	487.4615
mdb002	456.5552	225.6068	488.2861	483.0452	225.8962	484.3756	480.0112
mdb003	444.5899	454.5766	495.1159	493.5075	454.6607	493.9402	493.2732
mdb004	435.8009	441.3481	494.4904	473.2884	441.2787	490.9631	469.4346
mdb005	442.8517	320.7064	480.0494	502.1112	318.8519	468.6349	468.6349
mdb006	435.1348	316.3426	480.8935	504.8748	315.1964	475.1328	504.9635
mdb007	456.2851	275.9959	499.9168	472.8805	276.3717	497.6536	467.0325
mdb008	443.8662	358.0849	484.7778	491.8351	356.6469	458.7576	476.7657
mdb009	453.2929	452.0232	485.6506	508.7439	450.8196	477.6084	508.8224
mdb010	453.1954	253.6190	501.5770	484.7303	253.9863	500.5065	482.5789
Average of 322 mammogram images	**442.0961**	**368.8982**	**492.6976**	**485.9598**	**368.5635**	**487.1691**	**480.6658**

It is observed from the Table.1 that proposed method ground, skin and bone produces minimized fuzzy entropy values. Fig.1 shows a comparison of the Normal, Multi-level thresholding and proposed fuzzy entropy average values obtained by using the proposed methods. Fig.2 shows the segmented images using proposed algorithm.

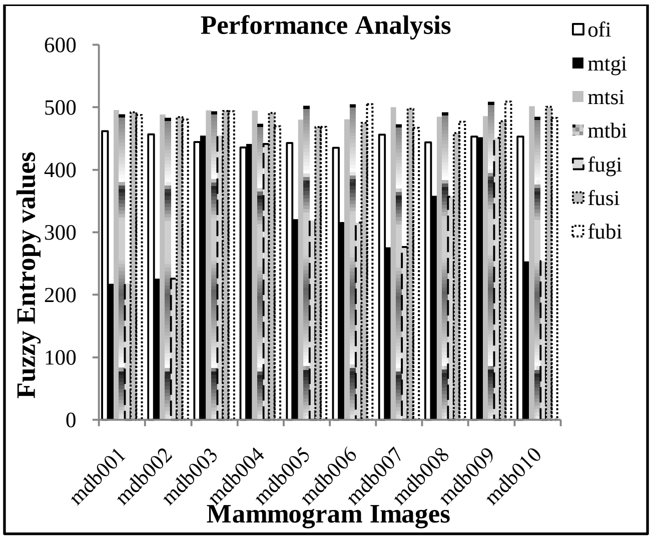

Fig.1. Performance Analysis for Fuzzy Entropy values of ten mammogram images

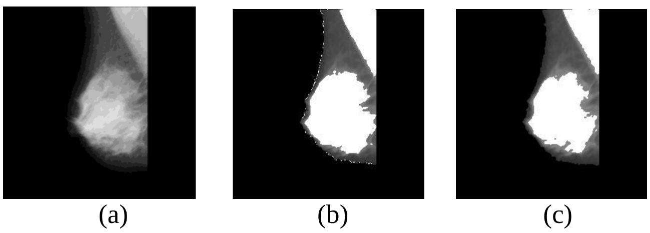

Fig.2(a) Original Mammogram Image(mdb001) (b) Image segmented using multi-level thresholding (mdb001) (c) using the proposed Fuzzy Entropy method (mdb001)

5. CONCLUSIONS

In this paper, we introduce a procedure for mammogram gray scale image segmentation which is based on the minimization of a fuzzy measure. The image histogram is divided into three fuzzy subsets using iterative approach to obtain subsets parameters. The obtained parameters were used as initial estimates the each pixel in the fuzzy regions were classified as belonging to one of the subsets by maximizing the fuzzy index. Results obtained show improvement over previous methods. The algorithm shows better performance when tested on images with no overlapping of regions.

REFERENCES

[1] Cheng H.D, Chen Y.H, and Sun Y, "A novel fuzzy entropy approach to image enhancement and thresholding", *Signal Processing*, Vol. 75, No. 3, pp. 277- 301, 1999.

[2] Chung-Hoon Rhee F and Yong-Shik S, "A fast numerical method for finding the optimal threshold for image segmentation", *The 12th IEEE International Conference on Fuzzy Systems*, Vol. 2, pp. 984-989, 2003.

[3] El-Feghi I, Sid-Ahmed M.A and Ahmadi M, "Automatic Localization of Craniofacial Landmarks for Assisted Cephalometry", *Pattern Recognition*, Vol. 37, No. 3, pp. 609-621, 2004.

[4] George J. Klir and Tina A. Folger, "*Fuzzy Sets, Uncertainty and Information*", Prentice-Hall of India Private Limited, 2007.

[5] Glasbey C.A, "An analysis of histogram-based thresholding algorithms", *Graphical Models and Image Processing*, Vol. 55, No. 6, pp. 532- 537, 1993.

[6] Hall L.O, Bensaid A.M, Clarke L.P, Velthuizen R.P, Silbiger M.S, and Bezdek J.C, "A comparison of neural

network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain", *IEEE Transactions on Neural Networks*, Vol. 3, No. 5, pp. 672-682, 1992.

[7] Hill P.R, Canagarajah C.N, and Bull D.R, "Image Segmentation Using a Texture Gradient Based Watershed Transform", *IEEE Transactions on Image Processing*, Vol. 12, No. 12, pp. 1618-1633, 2003.

[8] Kaufinann A, "*Introduction to the Theory of Fuzzy Subsets-Fundamental Theoretical Elements*", Academic Press, Inc., New York, 1975.

[9] Li X, Zhao Z, and Cheng H.D, "Fuzzy entropy threshold approach to breast cancer detection", *Journal on Information Science – Applications*, Vol. 4, No. 1, pp. 49-56, 1995.

[10] Liew A and Yan H, "An Adaptive Spatial Fuzzy Clustering Algorithm for 3-D MR Image Segmentation", *IEEE Transactions on Medical Imaging*, Vol. 22, No, 9, pp. 1063-1075, 2003.

[11] Otsu N, "A threshold selection method from gray- level histograms", *IEEE Transactions on Systems, Man and Cybernatics*, Vol. 9, No. 1, pp. 62-66, 1979.

[12] Pal N.R and Pal S.K, "A review on image segmentation techniques", *Journal of Pattern Recognition*, Vol. 26, No. 9, pp. 1277-1294, 1993.

[13] Panjabi M, Chang D, and Dvŏrák J, "An analysis of errors in kinematics parameters associated with in vivo functional radiographs", *Spine*, Vol. 17, No. 2, pp. 200-205, 1992.

[14] Raba D, Oliver A, Martí J, Peracaula M and Espunya J, "Breast Segmentation with Pectoral Muscle suppression on digital mammograms", *Proceedings of Iberian Conference on Pattern Recognition and Image Analysis*, pp. 471-478, 2005.

[15] Yapa R.D and Harada K, "Connected Component Labeling Algorithms for Gray-Scale Images and Evaluation of Performance using Digital Mammograms", *International Journal on Computer Science and Network Security*, Vol. 8, No. 6, pp. 33-41, 2008.

[16] Shannon C. E, "A mathematical theory of communication", *The Bell System Technical Journal*, Vol. 27, pp. 379-423, 1948.

[17] Sid-Ahmed M.A, "A Hardware Structure for the Automatic Selection of Multi- Level Thresholding in Digital Images," *Pattern Recognition*, Vol. 25, No. 12, pp. 1517-1528, 1992.

[18] Singh M, Patel P, Khosla D, and Kim T, "Segmentation of functional MRI by K-means clustering", *IEEE Transactions on Nuclear Science*, Vol. 43, No. 3, pp. 2030-2036, 1996.