EFFICIENT ROI SEGMENTATION OF DIGITAL MAMMOGRAM IMAGES USING OTSU'S N THRESHOLDING METHOD

Deepa S.[1], Subbiah Bharathi V.[2]

[1]Research Scholar, Sathyabama University, Chennai, India

[2]Principal, DMI College of Engineering, Chennai, India
Email: [1]dineshdeepas@yahoo.co.in

Abstract

Segmentation of the Region of Interest (ROI) is the first and crucial step in the analysis of digital mammogram images since the success of any Computer Aided Diagnostic (CADx) system depends greatly on the accuracy of the segmentation of the ROI from the mammogram images. Finding an accurate, robust and efficient ROI segmentation technique still remains a challenging problem in digital mammography analysis. In this paper we have proposed an efficient Otsu' s N thresholding method for segmenting regions of interest from the mammogram images. Digital Mammograms are taken from the mini MIAS(Mammographic Image Analysis Society) database for the purpose of experimentation and the results obtained are scaled to full color. Results show that the proposed method is efficient and is in concurrence with the ground truth table available in the database.

Key words: Digital Mammograms, MIAS database, Region Of Interest (ROI), OTSU's N thresholding, segmentation.

I. INTRODUCTION

Cancer begins when cells in a part of the body start to grow out of control. There are many kinds of cancer, but they all start because of out-of-control growth of abnormal cells. Instead of dying, cancer cells continue to grow and form new, abnormal cells. Cancer cells can also invade (grow into) other tissues, something that normal cells cannot do. Growing out of control and invading other tissues are what makes a cell a cancer cell. In most cases the cancer cells form a tumor. Breast cancer is a cancer that starts in the tissues of the breast. There are two main types of breast cancer, Ductal carcinoma and Lobular Carcinoma. Ductal carcinoma starts in the tubes (ducts) that move milk from the breast to the nipple. Most breast cancers are of this type. Lobular carcinoma starts in the parts of the breast, called lobules, that produce milk. Breast cancer may be invasive or noninvasive. Invasive means it has spread from the milk duct or lobule to other tissues in the breast. Noninvasive means it has not yet invaded other breast tissue. Noninvasive breast cancer is called "in situ." Ductal carcinoma in situ (DCIS), or intraductal carcinoma, is breast cancer in the lining of the milk ducts that has not yet invaded nearby tissues. It may progress to invasive cancer if untreated. Lobular carcinoma in situ (LCIS) is a marker for an increased risk of invasive cancer in the same or both breasts[1]. Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among females, accounting for 23% of the total cancer cases and 14% of the cancer deaths. GLOBOCAN 2008 [2] estimates shows that about 12.7 million cancer cases and 7.6 million cancer deaths are estimated to have occurred in 2008; of these, 56% of the cases and 64% of the deaths occurred in the economically developing world. The incidence of breast cancer in India is on the rise and is rapidly becoming the number one cancer in females pushing the cervical cancer to the second spot. The seriousness of the situation is apparent after going through recent data from Indian Council of Medical Research (ICMR). The rise is being documented mainly in the metros, but it can be safely said that many cases in rural areas go unnoticed. It is reported that one in 22 women in India is likely to suffer from breast cancer during her lifetime, while the figure is definitely more in America with one in eight being a victim of this deadly cancer. The problem with preventing breast cancer is that there is no one cause that can be pointed as being the culprit. Studies show that if detected early the survival rate is high. Mammography is currently the most effective imaging modality for breast cancer screening [1]. Mammograms are used as a screening tool to detect early breast cancer in women experiencing no symptoms and to detect and diagnose breast disease in women experiencing symptoms such as a lump, pain or nipple discharge. There are two types of mammography, Screening Mammography and

Diagnostic Mammography. Screening Mammography is a regular routine done annually for women with no symptom for early detection of breast cancers because it can show changes in the breast up to two years before a patient or physician can feel them. Diagnostic mammography is used to evaluate a patient with abnormal clinical findings, such as a breast lump or lumps that have been found by the woman or her doctor. Diagnostic mammography may also be done after an abnormal screening mammography in order to evaluate the area of concern on the screening exam. Many Computer-Aided Diagnosis (CADx) systems have been developed as a second opinion to assist radiologists. The research presented in this paper is part of an ongoing project to develop a reliable CADx system to classify the abnormalities in the mammograms as malignant or benign. While studying mammograms, radiologists look for abnormalities. These include masses of an approximately regular spherical or stellate shape, microcalcification clusters and asymmetry between breasts. Using CAD schemes as a second opinion in mammography has increased the number of early detected breast cancers, as well as it has improved the radiologist's performance. One of the crucial and important steps in CADx is the segmentation of the Region of Interest (ROI) from the mammogram image. Region of Interest is the region which covers the area of abnormalities seen in the mammograms. In one mammogram image there could be one or more regions of interest. The success or accuracy of any CADx system largely depends on the segmentation stage. In this paper we have extended OTSU's segmentation technique to threshold the image with multiple thresholds. Rest of the paper is organized as follows, section II contains the literature survey on various medical image segmentation techniques, section III describes OTSU's N thresholding method for segmentation. Section IV contains the results and section 5 the conclusion.

II. LITERATURE SURVEY

Segmentation of mammogram images refers to segmenting the regions in the image that contains abnormalities (lesions). Analysis of these regions helps us to classify the image under consideration to be benign or malignant. Hence these regions are referred to as Region Of Interest (ROI) in the process of analyzing the mammogram images. Various approaches have been proposed for the task of segmenting lesions in the mammogram images. The earliest approach is the use of global thresholding [3], since then extensive research effort is going on for the successful segmentation of masses or lesions from the mammogram images. The global segmentation approach proposed by Bick *et al.* [4] incorporates aspects of thresholding, region growing and morphological filtering. Yin *et al.* [5] proposed a method based on the deviation from the normal architectural symmetry of the right and left breasts, a bilateral subtraction technique is used to enhance the conspicuity of possible masses. Recent efforts, such as that of Masek *et al.* [6] using local thresholding have shown more promising results compared to global thresholding. Abdel- Mottaleb *et al.*[7] use a system of masking images with different thresholds to find the breast edge. Méndez *et al.* [8] found the breast contour using a gradient based method. Rangayyan [9] described standard operators for edge detection such as Prewitt operator, Sobels operator, Roberts operator and Laplacian of Gaussian (LoG) operator. Fauci et al. [10] developed an edge-based segmentation algorithm that uses iterative procedure, a ROI Hunter algorithm for selecting ROIs. Zou et al. [11] proposed a method that uses gradient vector flow field (GVF) which is a parametric deformable contour model. After the enhancement of mammographic images with adaptive histogram equalization, the GVF field component with the larger entropy is used to generate the ROI. A lesion segmentation algorithm developed by Sameti et al. [12] used fuzzy sets to partition the texture features with morphologic features mammographic image data. Sahiner et al [13] presented a fully automated and three-stage segmentation method that included clustering, active contour, and spiculation detection stages to detect spiculated and non spiculated masses. The results indicated that combining extracted features from automatically segmented mass regions was an effective approach for the automated characterization of mammographic masses. V.Grau et al [14] proposed improved watershed transform for medical image segmentation using prior information. Saheb Basha et al [15] have used Morphological Operators and Fuzzy C-Means clustering techniques for the segmentation of masses. Sheshadri et al [16] proposed segmentation based on statistical measures. Significant research effort is into medical image segmentation and in this paper we have proposed an extension of the OTSU's segmentation algorithm, Otsu's N thresholding for efficient segmentation of digital mammogram images.

III. OTSU'S N THRESHOLDING METHOD

OTSU's method of segmentation is an optimum global thresholding method proposed by Nobuyuki Otsu [17]. It is a nonparametric and unsupervised method of automatic threshold selection for segmentation of images. It is a simple procedure and utilizes only the zeroth and the first-order cumulative moments of the gray-level histogram. It is optimum in the sense that it maximizes the between class variance, a well known measure used in statistical discriminant analysis. Let the pixels of the input image be represented in L gray levels, [0,1, 2,…,$L-1$] for an image of size M $\times$ N pixels. The number of pixels having gray level value 'i' is denoted by n_i and the total number of pixels in the image, MN = $n_0 + n_1 + \ldots + n_{L-1}$. Suppose we select a threshold, k, 0 < k < $L-1$ and use it to threshold the image into two classes, C_1 and C_2, class C_1 consists of pixels with intensity values in the range [0, k] and class C_2 consists of the pixels with intensity values in the range [$k+1$, $L-1$]. Using this threshold, the probability, $P_1(k)$, that a pixel is assigned to class C_1 is given by the cumulative sum,

$$P_1(k) = \Sigma_{i=0}^{k} p_i \quad [1]$$

Similarly, the probability of Class C_2 occurring is,

$$P_2(k) = \Sigma_{i=k+1}^{L-1} p_i = 1 - P_1(k) \quad [2]$$

The mean intensity values of the pixels assigned to class C_1,

$$m_1(k) = \frac{1}{p_1(k)} \Sigma_{i=0}^{k} ip_i \quad [3]$$

and similarly the mean intensity values of the pixels assigned to class C_2,

$$m_2(k) = \frac{1}{P_1(k)} \Sigma_{i=k+1}^{L-1} ip_i \quad (4)$$

The global mean is given by,

$$m_G = \Sigma_{i=0}^{L-1} ip_i \quad (5)$$

The problem is to find an optimum value for k which will maximize the criterion defined by,

$$\gamma(k) = \frac{\sigma_B^2(k)}{\sigma_G^2} \quad [6]$$

Where $\sigma_B^2(k)$ is the between class variance defined as

$$\sigma_B^2(k) = P_1(m_1 - m_G)^2 + P_2(m_2 - m_G)^2 \quad [7]$$

and σ_G^2 is the global variance defined as,

$$\sigma_G^2 = \Sigma_{i=0}^{L-1} (i - m_G)^2 P_i \quad [8]$$

The optimum threshold is the value, k^*, that maximizes $\sigma_B^2(k)$.

The Otsu's segmentation method above explained segments the image into two classes using the optimum threshold value k^*. For our application of segmentation of mammogram images it is not sufficient if the image is segmented into two classes, hence the OTSU's segmentation method is extended to OTSU's N thresholding mehtod. We have used three (N=3) thresholds and segmented the image into four classes of varying intensities. The algorithm used is explained below:

A. *Otsu's N segmentation algorithm:*

1. Compute the normalized histogram of the input image. Denote the components of the histogram by p_i, i= 0,1,2,…,L-1.
2. Compute the cumulative sums, P_1, P_2, P_3, P_4

   Where,

   $$P_1 = \Sigma_{i=0}^{k_1} p_i \quad [9]$$

   $$P_2 = \Sigma_{i=k_1+1} p_i \quad [10]$$

   $$P_3 = \Sigma_{i=k_2+1} p_i \quad [11]$$

   $$P_4 = \Sigma_{i=k_3+1} p_i \quad [12]$$

3. Compute the global intensity mean, m_G, using equation [5].
4. Compute the mean intensity values of the classes using the following equations,

   $$m_1 = \frac{1}{P_1} \Sigma_{i=0}^{k_1} ip_i \quad [13]$$

   $$m_2 = \frac{1}{P_2} \Sigma_{i=k_1+1}^{k_2} ip_i \quad [14]$$

$$m_3 = \frac{1}{P_3} \Sigma_{i=k_2+1}^{k_3} ip_i \quad [15]$$

$$m_4 = \frac{1}{P_4} \Sigma_{i=k_3+1}^{L-1} ip_i \quad [16]$$

5. Compute the between class variance, $\sigma_B^2(k_1, k_2, k_3)$, for $\left\{ \begin{array}{l} k_1 = 1, \ldots, k_2 - 1, \\ k_2 = k_1 + 1, \ldots, k_3 - 1, \\ k_3 = k_2 + 1, \ldots, L - 2 \end{array} \right\}$ using the relation,

[17]

6. The optimum threshold values are the values corresponding to maximum between class variance i.e.,

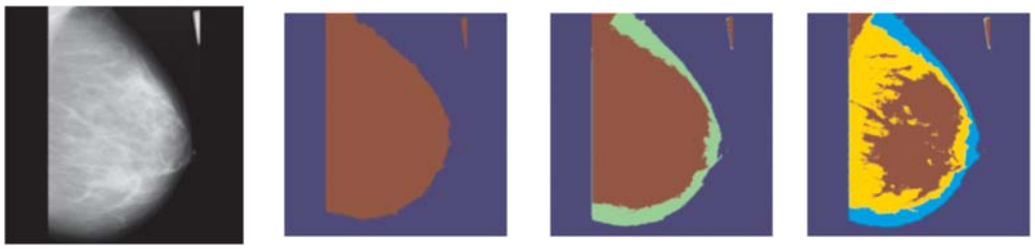
= max [18]

where,

If the maximum is not unique, obtain k* by averaging the values of k corresponding to the various maxima detected.

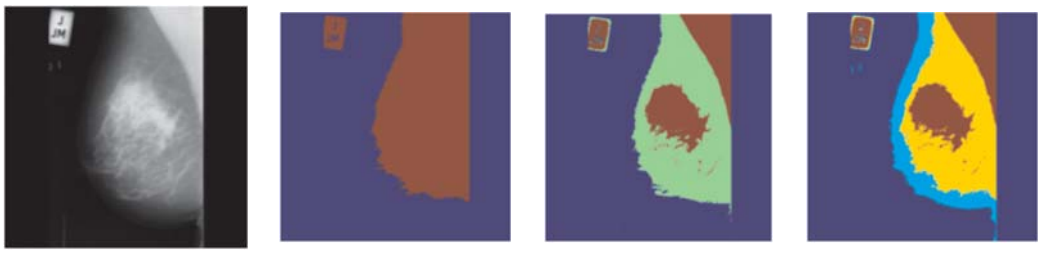
IV. RESULTS

The digital mammograms used for experimentation is taken from the digital mammography database offered by the Mammography Image Analysis Society (MIAS), which contains left and right breast mammogram images of 161 patients. The original images are of the size 1024 × 1024 pixels. The proposed algorithm was implemented on the database images and the results obtained are shown below. Figure 1(a) shows the original mammogram image mdb050 and 1(b) shows the segmented image using n=1, that is the image is segmented into two classes. Thus the foreground and the background are segmented. But our Region of Interest is the abnormalities in the mammogram image, hence segmentation at this level is not enough. Thus we have increased the n value to increase the number of classes. Figure 1(c) shows the result for n=2 and 1(d) shows the result for n=3. Thus from the results obtained it is clear that segmentation using n=3 gives better results. Further increase in the value of n will lead to oversegmentation making the results unfit for further processing as shown in figure of 2, 3 & 4.



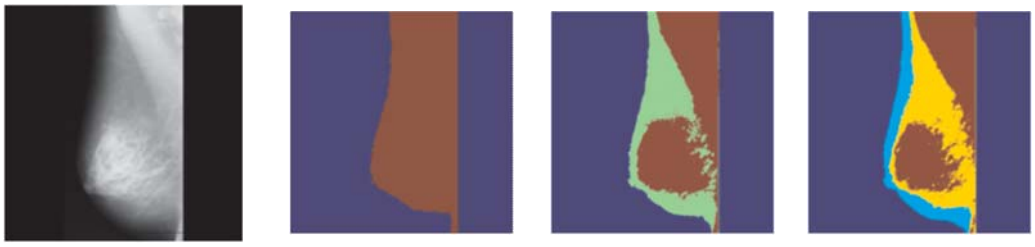
(a) (b) (c) (d)

Fig. 1. (a) mdb050 original mammogram image (b) Segmented output for n=1 (2 classes) (c) Segmented output for n=2 (3 classes) (d) Segmented output for n=2 (3 classes)



(a) (b) (c) (d)

Fig. 2. (a) mdb209 original mammogram image (b) Segmented output for n=1 (2 classes) (c) Segmented output for n=2 (3 classes) (d) Segmented output for n=2 (3 classes)



(a) (b) (c) (d)

Fig. 3. (a) mdb211 original mammogram image (b)Segmented output for n=1 (2 classes) (c)Segmented output for n=2 (3 classes) (d) Segmented output for n=2 (3 classes)

(a) (b) (c) (d)

Fig. 4. (a) mdb249 original mammogram image (b) Segmented output for n=1 (2 classes) (c) Segmented output for n=2 (3 classes) (d) Segmented output for n=2 (3 classes)

V. CONCLUSION

Segmenting the ROI from digital mammogram images for further processing is the first step in the process ofIn this paper we have extended OTSU's segmentation technique to segment the image into more number of classes. We have segmented the digital mammogram images using 3 threshold values and have obtained four classes. For better visual interpretation we have scaled the image to full color and displayed. The results obtained are comparable and are to be used for further processing of the images. This work is the initial part of our CADx system to classify the digital mammograms as normal, benign or malignant.

REFERENCES

[1] www.radiologyinfo.org

[2] Jemal A., Siegel R., Ward E., Hao Y., Xu J., Murray T., and Thun M.J., *"Cancer Statistics, 2008,"* CA: A Cancer Journal for Clinicians, vol.58, pp.71-96, 2008.

[3] Brzakovic, D., Luo, X.M., Brzakovic, P., "An approach to automated detection of tumors in mammograms". *IEEE Transactions on Medical Imaging,* Vol.9, Issue (3), pp.233-241 (1990).

[4] Bick U., Giger M.L., Schmidt R.A., Nishikawa R.M., Wolverton D.E., and Doi K., "Automated Segmentation of Digitized Mammograms", Academic Radiology, vol no.2, pp.1-9, 1995.

[5] Yin F.F., Giger M.L., Doi K., Metz C.E., Vyborny C.J., and Schmidt R.A., "Computerized Detection of Masses in Digital Mammograms: Analysis of Bilateral Subtraction Images", *Medical Physics*, vol.18, no.5, pp. 955-963, 1991.

[6] Masek M., Attikiouzel Y., and C.J.S. deSilva, "Skin-air Interface Extraction from Mammograms Using an Automatic Local Thresholding Algorithm", in Proc. of the 15th Biennial International Conference Biosignal, pp.204-206, 2000.

[7] Abdel-Mottaleb M., Carman C.S., Hill C.R., and Vafai S., "Locating the Bounday between the Breast Skin Edge and the Background in Digitized Mammograms", in Proc. of the 3rd International Workshop on Digital Mammography (WDM), pp. 467-470. 1996.

[8] Mandez A.J., Tahoces P.J., Lado M.J., Souto M., Correa J.L., and Vidal J.J., "Automatic Detection of Breast Border and Nipple in Digital Mammograms", Computer Methods and Programs in Biomedicine, vol.49, pp.253-262, 1996.

[9] Rangayyan, R.M.: Biomedical Image Analysis. CRC Press LLC, Boca Raton (2005).

[10] Fauci, F., Bagnasco, S. Bellotti, R., Cascio, D., Cheran, S.C., De Carlo, F., De Nunzio, G., Fantacci, M.E., Forno, G., Lauria, A., Torres, E.G., Magro, R., Masala, G.L., Oliva, P., Quarta, M., Raso, G., Retico,

A., Tangaro, S.: Mammogram Segmentation by Contour Searching and Massive Lesion Classification with Neutral Network. In: 2004 IEEE Nuclear Science Symposium Conference Record, Rome, Italy, October 16-22, vol.5, pp.2695-2699 (2004).

[11] Zou, F., Zheng, Y., Zhou, Z., Agyepong, K.: Gradient Vector Flow Field and Mass Region Extraction in Digital Mammograms. In: 21st IEEE International Symposium on Computer-Based Medical Systems, CMBS 2008, Jyvaskyla, June 17-19, pp. 41-43 (2008).

[12] Sameti M. and Ward R.K., "A fuzzy segmentation algorithm for mammogram partitioning", in Digital Mammography, International Congress Series. Amsterdam, the Netherlands: Elsevier, 1996, pp.471-474.

[13] Sahiner B, Chan HP, Petrick N, Improvement of mammographic mass characterization using spiculation measures and morphological features. Med Phys 2001; 28:1455-1465.

[14] Grau V., Mewes A.U.J., Alcañiz M., Kikinis R., and Warfield S.K., Improved Watershed Transform for Medical Image Segmentation Using Prior Information., IEEE Transactions on Medical Imaging, vol.23, no.4, April (2004).

[15] Saheb Basha S., Satya Prasad K., "Automatic Detection of Breast Cancer Mass in Mammograms using Morphological Operators and Fuzzy C-Means clustering", *Journal of theoretical and Applied Information Technology*, 2009, pp.704-709.

[16] Sheshadri H.S., Kandaswamy A., "Detection of breast cancer by mammogram image segmentation", *Journal of Cancer Research and Therapeutics* - December 2005 - Volume 1 - Issue 4, pp.232-234.

[17] Otsu N., "A Threshold Selection Method from Gray-level Histograms," *IEEE Transactions on Systems, Man & Cybernatics, 1979,* Volume 9, Issue 1 pp: 62-66.