IMPACT: International Journal of Research in Engineering & Technology (IMPACT: IJRET)
ISSN(E): 2321-8843; ISSN(P): 2347-4599
Vol. 2, Issue 3, Mar 2014, 31-36

EARLY BREAST CANCER DETECTION USING STATISTICAL PARAMETERS

H. C. NAGARAJ[1], PRASANNA PAGA[2] & KAMAL LAMICHHANE[3]

[1]Principal, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India

[2]Associate Professor, Department of ECE, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India

[3]B.E. Student, Department of ECE, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India

ABSTRACT

Breast cancer is the second most common cancer overall and the leading cause of cancer deaths in women. Studies proven that an early diagnosis of breast cancer can increase five year survival rate from 60% to 80+%. Mammography is, at present, the only viable method for detecting most of tumors early enough for effective treatment. The secret of setting up the accurate diagnosis is to detect and understand the most subtle signs of breast lesions. Analysis of different features of mammograms can provide clues about the presence of early signs of tumors. In this work we present an automated procedure for detection using image processing techniques. Many image processing methods were developed over the past two decades to help radiologists in diagnosing breast cancer. In this paper a new algorithm is introduced for Mammograms Region of Interest (ROI) identification using statistical properties of mammograms. The proposed algorithm has been verified using 100 mammograms from the MIAS databases and other sources. Simulation results show that the proposed algorithm achieved 70% true result.

KEYWORDS: Mass Detection, Breast Mammogram, Statistical Measures, Averaged Datum Moments

INTRODUCTION

Mammography (MMG) is widely used as a principal breast cancer screening method, however mass screening generates large number of images. Mammography is, at present, the only viable method for detecting most of tumors early enough for effective treatment, without unnecessary biopsies or other invasive procedures. Therefore, screening mammography in women aged 40 to 70 years is currently the effective strategy to reduce breast cancer mortality. Early detection of invasive breast cancers is associated with better prognosis than waiting for women to become symptomatic. However, detecting the early signs of breast cancer is challenging because the cancerous structures have many features in common with normal breast tissue. Moreover, the accuracy of interpretation of screening mammograms is affected by several factors, such as image quality, the radiologist's level of expertise, and the high volume of cases. According to recent statistics, in current breast cancer screenings, 10%−25% of the tumors are missed by the radiologists.

Computer Aided Detection (CAD) systems can support radiologists in the role of a second reader aiding the radiologist in finding the suspicious breast lesions and distinguishing between what is decidedly negative on a mammogram, as opposed to what needs regular monitoring and what requires a needle biopsy. The secret of setting up the accurate diagnosis is to detect and understand the most subtle signs of breast lesions [3]. According to the fourth edition of breast Imaging Reporting and Data System (BIRADS) [4], subtle signs of breast cancer are four: classifications, masses, architectural distortion, and bilateral asymmetry. The latest two signs do not necessarily mean that cancer is already present, but provide clues about the presence of early signs of tumors. However, a few works have been reported on the detection of various feature extraction using Matlab. In this paper a proposed algorithm is introduced to highlight suspected lesions to deduce the effective statistical properties of MMG.

MMGs used in this study are digitized images at 400x400 Dot per Inch (DPI) and 1024x1024 pixels in 8 bits per pixel in bit map format (BMP).

THE PROPOSED ALGORITHM

Tumors have higher x-ray attenuation coefficient than normal soft tissues, which means higher intensity in MMG image. Current CAD systems rely heavily on sophisticated methods in machine learning to address the area of pattern recognition and classification which has high computational load, leading to longer time for analysis of a single case. The proposed Effective Statistical Texture Detection algorithm (ESTD) is based on the weighted value of each bit in the 8-bits representation of a pixel in an MMG image. The least significant bit carries least significant information weight as its value changes more rapidly; whereas higher order bits carry most significant information weight and change at a slower rate, i.e. carrying more meaningful information. A thresholding step can reduce the effect of bits with low information content, by excluding them to simplify the process of ROI identification in later processing stages.

The processing steps of the ESTD algorithm are:

Preprocessing

This step is to identify the breast boarder, the Area Of Interest (AOI). In many MMGs background objects with high intensity values make breast boundaries identification a challenging task, especially for the scanned ones where the original film has some artifacts. Breast boundary identification algorithm given in [2] was applied to remove background objects (noise) as well as pectoral muscle. This was implemented as in the following steps:

- Identify breast boarder and AOI using a two dimension linked list technique as follows:
  - Small weighted value pixels, below a threshold value 32, are set to zero, where as higher (>32) weighted value bits set to 255 to create an image reference. Figure (1-a) shows the output of this step, it is the raw image reference.
  - For each row of the image reference identify linked pixels in that row with value above zero by giving each group an object index.
  - Link objects (column wise) of two adjacent rows by giving one object index (the smallest index) to connected objects.
  - Since breast is the biggest object, then the object with largest number of pixels is identified to be the breast and the remaining objects are set to zero. The resulting reference image is shown in figure (1-b).
- Identify the pectoral muscle. This is done by first identifying breast direction and then detect the biggest object touching straight and upper ends of identified breast boarder in previous step. This is shown in figure (1-c).
- Exclude the pectoral muscle from image reference as illustrated in figure (1-d).

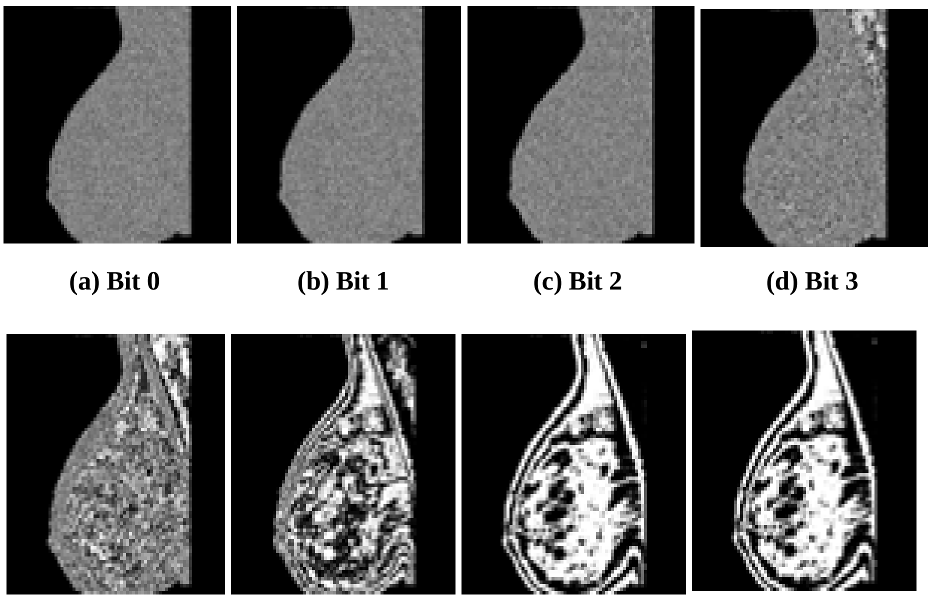
(a) (b) (c) (d)

Figure 1: Raw Image Reference: (a) After Thresholding (b) After Breast Boarder Identification (c) Pectoral Muscle Identification (d) AOI after Pectoral Muscle Exclusion

- The source mammogram with its image reference after pectoral muscle exclusion to get the AOI image, to be used in later processing stages.

Resolving of a MMG into 8 Images

This step is included to generate an image reference for objects with high intensity. Figure 2 Shows the output of resolving a sample MMG into 8 image bit planes, each image represents one bit plane of the MMG's pixels. From figure 2 it is obvious that bits 6&7 contain high intensity solid objects information while the remaining bits do not hold such information. So, the generated object image reference should consider objects represented in bit plane 6 and 7, this means pixels with weighted values of 192 or higher are to be considered as raw ROI, where as pixels with weighted value lower than this threshold are set to zero. The value 192 is to be defined as the high intensity objects datum threshold.



(a) Bit 0 (b) Bit 1 (c) Bit 2 (d) Bit 3

(e) Bit 4 (f) Bit 5 (g) Bit 6 (h) Bit 7

Figure 2: Resolving MMG into 8 Images Each Representing One Bit in All Pixels

Statistical Parameters Performance

The identified high intensity objects in the threshold datum, step (B), are filtered using statistical properties analysis of each object index. The following statistical measures were tested: Mean (Harmonic and Arithmetic), Median, Mode, Standard deviation (variance), Smoothness, Uniformity, and first, second, third, and fourth order moments (µ1, µ2, µ3, and µ4 respectively).The challenge here is to determine which measure is to be used to differentiate between normal and suspected objects and what is the border line between normal and suspected object value. Table 2 illustrates the statistical values of identified high intensity objects after step (B) in a sample MMG given in [12].

It is worthy to mention that the selected MMG sample well represent the MIAS MMGs being tested. Tumors are visually detected as high intensity adjacent cells in a MMG, which means high population of adjacent pixels with high intensity value. Mapping this property into statistical measures means:

- **Mean –Variance Relation:** a. High mean value; but if variance is also high that may not mean a solid group of pixels at high value. b. High mean value with small variance can be a "good" indicator of a suspected object.

  Table 1 illustrates the logic mapping between the Mean- Variance and the suspected objects.

- **Mode:** It gives only information about highest repetition value, but can't give any indication on intensity value.
- **Median:** It gives the mid value in the range, but does not reflect any differentiating figure.
- **Uniformity (U):** For suspected objects is very small compared to other objects. But it does not reflect the intensity of the object.
- **Smoothness (R):** It is nearest to 1 for suspected objects compared to normal objects, but again it is not related to intensity.
- **Moment:** It gives the relation between mean value and the distribution of pixels values around the mean.

Table 1: The Logic Mapping between Mean-Variance and Suspected Objects

Mean	Variance	Suspicious Possibility
Low	Low	Low
Low	High	Low
High	Low	High
High	High	Mid

Table 2: Statistical Values of Sample MMG Objects after Step(B)

Image	A-Mean	Median	Mode	σ	µ1	µ2	µ3	µ4
A_01	195.8152	195.50	192	2.99.007	-27.116373	175.431305	-7.98998	32.298679
B_02	194.2536	194	194	2.26535	-28.126	78.445	-10.55225	15.445
C_03	194.2845	193	192	1.02365	-25.125	79.4526	-15.159697	8.1945
D_04	195.2644	194	196	1.25645	-15.458	48.485	-5.78953	5.825
E_05	195.7879	194	195	3.21466	-5.963	16.152	-4.45968	10.465
F_06	195.7456	194	193	1.85680	-5.785	16.485	-4.45968	7.155
G_07	195.3465	194	195	2.45206	-4.452	13.152	-3.12987	5.231

Suspicious Object Selection Criteria

This step will define accept / reject criteria of identified high intensity objects in step (B) using statistical measures selected in step (C). Using central moments as a measure is very difficult to draw the line between normal and suspected objects values. The difficulty arises from the fact that central moments always refer to object's mean value (µ). The question should be how the moment is related to other objects in this MMG. If moment equation is slightly shifted to a common datum to all objects other than each object's (µ), this leads to a better differentiating factor. The new value will be a threshold value used to qualify these objects to this step, i.e. the 192 instead of (μ) in the central moment equation. But using the value 192 as it is includes also objects that are exactly on the boundary level, and based on the simulation done; those objects are not reflecting real masses. So, a datum value of 194 is used instead of 192 to move the base point slightly beyond the threshold limit, and the updated moments will be referred to as datum moments.

SIMULATION RESULTS AND DISCUSSIONS

A sample set of 100 MMGs were selected from the MIAS database and other sources of breast MMGs to test and evaluate the performance of the proposed ESTD algorithm. After pre- processing step, thresholding step identified many high intensity objects or suspected regions. The target from the use of feature extraction step is to represent the visual interpretation of the identified objects in numbers and then to select the proper parameters to be used in the last step. After applying many statistical measures; as Mean, Median, Mode, Variance, central moments, smoothness, uniformity, the spatial frequency autocorrelation function as in [4], and Moments (first, second, third, and fourth order). From the previous analysis it is found that the datum moments (first, Second, third, and fourth) are good representatives of visual features of the objects.

The final challenge was to define selection / rejection criteria of suspected objects. Simulation results show that Uniformity and Smoothness can be good indicators of homogeneity within

Object, but can't be a differentiating factor between suspected and normal regions as it does not link to object intensity. Also central moments had the same problem but when the central moments were changed to the introduced averaged datum moments, more differentiating results were achieved. The selection criteria step was done to define the differentiating values between normal and suspected objects. After applying the selected statistical measures, the averaged datum moments, first selection criterion is to have objects moments above the selected datum point; this mathematically means nonnegative first or third order datum moments. The remaining objects going to be selected based on the heavier distribution at higher intensity end. To avoid the effect of the object size (number of pixels), the datum moments were averaged by dividing its value by the number of pixels of the concerned object. The second selection criterion is to only consider objects with any averaged datum moment above 1.

CONCLUSIONS

A simple method, in terms of computational complexity, intensity based suspected cells detection was developed to find suspected lesions in breast MMGs. The proposed ESTD algorithm had four key steps. (A) De-noising and AOI identification using statistical properties. (B) The datum thresholding step to select the high intensity objects, it is a further reduction of number of pixels to be analyzed by the following steps. (C) Statistical measures to identify suspicious objects, which are the introduced Averaged Datum Moments ($\mu1$, $\mu2$, $\mu3$, and $\mu4$). (D) Identifying selection criteria with determined value to accept or to reject suspicious objects.

As a future work, the proposed ESTD algorithm can be applied to different imaging techniques (MRI, MRA, and US) considering their different visual specific properties. For example in MRI temporal and spatial resolution need to be considered after injecting contrast agent in the body, which is very different from MMG case. In Ultrasound (US) images the mass is represented by a dark region. The segmentation of a mass region on an US image is generally difficult because the signal is weak and noisy.

REFERENCES

1. Mazhar Tayel, Abdlemonem Mohsen, *"Breast Boarder Boundaries Extraction Using Statistical properties of Mammogram," IEEE 10th international conference on signal processing (ICSP 2010) Beijing, pp. 2468-2471, October 2010.*
2. A. Cao, Q. Song, and X. Yang, *"Robust information clustering incorporating spatial information for breast mass detection in digitized mammograms," Computer Vision and Image Understanding (109), pp.86-96, 2008.*
3. G.N. Srinivasan, and Shobha G., *"Statistical Texture Analysis," PWASET, vol. 36, pp. 1264-1269, December 2008.*
4. I. Abdel-Qader, L. Shen, C. Jacobs, F. Abu Amara, and S. Pashaie-Rad, *"Unsupervised detection of suspicious tissue using data modeling and PCA," International Journal of Biomedical Imaging, vol. 2006, Article ID 57850, 11 pages, 2006.*
5. Andrew Webb, *Introduction to biomedical imaging, Wiley Interscience press, 2003.*
6. Mark Nixon, Alberto Aguado, *Feature extraction & image processing, Newnes press, 2006.*
7. Richard A. Robb, *Biomedical imaging, visualization, and analysis, Wiley Liss, 2000.*
8. Tinku Acharya, Ajoy K. Ray, *Image processing, principles and examples, Wiley-Interscience, 2005*.
9. Elizabeth Berry, *A practical approach to medical image processing, CRC press, 2008.*
10. Griffiths & Tenenbaum, *"From algorithmic to subjective randomness," In advances in neural information processing, 2004.*
11. Mazhar Tayel and Abdelmonem Mohsen, *Statistical Measures and Criteria for ROI Identification in Breast Mammograms*
12. J Suckling et al., *"The Mammographic Image Analysis Society Digital Mammogram Database Experta Medica – MIAS," International Congress Series 1069, pp375-378, 1994. http://www.mammoimage.org/databases*