Multiscale Segmentation and Enhancement in Mammograms

M. M. ANGUH & A. C. SILVA[1]

[1]Departamento de Engenharia de Eletricidade
Universidade Federal do Maranhão (DEEE/UFMA)
Centro Universitário do Bacanga
Av. dos Portugueses S/N
São Luís, 65080-040
MA - Brasil
mma@dee.ufma.br, ari@dee.ufma.br

Abstract. A multiscale method for segmenting and enhancing lesions of various sizes in mammograms is presented. The method uses two stages. The first stage applies a multiscale automatic threshold estimator based on histogram moments to segment the mammogram at multilevels. The second stage converts the segmented image using pseudo-colour mapping to produce a colour image.

An algorithm is presented as well as experimental results. Mammograms are digitalised using a table scanners with a transparency adapter and the algorithm is implemented in Borland C++ 4.02 on a 486 PC. The results are analogous to a *breast map* which provide an adequate basis for radiological breast tissue differentiation and analysis in digital mammography. Experimental results and judgments from radiological experts are very encouraging.

1 Introduction

Mammography is an effective imaging modality for early breast cancer detection [3, 10]. Despite the advances in mammography, human interpretation still remains very difficult. An estimate of about 10% to 30% of tumors are missed during human interpretation and about 40% of the missed tumors appear as masses [8]. Computer aided diagnosis such as mathematical morphology [12], wavelet transforms [7], neural networks [1], filtering [8] have been proposed for enhancement [7, 11], microcalcification segmentation [4], detection and extraction [1, 2, 8] and classification [9, 12].

One of the most important steps in digital mammography is an adequate segmentation of possible tumors. This obviously minimises errors in further stages such as in classification. However, several factors, some of which are listed below affect the proper segmentation of mammograms.

> Mammograms contain low signal to noise ratio (low contrast) and a complicated structured background. Breast tissue contrast and density vary with age, thus mammography produces varying image qualities. In addition, mammographic images are not bimodal. As a result, any segmentation method which utilises an a-priori or single threshold value method is highly likely to generate serious segmentation errors. This paper overcomes these problems by utilising automatic thresholds derived from the image, and the multiscale segmentation approach eliminates errors which are usually presented by single threshold methods.
>
> Objects such as microcalcification have very small and varying sizes. The sizes could be about 0.7mm with a diameter of 0.3mm. An adequate method must be capable of segmenting objects with very small and varying sizes. This paper utilises a multiscale segmentation method based on histogram moments to overcome these problems. This permites the segmentation of objects with various forms and sizes.
>
> Tumors or calcifications are embedded in an inhomogeneous background. In mammograms, background objects may even appear brighter. Therefore, global threshold methods suffer considerable drawback. Adaptive neighbourhood segmentation methods attempt to overcome such drawbacks, but implementational issues such as neighbourhood sizes and the determination of regions where background objects are brighter still pose a difficult problem. These problems are resolved by using a multiscale segmentation method and the resulting thresholds are mapped using pseudo-colour image processing techniques to enhance human visualisation.

The proposed method utilises two stages. The first stage segments the mammogram at multiscale using an automatic threshold estimator based on histogram moments derived from the characteristics of the mammogram being processed. The second stage uses a pseudo-colour mapping scheme to generate a colour image of the segmented image. All objects of small and varying sizes are segmented and enhanced. An algorithm is presented as well as experimental results. Mammograms are digitalised using a table scanners with a transparency adapter and the algorithm is implemented on a 486 PC. The final results are analogous to a *breast map* which provide an adequate basis for radiological breast tissue differentiation and analysis in digital mammography. Experimental results and the judgment from radiological experts are very encouraging.

This paper is arranged as follows. Section 2 provides the definitions and the mathematical theory of moments for an automatic threshold value determination [13] based on mean, contrast and skewness. Section 3 gives a multiscale segmentation and enhancement algorithm. Section 4 presents experimental results and some advantages.

2 Mathematical Theory of Moments

This section provides the definitions and the mathematical theory of moments [5, 13] upon which multiscale segmentation and enhancement is based. The reason for using multiscale segmentation is based on the fact that mammograms are not bimodal, thus overcoming the problem of undesirable results produced by single threshold methods.

Definition 1

Given a histogram with two modes a and b, the segmentation error can be minimised by minimising the cost function,

$$e = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} (f(k,l) - a)^2 (f(k,l) - b)^2 \qquad (1)$$

where $f(k,l)$ is the image.

Definition 2

The r^{th} moment M_r is defined as

$$M_r = \frac{1}{N} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} f(k,l)^r = \frac{1}{N} \sum_{i=0}^{P-1} h(i) i^r \qquad (2)$$

where $N = KL$ is the total number of pixels in the image, $h(i)$ is the i^{th} histogram value which is the number of pixels which have $f(k,l) = i$, and P is the number of grey levels.

Definition 3

The central moments H_r of the histogram are defined as

$$\begin{aligned} H_r &= 1/N \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} (f(k,l) - m)^r &(3) \\ &= 1/N \sum_{i=0}^{P-1} h(i)(i - m)^r &(4) \end{aligned}$$

where $r = 1, 2, \ldots$ and m is the mean of the distribution.

- *H_1 is the central histogram mean of the distribution.*
- *H_2 is known as the variance and gives the measure of gray level contrast.*
- *H_3 is the measure of the skewness of the histogram.*

Further information on the derivation, analysis and properties of central moments can be obtained from the book by Gonzalez [6].

Minimisation [13, 14, 15] of the segmentation error e using the cost function in Equation (1) gives

$$\begin{aligned} a &= \frac{c_1 - \sqrt{c_1^2 - 4c_2}}{2} &(5) \\ b &= \frac{c_1 + \sqrt{c_1^2 - 4c_2}}{2} &(6) \end{aligned}$$

where

$$\begin{aligned} c_1 &= \frac{H_3 - H_1 H_2}{H_2 - H_1^2} &(7) \\ c_2 &= \frac{H_1 H_3 - H_2^2}{H_2 - H_1^2} &(8) \end{aligned}$$

The threshold value [13, 14, 15] is given by

$$T = 0.5(a + b) + m \qquad (9)$$

Let us examine the nature of T in equation (9). The deviation of a threshold value T from the mean of the distribution m depends solely on the variables a and b. Consider the following theorem.

Theorem 1

If an image has constant intensity, the central moments $H_r = 0$ for $r \geq 1$.

From Equation (3),

$$H_r = 1/N \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} (f(k,l) - m)^r \qquad (10)$$

Since the image has constant intensity, $f(k,l) = m$ and thus $H_r = 0$ for $r \geq 1$. Note also that $(a + b) = 0$ since both c_1 and c_2 are zero, hence the threshold value which minimises the segmentation error is m.

3 Multiscale Segmentation and Enhancement

The mathematical basis presented above can be iteratively used to automatically generate threshold values T_j, $j = 0, \ldots, P$ for multiscale segmentation. For such an automatic process, the following convergence criteria are required.

(1) $T_j > P-1$. The threshold value is greater than the maximum pixel intensity.

(2) $T_j < T_{j-1}$. In practice, this rarely occurs. This would occur if $(a+b) < 0$, and this case in treated below.

(3) $(a+b) < 0$. We note from Theorem 1 that m is the threshold value for an image with constant intensity. Equations (5) and (6) demonstrate that $(a < b)$, and from Equation (1), both $(a, b) \geq 0$. If $(a+b) < 0$ then $(a < 0)$ which contradicts Equation (1) and multiscale segmentation terminates.

Algorithm 1 computes the various different threshold values T_j, $j = 0, \ldots, k$ where $k < P$ for multiscale segmentation.

Algorithm 1

1 Compute the histogram $h[i], i = 0, \ldots, P-1$, *of* $f(l, k)$.

2 $T_0 = 0.5(a+b) + m$

3 If $((T_j > P-1)$ *or* $(a+b) < 0$ *or* $(T_j < T_{j-1}))$ *goto step 7 else goto step 4.*

4 $M_r = \frac{1}{W}\sum_{i=T_j}^{P-1} h(i)(i-m)^r$, *where* W *is the number of pixels in the image with grey level* $i \geq T_j$.

5 $T_{j+1} = 0.5(a+b) + m$.

6 Goto step 3.

7 exit.

However, for a proper segmentation, histogram post-processing is required. A adaptive histogram equalisation method is used to correctly assign pixels to their correct classes defined by the thresholds. To segment at threshold value T_j, a histogram equalisation is applied in the interval T_{j-1} to T_{j+1} before segmenting. This is based on the observation that intensity values may sometime be higher at the center of lesions and gradually decrease outwards to the periphery and spikes. This operation therefore attempts to segment the cancer and spikes into the same class. This can be regarded as region growing.

Multiscale segmentation produces a mammographic map $v(k, l) = T_j$ based on the various thresholds with varying object sizes. This mammographic map can be used by a radiologist to interpret the mammogram. All objects are identified even when the background is brighter than objects.

Minimisation of incomplete segmentation is achieved by an adaptive histogram equalisation method. However, noise appear as spike-like regions which can be eliminated on the basis of size.

The human visual system can only discern a limited number of gray scale. Therefore, representing the results from multiscale segmentation in gray scale mode does not actually produce sufficient contrast and tissue differentiation for a adequate radiological interpretation and analysis. An adequate manner is to map the multiscale segmented mammogram based on the various threshold values using pseudo-colour image processing techniques [6] to produce a colour image. Consider distinct threshold values T_j, and distinct colour values C_j for $j = 0, \ldots, n$ where $n \leq P$, then pseudo-colour mapping is achieved by

$$f(k, l) = C_j \;\; where \;\; T_j \leq f(k, l) < T_{j+1} \tag{11}$$

where $f(k, l)$ is the original image. There is no significance to the colours attributed. The main purpose is to improve visualisation for human mammographic interpretations. However, further research being undertaken aims to classify these results and attribute meaning to the colours attributed. Judgments from radiological experts are very encouraging with the use of pseudo-colour image representation.

4 Experimental Results

The algorithm was implemented on a 486 PC using Borland C++ 4.02. Experiments were conducted on two sets of images of a total of 50 mammograms. The maximum size of the digitalised mammograms considered was 128×128 pixels.

The first set of experiments on 25 images were conducted in a local clinic. The mammograms were digitalised by an ordinary table ScanJet 4C scanner from Hewlett Packard with a transparency adapter. Analysis of results by radiological experts were based on three criteria compared to direct radiological mammographic interpretation. Of the 25 mammograms tested, the results from expert radiological analysis were as follows:

- *Bad*. The multiscale segmentation and enhancement method did not detect all lesions visually identified by an expert radiologist in the process of visual inspection. No result was registered in this category.
- *Equal*. In addtion to the facility provided by colour enhancement, all the lesions identified by the radiologist were detected by the method and colour mapped. In this category, 21 mammograms were registered.
- *Better*. In addtion to the facility provided by colour enhancement, lesions were detected which the radiologist did not identify or had uncertainty in the visual inspection of the mammogram. 4 mammograms were registered.

The results presented in Figure 1 and Figure 2 represent the different multilevel thresholds obtained by Algorithm 1 starting with the original digitalised mammogram on the top-left corner.

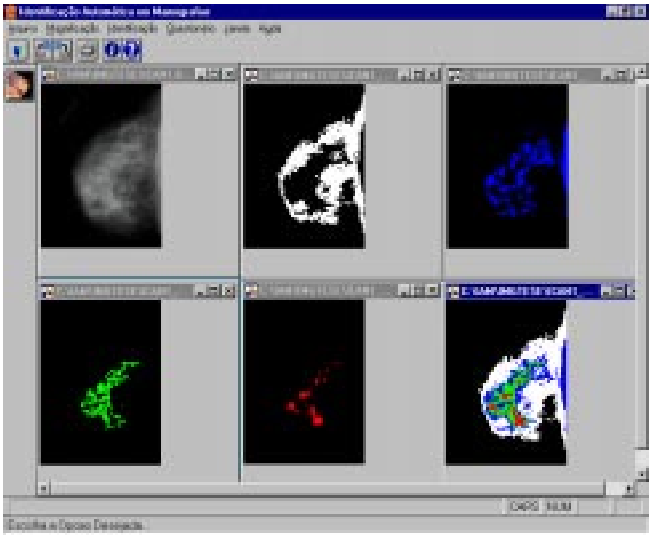

Figure 1: Category Equal

Figure 1 depicts the results in the category *Equal*, in addtion to producing a colour segmented image which facilitates radiological inspection. The multi threshold values obtained in this case by Algorithm 1 are $54, 96, 118$ e 131.

Figure 2 depicts the results in the category *Better* which show lesions not clearly detectable by visual radiological interpretation.

The second set of experiments were conducted on

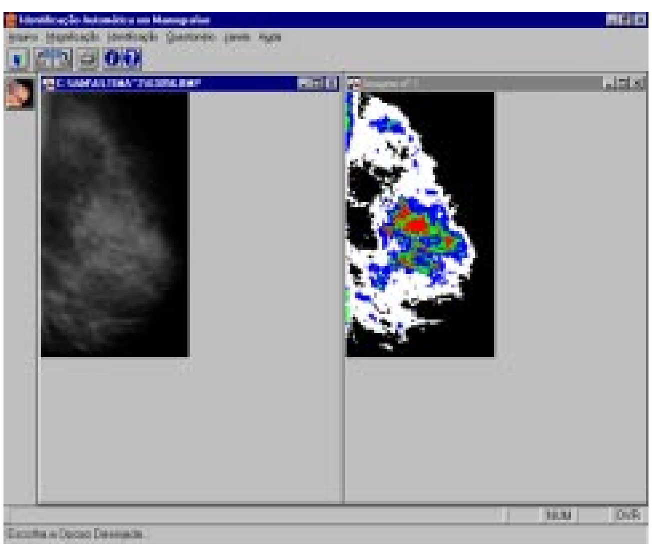

Figure 2: Category Better

images obtained from the mammographic database of the University of South Florida. These database contains mammograms and results provided by radiological experts. Of the 25 tests conducted, the results were in accordance with the information provided in the database by radiological experts. Because the images are of high resolutions, the results were even more impressive. Figure 3 depicts an axample of a multiscale segmented mammogram obtained from the database. In the course of all

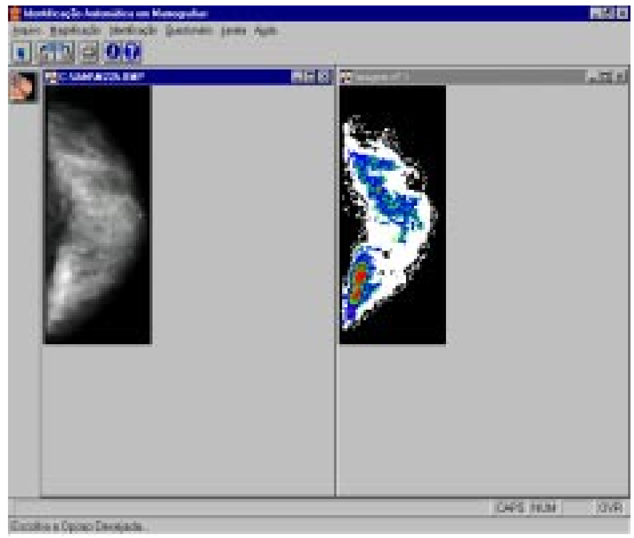

Figure 3: Segmented and Enhanced Image, Courtesy USF

experiments, all objects of small and varying sizes are segmented and enhanced. In addition to the algorithm, a user friendly window based graphical interface is provided. The final results produce a *breast map* as can be observed in the figures which provides an adequate basis for visual radiological breast tissue differentiation and analysis in digital mammography.

5 Conclusion

A multiscale method for segmenting and enhancing objects of various sizes in mammograms is presented. The method utilises an automatic threshold estimator based on histogram moments to segment the mammogram at multilevels. Finally, the segmented image is converted using pseudo-colour mapping to produce a colour image.

An algorithm is presented as well as experimental results. The results produce a *breast map* which provides an adequate basis for radiological breast tissue differentiation and analysis in digital mammography. Experimental results and judgments from radiological experts are very encouraging.

6 Acknowledgements

The authors wish to thank *Clinica Maranhense LtD* for their technical help and support and the University of South Florida for providing some of the test images.

References

[1] I. Bankman et al., *Detection of Microcalcification Clusters Using Neural Networks*, IEEE, 590-591, 1994.

[2] D. Brzakovic et al., *An Approach to Automate Detection of Tumors in Mammograms*, IEEE Trans. Medical Imaging, **9(3)**, 233-241, 1990.

[3] J. Caseldine, R. Blamey, E. Roebuck & C. Elston, *Breast Disease for Radiographers*, Wright, 1988.

[4] J. Dengler et al., *Segmentation of Microcalcification in Mammograms*, IEEE Trans. Medical Imaging, **12(4)**, 634-642, 1993.

[5] R. J. Fogler et al., *SAR Clustter Suppression Using Probability Density Skewness*, IEEE Trans. Aerospace & Electronic Systems, **30(2)**, 622-626, 1994.

[6] R. C. Gonzalez & R. C. Woods, *Digital Image Processing*, Addison-Wesley, 1992.

[7] A. F. Laine et al., *Mammographic Feature Enhancement by Multiscale Analysis*, IEEE Trans. Medical Imaging, **13(4)**, 725-738, 1994.

[8] N. Petrick et al., *An Adaptive Density-Weighted Contrast Enhancement Filter for Mammographic Breast Mass Detection*, IEEE Trans. Medical Imaging, **15(1)**, 59-67, 1996.

[9] L. Shen el al., *Application of Shape Analysis to Mammographic Calcification*, IEEE Trans. Medical Imaging, **13(2)**, 263-274, 1994.

[10] L. Tabár & P. B. Dean, *The Control of Breast Cancer through Mammography Screening: What is the Evidence*, Radiol. Clin. N. Amer., **25(5)**, 993-1005, 1987.

[11] P. G. Tahoces et al., *Enhancement of Chest and Brest Radiographs by Autimatic Spatial Filtering*, IEEE Trans. Medical Imaging, **10(3)**, 330-335, 1991.

[12] J. P. Thiran & B. Macq, *Morphological Feature Extraction for the Classification of Digital Images of Cancerous Tissues*, IEEE Trans. on Biomedical Engineering, **43(10)**, 1011-1020, 1996.

[13] S. L. Wood et al., *Segmentation of Gray Scale Sampled Images with Bimodal Source Models*, IEEE, 456-460, 1992.

[14] S. L. Wood, J. P. Marks & J. Pearlman, *A segmentation Algorithm for OCR Application to Low Resolution Images*, Proc. 14^{th} Asilomar Conf. on Circuits, Systems and Computers, 136-140, 1980.

[15] S. L. Wood, *Automatic Threshold Magnification Control of Images of Print*, Proc. 15^{th} Asilomar Conf. on Circuits, Systems and Computers, 197-201, 1981.