

ENHANCED MULTI-LEVEL THRESHOLDING SEGMENTATION AND RANK BASED REGION SELECTION FOR DETECTION OF MASSES IN MAMMOGRAMS

Alfonso Rojas Domínguez and Asoke K Nandi

Department of Electrical Engineering and Electronics, University of Liverpool, UK

ABSTRACT

A method for detection of masses in mammograms is presented. This method follows the general scheme of: (1) preprocessing of the image to increase the signal-to-noise ratio of the lesions being detected, (2) segmentation of all potential lesions, and (3) elimination of false-positive findings. An algorithm for enhancement of mammograms is proposed which has the objective of improving the segmentation of distinct structures in mammograms. The enhancement algorithm uses wavelet decomposition and reconstruction, morphological operations, and local scaling. After preprocessing, the segmentation of regions is performed via conversion to binary images at multiple threshold levels, and a set of features is computed from each of the segmented regions. A ranking system based on the features computed is also presented. This system is employed to select the regions representing abnormalities. The method was tested on 57 mammographic images of masses from the mini-MIAS database, including circumscribed, spiculated, and ill-defined masses. In this test, the proposed method achieved a sensitivity of 80% at 2.3 false-positives (FPs) per image.

Index Terms— Medical image processing, breast cancer, breast masses, mammography, tumor detection

1. INTRODUCTION

Breast cancer is the most common form of cancer in the female population, affecting one in approximately eleven women at some stage of their life in the Western world [1, 2]. Early detection of breast cancer can be achieved through mammography screening programs assisted by computers [3]. In the past, several researchers have studied and proposed methods for computer-aided detection and classification of abnormalities related to breast cancer in mammograms [4–9].

Kegelmeyer *et al.* [10] investigated the detection of spiculated lesions on mammograms. Their study included 36 positive cases and 49 negative cases. Their method achieved 97% sensitivity with an average of 0.28 FPs per image.

Polakowski *et al.* [8] presented a model-based vision algorithm to detect and classify masses in mammograms. Their

algorithm was tested on 272 images, achieving a 92 % sensitivity in locating malignant masses at an average of 1.8 FPs per image.

Mudigonda *et al.* [11] presented a mass detection method that performs segmentation of objects based on isointensity contours and texture flow-field analysis. Their study included 43 masses and 13 normal cases from the Mini-MIAS database [12]. The performance of their method was reported as 81% of detection success with an average of 2.2 FPs per image.

A density-weighted adaptive contrast enhancement filter was used by Petrick *et al.* [7] as part of a mass-detection algorithm. Petrick *et al.* tested this algorithm on a dataset including 156 malignant masses. The reported detection rate was 87%, at 1.5 marks per mammogram.

In this paper our interest is focused on the detection of masses, either benign or malignant, based on a novel ranking system for the selection of regions representing masses, including well-defined circumscribed, spiculated, and ill-defined masses.

2. MATERIALS

The database of mammograms used in this study is known as MIAS (Mammographic Image Analysis Society) Mini Mammographic Database [12]. In the Mini-MIAS database, the MIAS Database (an earlier version digitized at 50 μm pixel size) has been downsampled to 200 μm pixel size and adjusted to 1024×1024 pixels.

Functions involving wavelets were implemented with code from [13]. Some functions for shape properties are from [14]. The functions used to shift the images are part of the TEMPLAR Software Package [15].

3. METHODOLOGY FOR MASS DETECTION

A method for the detection of masses in mammograms is proposed, which is divided into three main stages. The first stage is an enhancement procedure, which is different from others in the literature in that it incorporates morphological, wavelet, and histogram-based operations. After enhancement and segmentation, several shape and gray-level characteristics of the segmented regions are computed, and a ranking system is employed to select suspicious regions. This ranking system is a

A. Rojas is supported by the National Council of Science and Technology (CONACYT) of Mexico.

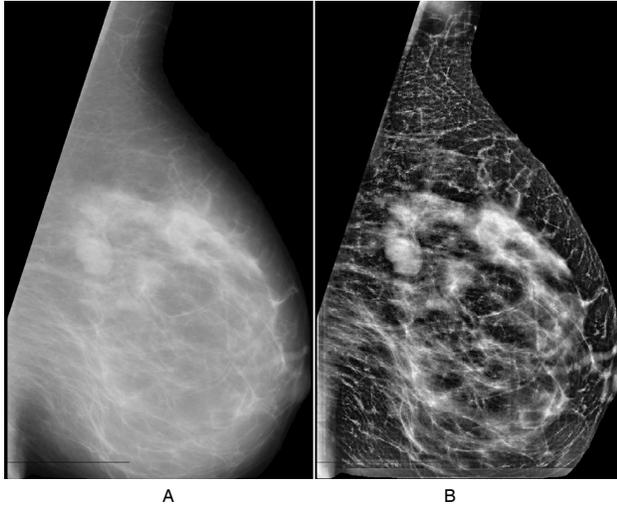


Fig. 1. Example of the effect of the proposed enhancement routine. A: Original mammogram. B: Enhanced version.

Table 1. Possible cases and subroutines for block processing.

Value of k	Shift direction	Subroutine
3	vertical	A
3	horizontal	B
5	vertical	B
5	horizontal	A
10	vertical	B
10	horizontal	B

novel approach to the problem of region selection (i.e., elimination of FPs) that does not require training, and implements a type of on-the-fly feature selection.

3.1. Mammogram enhancement

An image enhancement procedure is proposed which has the objective of increasing the contrast between mammographic structures and their background while providing a relatively uniform intensity to all of the structures. Below, the procedure is described in detail:

First, the images are filtered with a Gaussian smoothing filter to eliminate noise and decrease the effect of outliers. Secondly, the top-hat operation is applied to eliminate the background (using a disk with radius equal to 80 pixels as the structuring element). In the next step, the output of the top-hat operation is decomposed in three scales using wavelet decomposition, and the image is reconstructed using only the detail component of the second scale since this contains most of the mass-boundary information. Following this, the image is processed as shown in Figure 2 with the parameters: $k = 3$,

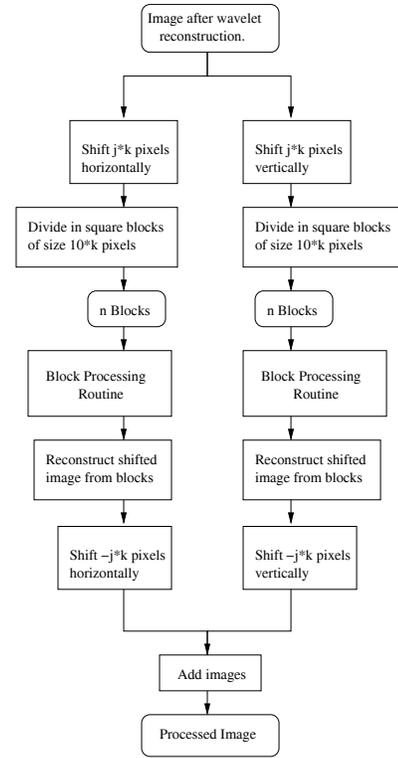


Fig. 2. Block diagram of the proposed enhancement routine.

5, and 10 (k was selected with base on the size of the masses we wish to detect); and $j = 1$ to 9 in steps of 1. Table 1 indicates which of the block processing routines shown in Figure 3 is employed. The choice of routines is designed to avoid artifacts in the output image. Finally, the maximum value between the images processed with each k value is chosen for each pixel.

Figure 1 illustrates the effect of the enhancement routine with an original mammogram and the corresponding enhanced version. It can be observed that all structures at different scales are easily distinguishable; in particular, the structures close to the breast boundary appear much clearer than in the original mammogram.

3.2. Segmentation and feature extraction

The enhanced images are converted to binary images through thresholding at different values starting from the top level. It was found that for the enhanced images in this study, with gray values in the range $[0, 1]$, 30 levels with a step size of 0.025 were adequate to segment all the mammograms.

Once the segmentation procedure is completed, the binary images are filtered with a Gaussian smoothing filter (parameters $\mu = 9$ pixels and $\sigma = 5$ pixels) to eliminate noise (any isolated pixels) and split regions that are joined by single pixels or by a small group of pixels. The images remain binary

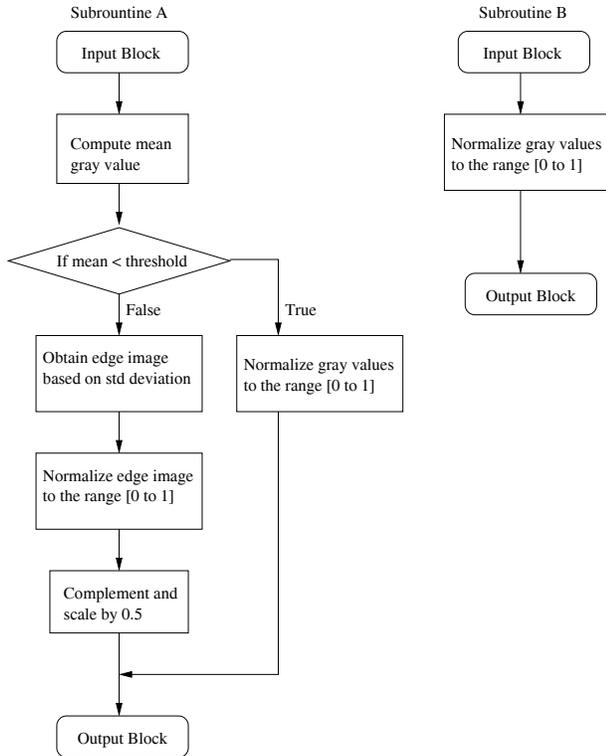


Fig. 3. Block processing subroutines. The threshold in Subroutine A was set to 0.6 for images with pixel values in the range [0, 1].

images.

To complete this stage of the detection method, a set of properties of the remaining regions are computed and stored together with the binary image containing all regions at the corresponding segmentation levels. The properties obtained from each region are: 1. area, 2. perimeter, 3. major axis length, 4. minor axis length, 5. eccentricity, 6. orientation, 7. equivalent diameter, 8. solidity, 9. extent, 10. compactness, 11. dispersion-I, 12. dispersion-II (a variation), 13. mean gradient within region, 14. mean gradient of boundary, 15. gray value variance, 16. edge distance variance, 17. mean intensity difference, and 18. fractal dimension. All of these measures were computed using the gradient and intensity values of the enhanced mammogram except for the *Fractal Dimension*, which was computed using an adaptation of the method of Caldwell *et al.* [16], from the original mammogram.

3.3. Selection of suspicious regions

The selection of suspicious regions is performed by means of a ranking system. By considering how many of the properties of each given region are concentrated around a reference value and within a fixed range (called the *scoring zone*), a rank can be assigned to each region. The reference value for

each property is the mean value of that property computed over the set of masses, and is located at the center of the scoring zone. The range defining the extent of the scoring zone is the standard deviation of the property times a regularization factor α . The rank of the i -th region is mathematically expressed as

$$\begin{aligned} \bar{Z}_i &= [|\bar{x}_i - \bar{\mu}| \leq \alpha \bar{\sigma}], \\ \text{rank}_i &= \|\bar{Z}_i / \bar{\sigma}\|, \end{aligned} \quad (1)$$

where \bar{x}_i the set of properties, $\bar{\mu}$ is the set of means and $\bar{\sigma}$ is the set of standard deviations. We use $[\cdot]$ to clarify that \bar{Z}_i receives the outcome of the test condition $|\bar{x}_i - \bar{\mu}| \leq \alpha \bar{\sigma}$, which is 1 if the condition is true and zero otherwise. The value of the parameter α was chosen empirically, and fixed to 1.9 for all experiments. Once the ranks of all regions are computed, the algorithm selects the ones with high ranks up to a desired number of regions.

4. RESULTS AND DISCUSSION

The algorithm for mass detection was tested on a set of 57 mammograms from the Mini-MIAS database including circumscribed, spiculated and ill-defined masses. A true positive (TP) was recorded for a segmented region when the region overlapped the centroid of a mass, represented by a circular area with a radius of five pixels. Otherwise, the region was considered as FP.

The algorithm was tested with four sets of properties to test their discrimination power. One set included all the properties, whereas the other three included a subset of these. Subset *A* included all properties except the very basic shape descriptors (i.e., properties 7 to 18). Subset *B* included only the measures corresponding to gray-level characteristics (properties 13 to 18). Subset *C* included only the more advanced shape descriptors (properties 7 to 12). Figure 4 presents a plot of the true positive (TP) fraction achieved using each of the four sets of properties versus the number of FPs per image.

The detection of masses used in this study follows the general scheme of first finding all possible distinguishable regions, and then sorting out which of them actually represent masses in the mammograms. This scheme has the disadvantage that a very large number of regions must be processed, which is costly in computing time and resources. The clear advantage is that the initial sensitivity is high; other advantages are that the design of the algorithm is simple and the implementation does not require complex computations.

5. CONCLUSION AND FUTURE WORK

A computer-aided method for the detection of masses in mammograms has been presented. With a performance of 80% of all types of masses in the test database being successfully

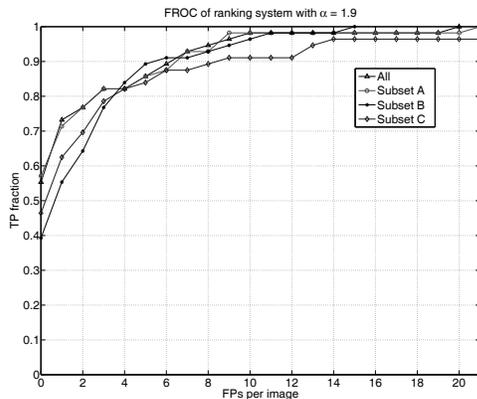


Fig. 4. True positive (TP) fraction vs false positives (FPs) per image.

detected at 2.3 FPs per image, this algorithm compares well with other methods in the literature. Combining this algorithm with other detection methods, refining the system for FP reduction, and including a feature selection step are being considered for future work.

ACKNOWLEDGEMENT

The authors acknowledge the effort of the researchers who made the software and database used in this study freely available. We would also thank Prof. Rangaraj M. Rangayyan of the University of Calgary, Alberta, Canada and Dr. Sonu Punnoose of the University of Liverpool, UK, for their insightful observations.

6. REFERENCES

- [1] Cancer Research UK, *Breast Cancer Factsheet - February 2004*, February 2004.
- [2] National Cancer Institute of Canada, *Canadian cancer statistics 2006*, 2006.
- [3] C. Di Maggio, "State of the art of current modalities for the diagnosis of breast lesions," *European Journal on Nuclear Medicine and Molecular Imaging*, vol. 31, no. Supplement 1, pp. S56–S69, June 2004.
- [4] R. M. Rangayyan, L. Shen, Y. Shen, M. S. Rose, J. E. L. Desautels, H. E. Bryant, T. J. Terry, and N. Horeczko, "Region-based adaptive contrast enhancement," in *Image Processing Techniques for Tumor Detection*, R. N. Strickland, Ed., pp. 213–241. Marcel Dekker, Inc., New York, USA, 2002.
- [5] H. Li, Y. Wang, K. J. R. Liu, S. B. Lo, and M. T. Freedman, "Computerized radiographic mass detection - Part I: Lesion site selection by morphological enhancement and contextual segmentation," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 289–301, April 2001.
- [6] N. Petrick, H.-P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Trans. Med. Imag.*, vol. 15, no. 1, pp. 59–67, February 1996.
- [7] N. Petrick, B. Sahiner, H.-P. Chan, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evaluation of a mass-detection algorithm for computer-aided diagnosis - experience in 263 patients," *Radiology*, vol. 224, no. 1, pp. 217–224, July 2002.
- [8] W. E. Polakowski, D. A. Cournoyer, S. K. Rogers, M. P. De Simio, D. W. Ruck, J. W. Hoffmeister, and R. A. Raines, "Computer-aided breast cancer detection and diagnosis of masses using difference of gaussians and derivative-based feature saliency," *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 811–819, December 1997.
- [9] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, no. 5, pp. 369–378, May 1999.
- [10] W. P. Kegelmeyer Jr, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology*, vol. 191, no. 2, pp. 331–337, May 1994.
- [11] N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1215–1227, December 2001.
- [12] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, "The Mammographic Image Analysis Society digital mammogram database," *Excerpta Medica, International Congress Series 1069*, pp. 375–378, 1994.
- [13] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Prentice Hall, New Jersey, USA, 2004.
- [14] M. Nixon and A. Aguado, *Feature Extraction & Image Processing*, Newnes, Oxford, UK, 2001.
- [15] C. Scott, "TEMPLAR software package, copyright 2001, Rice University," 2001.
- [16] C. B. Caldwell, S. J. Stapleton, D. W. Holdsworth, R. A. Jong, W. J. Weiser, G. Cooke, and M. J. Yaffe, "Model-based detection of spiculated lesions in mammograms," *Physics in Medicine and Biology*, vol. 35, no. 2, pp. 235–247, February 1990.