FAST DETECTION OF MASSES IN DIGITIZED MAMMOGRAMS

I. Christoyianni, E. Dermatas and G. Kokkinakis

WCL, Electrical & Computer Engineering Dept.
University of Patras, 26100 Patras, HELLAS
e-mail: ioanna@ios.wcl2.ee.upatras.gr

ABSTRACT

A novel method for fast detection of regions of suspicion (ROS) that contain circumscribed lesions in mammograms is presented. The position and the size of ROS are first recognized with the aid of a Radial-Basis-Function neural network (RBFNN) by performing windowing analysis. Then a set of criteria is employed to these regions to make the final decision concerning the abnormal ones. Accelerated estimation of the high-order statistical features decreases the computational complexity 55 times in multiplication operations. The proposed method detects the exact location of the circumscribed lesions with accuracy of 72.7% (overlap between groundtruthed and detected regions greater than 50%) for mammograms containing masses, while the recognition rate for the normal ones reaches 77.7% in the MIAS database.

1. INTRODUCTION

Breast cancer is a leading cause of fatality in women, with approximately 1 in 12 women being affected by the disease during their lifetime. Mass screening using X-ray mammography is currently the most effective method of early detection of the disease. The radiographs are searched for signs of abnormality by expert radiologists but mammograms are complex in appearance and signs of early disease are often small or subtle. That's the main cause of many missed diagnoses, that can be mainly attributed to human factors such as subjective or varying decision criteria, distraction by other image features, or simple oversight [1,2]. However, the consequences of errors in detection or classification are costly. Since the advent of mass screening, there has been a considerable interest in developing methods for automatically detecting mammography abnormalities, as means of aiding radiologists and improving the efficacy of screening programs. Masses and clustered microcalcifications often characterize early breast cancer [3]. Masses appear as dense regions of varying sizes and properties. The lesions can be circumscribed, lobulated, spiculated, or ill-defined. The emphasis of this paper is given to the diagnosis of the abnormal mammograms that contain circumscribed masses.

The use of computer-aided diagnosis (CAD) has been recently proposed as a "second-opinion" strategy for breast screening [4,5]. Specifically, CAD for detecting lesions in mammograms is of great interest to many researchers worldwide [6-9]. In particular, neural network based CAD systems have already been applied to a variety of pattern-recognition tasks such as microcalcifications detection and specification and have proven as a potentially powerful tool. However, the use of neural networks in case of other abnormalities and particularly in the most frequently encountered circumscribed masses, has only been very limited.

In this paper, we present a complete method for fast detection of circumscribed mass lesions in mammograms employing a Radial-Basis Function Neural Network (RBFNN). This method is able to make a decision whether a mammogram is normal or not and then detects the masses' position by performing sub-image windowing analysis. In the latter case, with the implementation of a set of criteria, square regions containing the masses are marked as region of suspicion. A fast feature extraction reduces significantly the overall processing time allowing implementation of the method in low cost PCs.

The structure of this paper is as follows: In the next section a detailed description of the proposed method is given. In section 3 the computational efficiency of the features extraction module is estimated. In section 4 we present the data set and our experimental results and finally in section 5 some conclusions are drawn.

2. THE PROPOSED METHOD

The basic scheme of the proposed method is shown in Figure 1. It consists of the image preprocessing and feature extraction

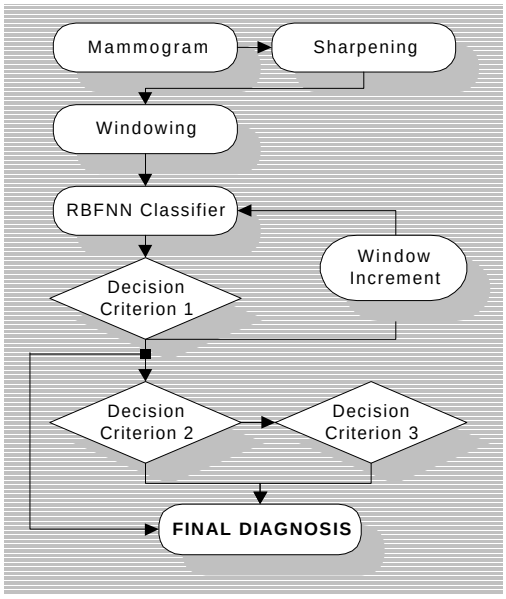

Figure 1. Structure of the proposed detector of circumscribed masses

steps, the neural network classifier and the detection criteria for circumscribed mass identification.

2.1 Image pre-processing

Lesions in mammograms usually have different gray level values from their background. As a first step, a sharpening filter is applied to the mammogram in order to maximize the contrast value between the masses and the local background.

2.2 Feature Extraction

The implemented feature extraction procedure relies on the texture, which is the main descriptor for all kinds of mammograms. Therefore, statistical descriptors that depend on calculating averages, standard deviations, and higher-order statistics of intensity values are used for texture description.

After image sharpening, a successive windowing analysis is performed by moving a testing window in 5-pixel increments. The window on the tested mammogram follows the path as shown in Figure 2 in order to scan the overall image area. The applied path permits the isolation of three image areas at any time of the windowing analysis (Figure 2). The successive windows W(t-1) and W(t) are the union of the following areas:

$$W(t-1) = S_o(t) \cup S(t), \qquad W(t) = S(t) \cup S_n(t)$$

The overall feature extraction process includes the initialization module and the recursive part as follows:

Initially, four statistical features are extracted from the window W(0):

$$\mu_1(0) = \frac{1}{N+n} \sum_{x_i \in W(0)} x_i \qquad \mu_2(0) = \frac{1}{N+n} \sum_{x_i \in W(0)} (x_i - \mu_1(0))^2$$

$$\mu_3(0) = \frac{1}{N+n} \sum_{x_i \in W(0)} (x_i - \mu_1(0))^3 \qquad \mu_4(0) = \frac{1}{N+n} \sum_{x_i \in W(0)} (x_i - \mu_1(0))^4$$

where, *N* denotes the number of pixels in the image area S(t), n is the number of pixels in the areas $S_o(t)$, $S_n(t)$, and x_i is the gray level value of the i^{th} pixel.

Taking into account that the first four statistical moments of the image areas $S_o(t)$ and $S_n(t)$ can be estimated using trivial computing resources (N>>n),

$$\alpha_k(t) = \sum_{x_i \in S_o(t)} x_i^k \qquad c_k(t) = \sum_{x_i \in S_o(t)} x_i^k$$

the following high-order statistical features can be estimated using the recursive equations:

$$\mu_1(t) = \mu_1(t-1) + \frac{1}{N+n}(c_1(t) - a_1(t))$$

$$\mu_2(t) = \mu_2(t-1) + \mu_1^2(t-1) - \mu_1^2(t) + \frac{1}{N+n}(c_2(t) - a_2(t))$$

$$\mu_3(t) = \mu_3(t-1) + \mu_1(t-1)(\mu_1^2(t-1) + 3\mu_2(t-1)) - \mu_1^2(t) - \mu_1(t)(\mu_1^2(t) + 3\mu_2(t)) + \frac{1}{N+n}(c_3(t) - a_3(t))$$

$$\mu_4(t) = \mu_4(t-1) - 4\mu_3(t)\mu_1(t) + 4\mu_3(t-1)\mu_1(t-1) - 6\mu_2(t)(\mu_1^2(t) - \mu_1^2(t-1)) - \mu_1^4(t) - 5\mu_1^4(t-1) + 6\mu_1^2(t)\mu_1^2(t-1) + \frac{1}{N+n}(c_4(t) - a_4(t) + 6\mu_1^2(t-1)(c_1(t) - a_3(t)))$$

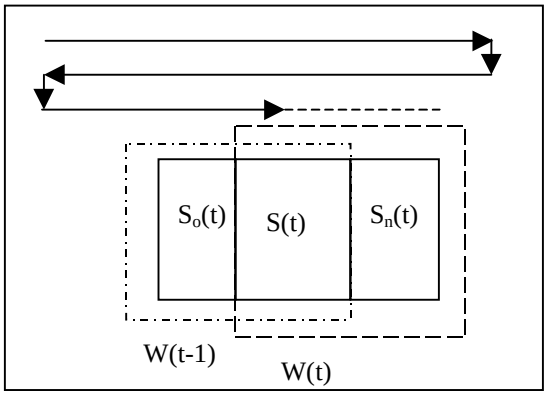

Figure 2. Path of the Sub-image windowing analysis and the overlapping areas of successive windows

From the above, the mean, variance, skewness and the kurtosis statistical features employed in our method are estimated for each window:

$$Mean(t) = \mu_1(t) \qquad Variance(t) = \mu_2(t)$$

$$Skewness(t) = \frac{\mu_3(t)}{\mu_2(t)} \qquad Kurtosis(t) = \frac{1}{4}\mu_4(t) - 3$$

All features are normalised by their sample means and standard deviations. The whole mammogram is scanned repeatedly by increasing the window size from a minimum of 35x35 pixels to its maximum of 400x400 pixels, with a step size of 5 pixels.

2.3 Neural Network Classifier

Neural networks have been widely used in situations where the expert knowledge is not explicitly defined and cannot be described in terms of statistically independent rules. A radial-basis-function neural network (RBFNN) is employed as proposed in [10,11]. The input layer handles the four features extracted from each testing-window. Two output units denote the presence or absence of a lesion. A hidden layer with five nodes is located between the input and the output layer. The number of hidden nodes was estimated experimentally for the optimal diagnosis of the circumscribed lesions.

2.4 Decision Criteria

Normally the neural classifier detects a great number of suspicion regions in a tested mammogram; i.e. the neuron that corresponds to the presence of a lesion is the most activated output neuron. The main goal of the implemented criteria is to select the most important region of suspicion (ROS), otherwise a normal mammogram is diagnosed. This comes out progressively as a result of a sequential evaluation of three acceptance/rejection criteria. The proposed diagnosis procedure is presented schematically in Figure 1.

- **Decision Criterion 1**. *Each ROS is considered to be abnormal if and only if the neural network classifies as normal all its neighbours with the same size (8 windows).* The above criterion is applied after the scanning of the mammogram with a variable window size is completed and reduces dramatically the amount of the abnormal detected regions.
- **Decision Criterion 2.** *A ROS is considered as a candidate circumscribed mass (CCM) if the value of the most activated neuron output is greater than a threshold Th. Otherwise, the mammograph is classified as normal.* Th is experimentally defined (0.65 in our experiments).
- **Decision Criterion 3.** *If the candidate-circumscribed masse's mean intensity is higher than a threshold value Tm, the method diagnoses a circumscribed lesion. Otherwise, the mammograph is classified as normal.* This threshold value is chosen according to the character of the mammogram background tissue as can be seen in Table 1.

Background Tissue	Threshold value Tm
Fatty	> 138
Fatty-Glandular	> 170
Dense-Glandular	> 180

Table 1. *Threshold values for three types of background tissue of digital mammograms*

3. FEATURE EXTRACTION COMPLEXITY

The computational complexity of the feature extraction module is significantly reduced when the proposed recursive equations are used. Specifically, for each square window the complexity in multiplication operations is decreased by a factor:

$$ComFactor(x) = \frac{3x^2+15x+6}{15x+21} \cong 0.2x+1$$

where x is the window side. The approximation is valid for window sides greater than 30 pixels. The overall reduction in multiplication operations of the proposed feature extraction method is estimated by the following equation:

$$Overall\,\mathrm{Re}duction = \frac{\sum_{x=7}^{80}(75x^2+75x+6)}{\sum_{x=7}^{80}(75x+21)} = 54.6386$$

4. EXPERIMENTAL RESULTS

4.1 Data set

For our experiments the MIAS MiniMammographic Database [12] provided by the Mammographic Image Analysis Society (MIAS) was used. The mammograms are digitized at 200-micron pixel edge, resulting to a 1024x1024-pixel resolution.

There are a total of 22 mammograms containing circumscribed lesions. The smallest lesion extends to 18 pixels in radius, while the largest one to 198 pixels. For the training procedure 22 groundtruthed abnormal regions from the 22 mammograms, along with 22 randomly selected normal regions were used. This results in a training data subset of 44 regions.

For the evaluation of the proposed method we used all the abnormal mammograms from the MIAS database that contain circumscribed masses (22 images) together with 54 entirely normal mammograms that were randomly selected.

The MIAS database provides groundtruth for each lesion in the form of circles, which indicate the approximate center and radius of each abnormality. Therefore, since circumscribed lesions are rarely perfectly circular, and since the MIAS policy was to err on the side of making the groundtruth circles completely inclusive rather than too small, these regions often contain a substantial amount of normal tissue as well.

4.2 Classification Results

For the validation of the circumscribed lesion detection method we employed an objective 50% overlap criterion. In particular, if the area of the groundtruth circle, approximated with a square region for reasons of compatibility with our testing windows form, overlaps the area of the detected window by at least 50%, then the detection is considered as a true positive (TP), otherwise the detection is a false positive (FP). This is similar to the validation strategy employed by Woods [13] and Kegelmeyer [14].

For the case of the abnormal mammograms, the proposed method diagnosed correctly 16/22 mammograms with circumscribed lesions that satisfied the above validation criterion resulting to 72,72% True Positive Rate (TPR). On the other hand, the method missed 6 cases; in five of them the detected regions had a common area less than 50% of the groundtruthed region while the last one was miss-classified as normal. For the normal mammograms, 42/54, a percentage of 77.7%, was correctly classified as normal (no suspicion region was found) and the remaining 12/54 (22.22%) was misdiagnosed as abnormal. Table 2 shows analytically the experimental results for the testing set of the 76 mammograms. The mean overlap value for the true positive mammograms was found to be 0.868.

In addition, the method achieved to diagnose successfully even cases that were *hard-to-diagnose*. As shown in Figure 3 our method detected the abnormality in usual (case1, case2) as well as *hard-to-diagnose* cases (case3). However, a significant factor that affects the performance of the overall method is the character of the mammogram's background tissue. In the case of dense background tissue the methods effectiveness decreases which can be attributed to the small number of the dense tissue mammograms with circumscribed masses in the MIAS Database.

	Fatty	Glandular	Dense	**TOTAL**
Abnormal	90.9 %	62.5 %	33.3 %	**72.7 %**
Normal	73.6 %	100 %	66.6 %	**77.7 %**
TOTAL	78 %	83.33 %	55.5 %	**75.2 %**

Table 2 *Recognition results*

5. CONCLUSION

In this paper we presented a novel method based on the RBFNN classifier and a set of decision criteria capable of making a decision whether any given mammogram contains circumscribed masses or not. The achieved results are promising so that work is continued towards increasing the location accuracy of the circumscribed lesions and reducing the number of false diagnoses of mammograms. To this end training and testing of the RBFNN with a great number of mammograms is carried out while extending and refining the decision criteria.

REFERENCES

[1] J. Martin, M. Moskowitz, and J. Milbrath, "Breast cancer missed by mammography", *AJR*, Vol.132, p. 737, 1979.

[2] L. Kalisher, "Factors influencing false negative rates in xero-mammography", *Radiology*, Vol.133, p. 297, 1979.

[3] L. Tabar and P.B.Dean, *Teaching Atlas of Mammography*, 2nd edition, Thieme, NY, 1985.

[4] M. Giger, "Computer-aided diagnosis", *RNSA Syllabus: A Categorical Course in Physics, Technical Aspects of Breast Imaging*, eds, A Haus, M Yaffe, 1993; 283-298.

[5] C. Vyborny, M. Giger, "Computer vision and intelligence in mammography ", *AJR*, vol.162, pp.699-708, 1994.

[6] Z. Huo, M. Giger, C. Vyborny, D. Wolverton, R. Schmidt, K.Doi, "Computer-Aided Diagnosis: Automated Classification of Mammographic Mass Lesions", *Proc. of the 3rd Int. Workshop on Digital Mammograph*, Chicago, IL, 1996, pp.207-211.

[7] N. Petrick, H. Chan, B. Sahiner, M. Helvie, M. Goodsitt, and D. Adler, "Computer-Aided Breast Mass Detection: False Positive Reduction using Breast Tissue Composition", *Proc. of the 3rd Int. Workshop on Digital Mammograph*, Chicago, IL, 1996, pp.373-378.

[8] H. Chan, D. Wei, M. Helvie, B. Sahiner, D.Adler, M. Goodsitt, and N. Petrick, "Computer-Aided Classification of Mammographic Masses and Normal Tissue: Linear Discriminant Analysis in Texture Feature Space", *Phys. Med. Biol.*,Vol.40, pp.857-876, 1995

[9] U. Bick, M. Giger, R. Schmidt, R. Nishikawa, D. Wolverton, K. Doi, "Computer-Aided breast cancer detection in screening mammography", *Proc. of the 3rd Int. Workshop on Digital Mammograph*, Chicago, IL, 1996, pp.97-103.

[10] X. Zong, A. Meyer-Baese, and A.Laine "Multiscale segmentation through a radial basis neural network", *ICIP*, 1997, Vol. III, pp.400-403.

[11] A.P. Dhawan, Y. Chite, C. Bonasso and K. Wheeler "Radial-Basis-Function Based Classification of Mammographic Microcalcifications Using Texture Features", *IEEE Engineering in Medicine and Biology & CMBEC*, 1995, pp.535-536.

[12] http://s20c.smb.man.ac.uk/services/MIAS/MIASmini.html

[13] K.S. Woods, "Automated Image Analysis Techniques for Digital Mammography", *Ph.D. Dissertation*, University of South Florida, December 1994.

[14] W. Kegelmeyer, J. Pruneda, P. Bourland, A. Hillis, M. Riggs, and M. Nipper, " Computer-Aided mammographic screening for spiculated lesions", *Radiology*, vol. 191, pp. 331-337, 1994.

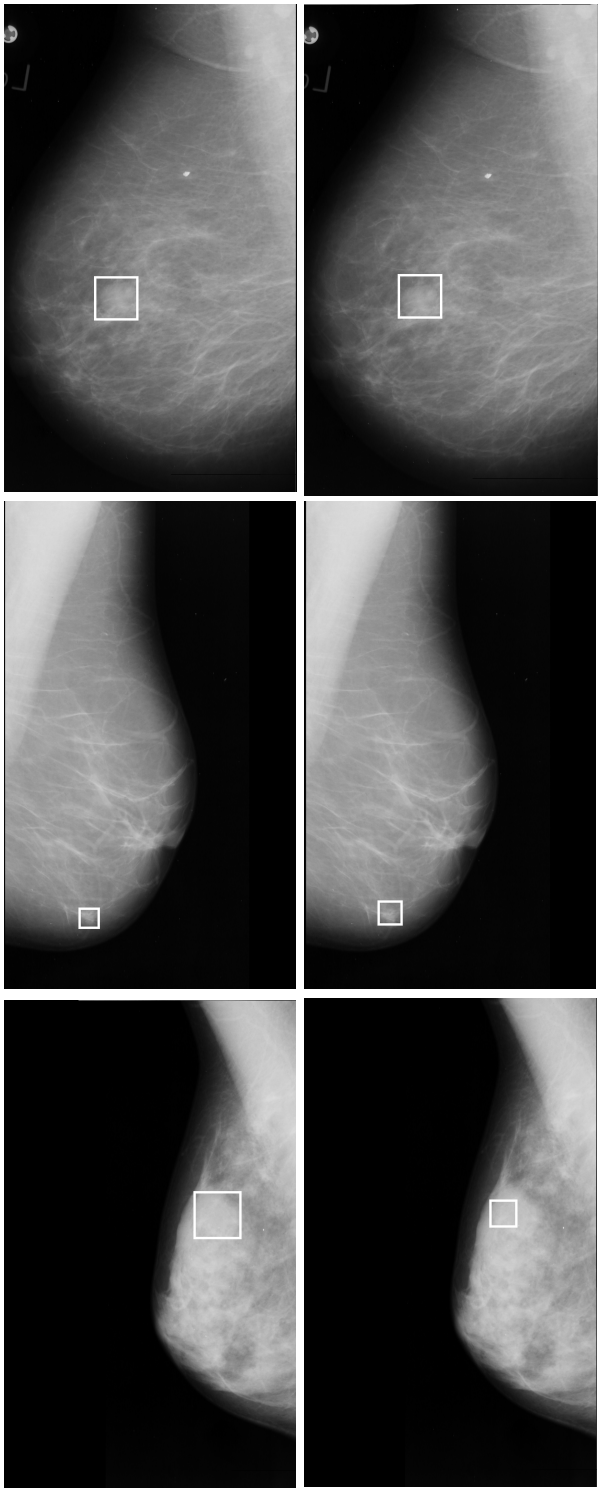

Figure 3 Left: Mammograms with tumor as determined by expert radiologists; Right: The same mammogram with tumor as detected by the proposed method.