İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl: 4 Sayı: 8 Güz 2005/2 s.87-97

COMPUTER AIDED DETECTION OF MAMMOGRAPHIC MASSES ON DIGITAL MAMMOGRAMS

Serhat ÖZEKES [*], A.Yılmaz ÇAMURCU [**]

ABSTRACT
This paper presents an automated system for detecting masses in mammogram images. The proposed method is based on a two-step procedure: a. regions of interest (ROI) specification, b. rule based classification of regions of interest. In the first step, the intensity values of pixels in mammogram images are used and scanning the pixels in 8 directions is evaluated. By using various thresholds while scanning the pixels, ROIs are specified. In the second step, all ROIs are labeled using Connected Component Labeling (CCL) and two rules are used to categorize ROIs as true masses or not. These rules are based on euclidean distance and regularity values of the ROIs. To test the system's efficiency, we applied it to images from the Mammographic Image Analysis Society database. The accuracy of the system reaches 88.37% with 0.292 false positives per image.

Keywords: Mammographic Mass Detection, Computer Aided Detection, Mammography

DİJİTAL MAMMOGRAMLARDAKİ MEME KİTLELERİNİN BİLGİSAYAR DESTEKLİ TESBİTİ

ÖZET
Bu çalışmada, mammogram görüntülerindeki kitlelerin otomatik olarak tesbit edilebilmesi için bir sistem geliştirilmiştir. Önerilen yöntem iki basamaklıdır: a. ilgi alanlarının belirlenmesi, b. ilgi alanlarının kural tabanlı sınıflandırılması. İlk aşamada görüntü kesitlerindeki piksellerin yoğunluk değerleri hesaplanmış ve her piksel için 8 yönlü tarama işlemi gerçekleştirilmiştir. Bu tarama işlemi sırasında çeşitli eşik değerleri kullanılarak, ilgi alanları belirlenmiştir. İkinci aşamada, tüm ilgi alanları bağlantılı bileşen etiketleme (BBE) yöntemiyle tanımlanmış ve iki kural kullanılarak ilgi alanları sınıflandırılmıştır. Bu kurallar ilgi alanlarının öklid uzaklıkları ve biçim değerlerini sorgulamaktadır. Sistemin performansı Mammogram Görüntü Analizi Topluluğu veritabanına uygulanarak ölçülmüştür. Sistemin duyarlılığı görüntü başına 0.292 yanlış pozitif değeriyle %88.37'ye ulaşmaktadır.

Anahtar Kelimeler: Meme Kitle Tesbiti, Bilgisayar Destekli Tesbit, Mammografi

[*] *Istanbul Commerce University, Vocational School, Uskudar, Istanbul, serhat@iticu.edu.tr*
[**] *Marmara University, Technical Education Faculty, Goztepe, Istanbul, camurcu@marmara.edu.tr*

1. INTRODUCTION

Breast cancer is the most frequently diagnosed form of cancer and the most common cause of cancer related deaths amongst women in the world. Early detection of breast cancer is the key successful treatment. X-ray mammography is currently the most popular, cost-effective, low radiation dose and relatively accurate method of early detection of the disease. Radiologists carefully search each image for any visual sign of abnormality. However, abnormalities are often embedded in and camouflaged by varying densities of breast tissue structures (Baines et al., 1990; Wallis et al., 1991). Indeed, estimates indicate that between 10 and 30% of breast cancers are missed by radiologists during routine screening (Bird, 1990; Brenner, 1991). Thus, a variety of computer-aided diagnosis (CAD) systems have been proposed to increase the efficiency and effectiveness of screening procedures by using a computer system, as a "second opinion", to aid the radiologist by indicating locations of suspicious abnormalities in mammograms, leaving the final decision regarding the likelihood of the presence of a cancer and patient management to the radiologist (Thurfjell et al.,1994; Vyborny and Giger,1994).

Many research groups in either academia or industry are developing computerized techniques for the detection of abnormalities (e.g., mass lesion or cluster of microcalcifications) in digital mammograms. Many involve the use of classifiers to distinguish between actual lesions and false-positive detections. Examples of these classifiers in CAD techniques include rule-based methods (Nishikawa et al., 1995; Polakowski et al., 1997), Bayesian methods (Kupinski and Giger, 1997), artificial neural networks (Nagel et al., 1998; Kalman et al., 1997), and fuzzy logic (Cheng et al., 1998).

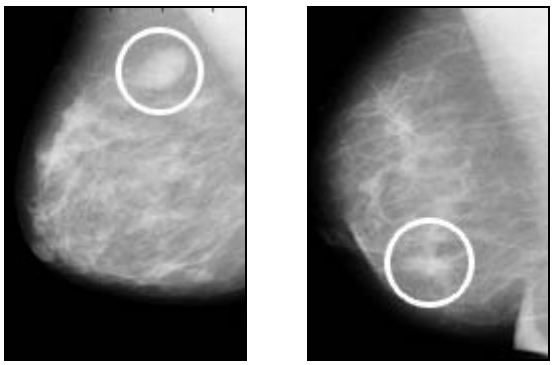
The objective of this research is to automatically detect masses from digital mammograms. Using the intensity values of pixels in mammograms and scanning these pixels in 8 directions with distance thresholds, ROIs are identified. All ROIs are labeled using CCL. Then, two rules are used in the characterization of each ROI as mass or non-mass.

2. MATERIALS AND METHODS

2.1 Image Dataset

For the development and evaluation of the proposed system we used the MiniMIAS (Suckling et al., 1994) database. This database contains left and right breast images for a total of 161 patients with ages between 50 and 65. All images are digitized at a resolution of 1024 X 1024 pixels and at 8-bit grey scale level. All the images also

include the locations of any abnormalities that may be present. The existing data in the collection consists of the location of the abnormality (like the centre of a circle surrounding the tumor), its radius, breast position (left or right), type of breast tissues (fatty, fatty-glandular and dense) and tumor type if it exists (benign or malign). In this study we considered 41 images whose abnormal cases are circumscribed mass (22 images) and spiculated mass (19 images). Figure 1 gives examples to these cases.



a b

Figure 1. a. Mammogram with a Circumscribed Mass, b. Mammogram With a Spiculated Mass.

2.2 Regions of Interest Specification

The mammograms of miniMIAS database present several different areas such as the image background, the tissue area, and informative marks. To segment the ROIs from breast tissue, it is assumed that pixels which form a ROI must be members of a set of adjacent neighbor pixels with suitable intensities. To identify the suitable intensities two thresholds are used that are *"minimum intensity threshold"* and *"maximum intensity threshold"*. It has been observed that diameters of masses are between upper and lower boundaries. So, to understand whether a pixel is in the center region of the ROI, first, diameter of the ROI (assuming the pixel in question is the center) should be considered. In this stage, we introduce two thresholds which form the boundaries. As seen in Figure 2, one is the "*minimum distance threshold*" representing the lower boundary and the other is the "*maximum distance threshold*" representing the upper boundary. If a pixel has adjacent neighbors that are less than "*minimum distance threshold*" or more than "*maximum distance threshold*" in 8 directions, it could be concluded that this pixel couldn't be a part of the ROI. Otherwise, it could be a part of the ROI.

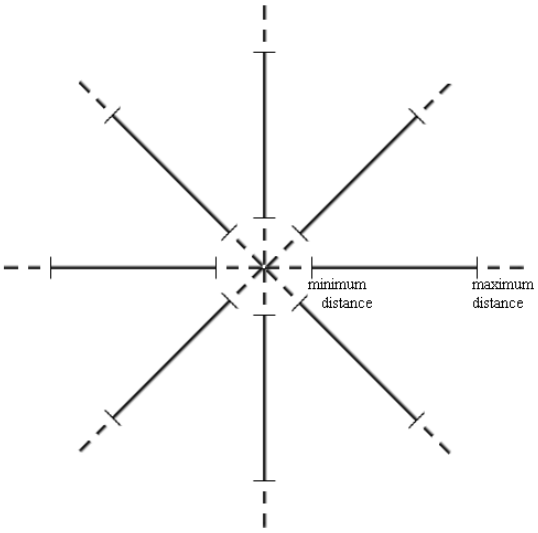
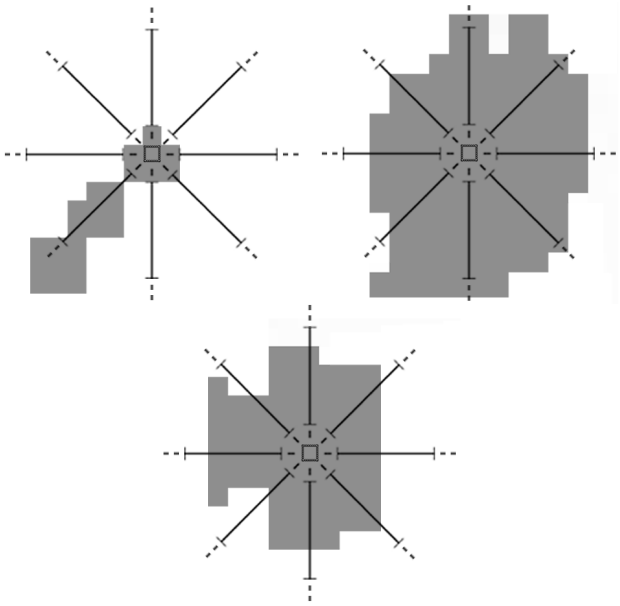


Figure 2. Minimum and Maximum Distance Thresholds in 8 Directions.

Examples of determining the pixels to be a part of ROI can be seen in Figure 3. Assume that in Figure 3 a, b and c, grey pixels have suitable intensities. As seen in Figure 3a, if a grey pixel doesn't have a number of adjacent neighbor grey pixels greater than or equal to the value of "*minimum distance threshold*", or as seen in Figure 3b, if a grey pixel doesn't have a number of adjacent neighbor grey pixels less than or equal to the value of "*maximum distance threshold*" in all directions, it could be considered that the pixel under investigation is not a part of the ROI. Otherwise, as seen in Figure 3c, it could be concluded that the pixel is a part of the ROI. The values of minimum and maximum distance thresholds are dealt with the resolution of the mammogram image. These thresholds are used to avoid very big or very small structures corresponding to artifacts and blood vessel type objects.



a b c

Figure 3. a. A Pixel Which Doesn't Have a Number of Adjacent Neighbor Pixels Greater Than or Equal to the Value of "*Minimum Distance Threshold*", so It is Not a Part of the ROI, b. A Pixel Which Doesn't Have a Number of Adjacent Neighbor Pixels Less Than or Equal to the Value of "*Maximum Distance Threshold*", so It is Not a Part of the ROI, c. A Pixel Which Has a Number of Adjacent Neighbor Pixels Greater Than or Equal to the Value of " *Minimum Distance Threshold* ", and Less Than or Equal to the Value of "*Maximum Distance Threshold*", so It is a Part of the ROI

2.3 Rule Based Classification of Regions of Interest

The objective of the classification module is to categorize the specified ROIs as true masses or non-masses. To find the masses by distinguishing them from normal structures we label each ROI by using connected components labeling (CCL) and record the coordinates of each pixel of each ROI.

CCL works by scanning an image pixel by pixel (from top to bottom and left to right) in order to identify connected pixel regions - *i.e.* regions of adjacent pixels which share the same set of intensity values *V* (Ronse and Devijver, 1984). The CCL

operator scans the image by moving along a row until it comes to a point *p* (where *p* denotes the pixel to be labeled at any stage in the scanning process) for which *V={1}*. When this is true, it examines the four of the neighbors of *p* which have already been encountered in the scan (*i.e.* the neighbors (i) to the left of *p*, (ii) above it, and (iii and iv) the two upper diagonal terms). Based on this information, the labeling of *p* occurs as follows (Manohar and Ramapriyan, 1989; Stefano and Bulgarelli, 1999):

- if all four neighbors are 0, assign a new label to *p*, else
- if only one neighbor has *V={1}*, assign its label to *p*, else
- if one or more of the neighbors have *V={1}*, assign one of the labels to *p* and make a note of the equivalences.

After completing the scan, the equivalent label pairs are sorted into equivalence classes and a unique label is assigned to each class. As a final step, a second scan is made through the image, during which each label is replaced by the label assigned to its equivalence classes. For display, the labels might be different grey levels or colors (see Figure 4).

0	1	1	0	1	0	0
0	1	1	0	1	0	1
1	1	1	0	1	0	1
0	0	0	0	1	1	1
0	1	0	0	0	0	0
0	1	1	1	1	1	0
0	1	1	1	0	0	0

0	1	1	0	2	0	0
0	1	1	0	2	0	2
1	1	1	0	2	0	2
0	0	0	0	2	2	2
0	3	0	0	0	0	0
0	3	3	3	3	3	0
0	3	3	3	0	0	0

Figure 4. Labeling Examples of CCL System.

We observed that ROIs have different morphologies. While masses are thicker and more circular, other structures are thinner and longer. So, to distinguish masses from normal breast structures by using their morphologies, two rules are taken into account. In the first rule, the euclidean distance of the ROI and in the second rule the regularity which is the ratio of euclidean distance to thickness of the ROI, is considered. We now explain these rules explicitly.

Rule 1: The first step of classifying ROIs is calculating their euclidean distances. We use an "*euclidean distance threshold*" and assume that, for a ROI to be a mass candidate, its euclidean distance must be bigger than the "*euclidean distance threshold*". Otherwise, it would be a normal breast structure or artifact.

To calculate the euclidean distance, first, the pixel coordinates of the ROI are searched and maximum x (x_{max}), minimum x (x_{min}), maximum y (y_{max}) and minimum y (y_{min}) coordinates are found. Second, the euclidean distance (E_d) of the ROI is calculated as follows:

$$E_d = \sqrt{(x_{\max} - x_{\min})^2 + (y_{\max} - y_{\min})^2} \tag{1}$$

Rule 2: The second step of classifying ROIs is calculating their thickness and regularity values. The regularity of a ROI is the ratio of its euclidean distance to its thickness. Because masses are thicker and more circular than other structures, it is expected that their thickness is bigger and their regularity is smaller. So, we look for the thickness to compare the regularity of the ROI with "*regularity threshold*". For a ROI to be a candidate mass, its regularity must be smaller than the "*regularity threshold*".

To calculate the thickness of the ROI, we first find the center coordinates. The center coordinates (x_c, y_c) are calculated as follows:

$$x_c = fix\left(x_{\min} + \frac{x_{\max} - x_{\min}}{2} \right) \tag{2}$$

$$y_c = fix\left(y_{\min} + \frac{y_{\max} - y_{\min}}{2} \right) \tag{3}$$

As a further step, $y_{c\min}$ (minimum y coordinate corresponding to x_c), $y_{c\max}$ (maximum y coordinate corresponding to x_c), $x_{c\max}$ (minimum x coordinate corresponding to y_c) and $x_{c\max}$ (maximum x coordinate corresponding to y_c) coordinates are found. Using these values, the height and width are calculated as follows:

$$height = y_{c\max} - y_{c\min} \tag{4}$$

$$width = x_{c\max} - x_{c\min} \tag{5}$$

To find the minimum thickness of the ROI which we denote by S, the smaller one of height and width is chosen for the reasons explained above:

$$S = \min(height, width) \tag{6}$$

And finally, the regularity (R) is calculated as follows:

$$R = \frac{E_d}{S} \tag{7}$$

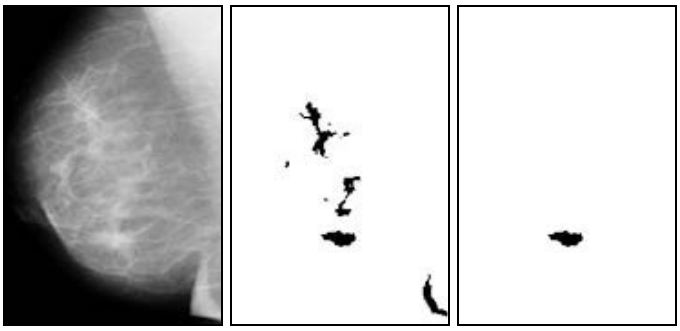
If these two rules are both valid for the same ROI, it means that the euclidean distance of the ROI is bigger than the "*euclidean distance threshold*" and the regularity of the ROI is smaller than the "*regularity threshold*", the ROI is a mass. These operations are repeated for all ROIs and if it is non-mass, it is erased.

3. RESULTS

Our CAD system is applied to 41 mammogram images with abnormal cases circumscribed mass (22 images) and spiculated mass (19 images). Using the regions of interest specification methods 289 ROIs are specified. Most of them are normal ROIs corresponding to artifacts and blood vessel type objects. By the rule based sub-system ROIs are classified as follows: 38 ROIs are true positive (TP), 5 are false negative (FN), 12 are false positive (FP) and 234 are true negative (TN). Figure 5 and 6 give examples on specifying the ROIs and detecting the masses. The experimental results show that the sensitivity of the system reaches 88.37% with 0.292 false positives per image.

a b c

Figure 5. a. The Mammogram With A Circumscribed Mass, b. The Specified ROIs, c. The Detected Circumscribed Mass



a b c

Figure 6. a. The Mammogram with a Spiculated Mass, b. The Specified ROIs, c. The Detected Spiculated Mass

4. CONCLUSION

It is well known that mammogram interpretation is a very difficult task even for experienced radiologists. We proposed a new algorithm for the detection of masses on mammograms. Every suspicious object is labeled and their morphologies are considered in the extraction of masses from the original image. The evaluation of the system was carried out on MIAS. By obtaining high sensitivity with acceptable number of false positives per image, our CAD system has the potential of improving doctors' diagnostic performances.

5. REFERENCES

Baines C. J., McFarlane D. V., and Miller A. B., (1990), "The Role of the Reference Radiologist: Estimates of Interobserver Agreement and Potential Delay in Cancer Detection in the National Screening Study", Investiga Radiology, 25, 971-976.

Bird R. E., (1990), "Professional Quality Assurance for Mammography Screening Programs", Radiology, 175, 587.

Brenner R. J., (1991), "Medicolegal Aspects of Breast Imaging: Variable Standards of Care Relating to Different Types of Practice", AJR, 156, 719-723.

Cheng H. D., Lui Y. M., and Freimanis R. I., (1998), "A Novel Approach to Microcalcification Detection Using Fuzzy Logic Technique", IEEE Trans Med Imaging, 17, 442-450.

Kalman B. L., Reinus W. R., Kwasny S. C., Laine A., and Kotner L., (1997), "Prescreening Entire Mammograms for Masses with Artificial Neural Networks: Preliminary Results", Acad Radiol, 4, 405-414.

Kupinski M. A. and Giger M. L., (1997), "Investigation of Regularized Neural Networks for the Computerized Detection of Mass Lesions in Digital Mammograms", Proc IEEE Eng Medicine & Biology Conference, 1336-1339.

Manohar M. and Ramapriyan H. K., (1989), "Connected Component Labeling of Binary Images on a Mesh Connected Massively Parallel Processor," Computer Vision, Graphics, and Image Processing, 45, 133-149.

Nagel R. H., Nishikawa R. M., and Doi K., (1998), "Analysis of Methods for Reducing False Positives in the Automated Detection of Clustered Microcalcifications in Mammograms", Med Phys, 25, 1502-1506.

Nishikawa R. M., Giger M. L., Doi K., Vyborny C. J., and Schmidt R. A., (1995), "Computer-Aided Detection of Clustered Microcalcifications on Digital Mammograms", Medical and Biological Engineering and Computing, 33, 174-178.

Polakowski W. E., Cournoyer D. A., Rogers S. K., DeSimio M. P., Ruck D. W., Hoffmeister J. W., and Raines R. A., (1997), "Computer-Aided Breast Cancer Detection and Diagnosis of Masses Using Difference of Gaussians and Derivative-Based Feature Saliency", IEEE Trans Med Imaging, 811-819.

Ronse C. and Devijver P. A., (1984), Connected Components in Binary Images: the Detection Problem, Research Studies Press, NY, Wiley.

Stefano L. D. and Bulgarelli A., (1999), "A Simple and Efficient Connected Components Labeling Algorithm", Proceedings of International Conference on Image Analysis and Processing, 322-327.

Suckling J., Parker J., Dance D., Astley S., Hutt I., and Boggis C., (1994), "The Mammographic Images Analysis Society Digital Mammogram Database", Exerpta Medica, 1069, 375–8

Thurfjell E. L., Lernevall K. A., and Taube A. A., (1994), "Benefit of Independent Double Reading in a Population-Based Mammography Screening Program", Radiology, 191, 241-244.

Vyborny C. J. and Giger M. L., (1994), "Computer Vision and Artificial Intelligence in Mammography", AJR, 162, 699-708.

Wallis M., Walsh M., and Lee J., (1991), "A Review of False Negative Mammography in a Symptomatic Population", Clinical Radiology, 44, 13-15.