computing@tanet.edu.te.ua
www.tanet.edu.te.ua/computing

ISSN 1727-6209
International Scientific
Journal of Computing

FAST DETECTION OF MASSES IN MAMMOGRAMS WITH DIFFICULT CASE EXCLUSION

Gábor Takács [1], Béla Pataki [2]

Department of Measurement and Information Systems,
Budapest University of Technology and Economics,
H-1117 Budapest, Magyar tudósok körútja 2.
[1] gtakacs@mit.bme.hu, http://www.mit.bme.hu/~gtakacs
[2] pataki@mit.bme.hu, http://www.mit.bme.hu/~pataki

Abstract: *Breast cancer is one of the most common forms of cancer among women. Currently mammography is the most efficient method for early detection. A simple and fast mammographic mass detection system and two different methods for difficult case exclusion are presented in this paper. The mass detection system uses a modified version of a known algorithm for small masses and a new algorithm for large masses. The first difficult case filtering method is based on tissue density estimation, the second one on mass candidate count. The system was tested with 600 mammographic cases, each containing 4 images. Case-level performance was measured for malignant mass detection first without and then with difficult case exclusion.*

Keywords: *Mammography, Mass Detection, Image Processing, Computer-Aided Diagnosis*

1. INTRODUCTION

Breast cancer is one of the most frequent cancerous diseases among women. Every 12th woman suffers from this disease at least once in her lifetime [1]. Since the cause of the disease is unknown, early detection is very important.

Currently mammography (X-ray examination of the breast) is the most efficient method for early detection. In a mammographic session usually two images are taken of both breasts. Left craniocaudal (LC) is a top view, left mediolateral (LM) is roughly a side view image of the left breast, right craniocaudal (RC) and right mediolateral (RM) are the same views of the right breast.

With regular mammographic screening examinations the mortality of the disease can be significantly decreased. (If breast cancer is detected early, the five-year survival rate exceeds 95 %.)

The evaluation of the images taken at the screening examinations needs a large amount of human resource and money. Therefore computer-aided diagnosis (CAD) for mammography has been an active area of research (e.g. [2], [3]). The main goals of a CAD system are to increase the accuracy of examination by aiming radiologists' attention to suspicious cases and to decrease the cost by filtering out normal cases.

The most important mammographic symptoms of breast cancer can be divided into two main classes:

- Microcalcification: group of small white calcium spots.
- Mass: circumscribed object brighter than its surrounding tissue.

Not all microcalcifications and masses are cancerous, they can also be benign. The two main classes can be divided into subclasses, for example the ACR (American Collage of Radiology) BI-RADS recommendation [4] defines 9 mass and 13 microcalcification subtypes. Combined mass-microcalcification lesions are possible too.

The recognition of these structures is a hard and challenging task that needs intelligence. Mass detection is particularly difficult, because masses show a great diversity in optical density, shape, position, size and characteristics at the edge. Humans and computer algorithms have to deal with a number of difficulties: for example the boundary can be fuzzy or partially missing, irrelevant objects can overlap the mass, some benign findings (e.g. cysts) also appear as masses, normal architectural structures of the breast superimposed on each other can look like real masses.

Figures 1 – 4 show some typical forms of microcalcifications and masses appearing in real mammograms.

Fig. 1 – A typical benign microcalcification

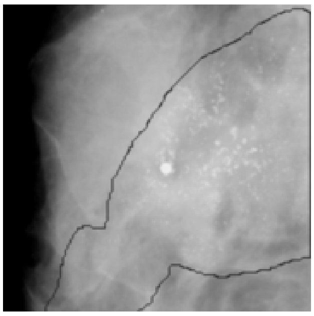

Fig. 2 – A typical malign. microcalcifiction

Fig. 3 – A typical benign mass

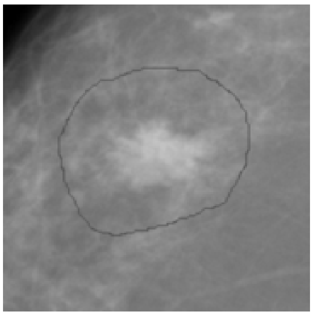

Fig. 4 – A typical malignant mass

This paper presents a simple and fast system for detecting masses in digitalized mammograms. Two size classes were defined and different algorithms are used for small and large mass detection. The precision of the system was improved by applying difficult case exclusion. A tissue density estimation based and a mass candidate count based method were developed and tested in the difficult case filtering experiments.

2. DETECTION OF SMALL MASSES

The size of mammographic masses varies in a wide range (~5 mm to ~50 mm in diameter). An interesting question of automated breast cancer detection is how to handle this size variability. The proposed system defines two size classes and uses different algorithms for small and large mass detection. Most mammographic masses belong to the small class so the small mass detector is the critical part of the system. Large masses are rare and easier to detect.

For the detection of small masses (smaller than 20 mm in diameter) a slightly modified version of the AFUM mass detection algorithm [5] was applied. A short description of our modified algorithm:

At each pixel position (x, y) the minimal intensity at distance r_1 from location (x, y) is computed (m_1), then the fraction of pixels at distance r_2 from (x, y) that have lower intensity value than m_1 is measured. This fraction under the minimum (FUM) calculation is done over many scales using a range of r_1 and r_2 values and the average of those calculations yields the average FUM (AFUM) value.

This AFUM algorithm variant slightly differs from the original one, because in the original algorithm the minimal intensity at distance *less than or equal* r_1 from (x, y) is compared to intensity values at distance r_2 from (x, y).

In real mammograms some masses contain small dark dots inside. The original AFUM algorithm prohibits this case while the proposed variant tolerates it to some degree.

If $r_1 = R_{min}, R_{min} + 1, R_{min} + 2, ..., R_{max}$ and $r_2 = r_1 + D$ then the AFUM value calculation can be written as:

$$\frac{1}{R_{max} - R_{min} + 1} \cdot \sum_{r_1 = R_{min}}^{R_{max}} FUM(r_1, r_1 + D) \quad (1)$$

Values R_{min} R_{max} and D are fixed a priori choices based on the problem definition. An advantage of this algorithm is simplicity and therefore computational efficiency. A nice property of the algorithm is invariance to any monotonically increasing intensity transformation of the input image, since only logical operators (min. finding and comparison) are applied to the intensity values.

A fast mass detector can be obtained by running the AFUM algorithm for each non-background pixel of a mammogram. The filtered image is thresholded and continuous regions are identified by a region-filling algorithm. Regions with a too high perimeter-area ratio are excluded from further examinations. The location of the maximal AFUM value is computed for each region, and an “energy” value is assigned for each maximum location based on the AFUM value of that position and its neighboring pixels. A structure is accepted as a mass if this energy is higher than a limit. Finally the locations of the highest N energy maxima are returned. Figure 5 illustrates the steps of small mass detection:

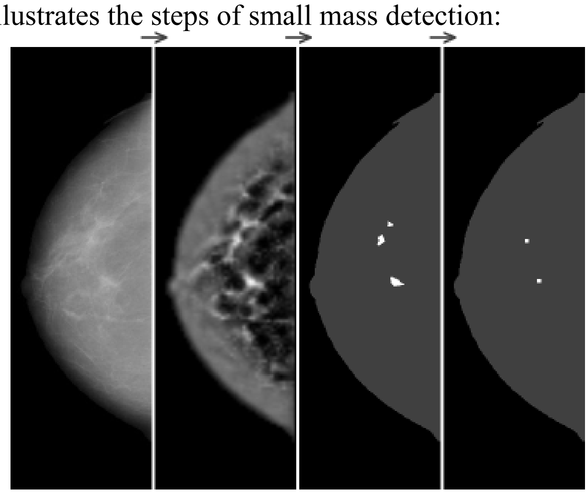

Fig. 5 – The steps of small mass detection

3. DETECTION OF LARGE MASSES

The a priori parameters of the AFUM algorithm ($R_{\min}$, $R_{\max}$ and D) could not be set to deal with arbitrary mass size. At the resolution of 400 microns $R_{\min} = 0$, $R_{\max} = 6$ and $D = 12$ proved to be a good choice but worked well only for masses smaller than 20 mm in diameter.

For the detection of larger masses the following simple and fast algorithm was developed: At a given pixel position 8 lines are started from the center (vertically, horizontally and diagonally) and a mass boundary point is estimated for each direction based on some simple intensity change constraints. Then a "conspicuousness" value can be obtained from the line lengths (l_i), average intensity along the lines (*Brightness*) and average contrast at the end of the lines (*Contrast*).

$$Conspicuousness = \min_i l_i \cdot Brightness \cdot Regularity \cdot Contrast \quad (2)$$

$$Regularity = \min_i l_i \Big/ \max_i l_i \quad (3)$$

Since the detection of large masses is easier than that of the small ones, the large mass detector returns only the location of the highest conspicuousness value when processing a whole mammogram. Obviously a lower threshold for the conspicuousness can be applied to control the sensitivity of the large mass detector.

4. DIFFICULT CASE EXCLUSION BASED ON TISSUE DENSITY ESTIMATION

Breast tissue density is an important feature of a mammographic case for human experts. There are standards for classifying mammogaphic cases into tissue density classes. For example BI-RADS recommends four breast tissue density types numbered from 1 to 4. Class 1 means fatty tissue (dark, homogenous background), class 4 means dense tissue (that can mask interesting structures).

Human experts say that mammographic mass detection (and microcalcification detection) is a much more difficult task on dense cases than on fatty ones. For cases with an extremely dense tissue mammogaphy is not applicable at all.

Figures 6 – 7 show a fatty and a dense case, each containing a malignant mass.

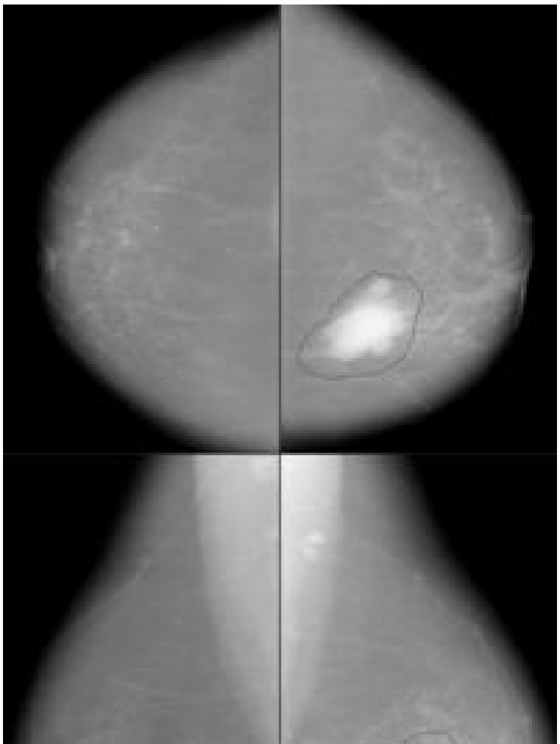

Fig. 6 – A fatty case containing a malignant mass

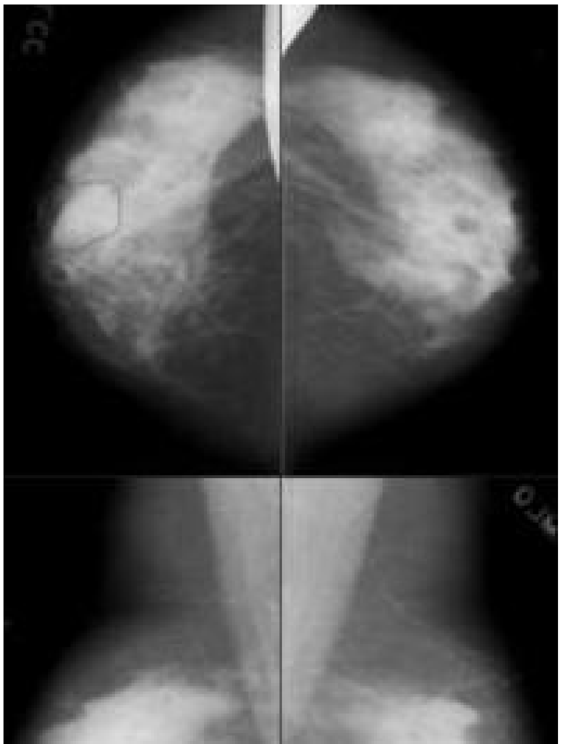

Fig. 7 – A dense case containing a malignant mass

According to this observation a density estimation algorithm was developed to improve the accuracy of the mass detector by filtering out difficult cases. The density estimator measures the intensity mean and variance of the four images (LC, LM, RC and RM) of the input case. Then the estimated density value is computed with the following heuristic formulae:

$$EstimatedDensity = (2 \cdot NSumSd + NSumMean)/3 + 2.5 \quad (4)$$

$$NSumSd = (\sqrt{Var_{LC}} + \sqrt{Var_{LM}} + \sqrt{Var_{RC}} + \sqrt{Var_{RM}} - 166.8)/24.5 \quad (5)$$

$$NSumMean = (Mean_{LC} + Mean_{LM} + Mean_{RC} + Mean_{RM} - 492.6)/73.0 \quad (6)$$

NSumSd denotes statistically normalized sum of the standard deviations, *NSumMean* denotes statistically normalized sum of the means. Constants are set to make the estimated density value directly comparable with the BI-RADS density value.

Experiments showed that our estimated density value (rounded to the nearest integer) usually does not equal exactly with the BI-RADS density value, but the two parameters are in positive correlation with each other (Figure 7). The measurement was based on 157 mammographic cases of the DDSM database [6].

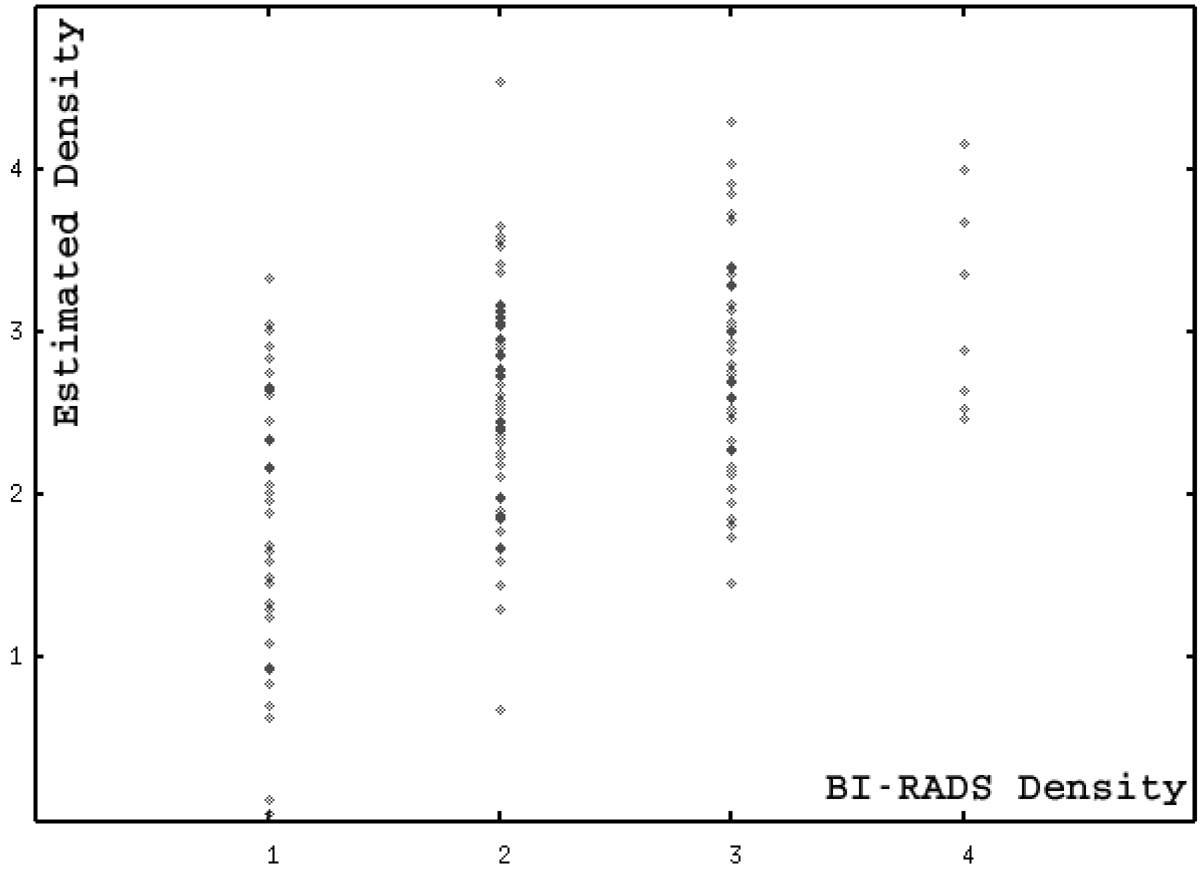

Fig. 8 – The correlation between the BI-RADS and the estimated density value

5. DIFFICULT CASE EXCLUSION BASED ON MASS CANDIDATE COUNT

The first difficult case filtering method was developed according to a human experts' observation: dense cases are more difficult than fatty ones. Another possible way is to get the case difficulty information from the mass detection system itself.

For example the number of energy maxima after the perimeter-area ratio filtering step in the small mass detector (mass candidate count) seems to be a good parameter. If the number of energy maxima is high then real masses have lesser chance to be among the *N* highest maxima, so the film is difficult for the system. The case-level mass candidate count value can be obtained by the summation of film-level mass candidate counts.

The average mass candidate count of the BI-RADS density classes can be seen in Table 1 (based on 157 cases of the DDSM database [6]).

Table 1. Average mass candidate count of the BI-RADS density classes

BI-RADS Density	Avg. Mass Candidate Count
1	71.1
2	93.1
3	112.9
4	85.5

6. RESULTS

The mass detection system was tested with 600 cases (600 · 4 = 2400 images) of the DDSM database [6]. 424 cases contained malignant masses, 10 cases contained benign (but no malignant) masses, 166 cases contained no masses. A malignant case was counted as recognized if one of the pixel positions returned by the mass detector was inside the radiologist-drawn boundary of a malignant mass on any image of the case. An output pixel position was counted as a false mark (FM) if it was not inside the radiologist-drawn boundary of any (malignant or benign) mass. Table 2 shows the performance parameters of the system without difficult case exclusion:

Table 2. Results of the mass detector

MCRR	# FMs / Image
90.3 %	5.45

The malignant case recognition rate (MCRR) is nearly 90 % that is acceptable in itself, but it comes together with a high number of false marks per image. At this false mark level the system can be used for increasing the accuracy of the examination but cannot be used for decreasing the cost by filtering out normal cases.

Results with density estimation based difficult case exclusion are summarized in Table 3:

Table 3. Results with density estimation based difficult case exclusion

Rejection Rate	MCRR	# FMs / Image
18.7 %	90.4 %	5.46
33.3 %	90.6 %	5.46
53.8 %	90.8 %	5.43
69.3 %	90.6 %	5.45
82.3 %	92.9 %	5.45

The false mark level remained nearly the same. The malignant case recognition increases with the rejection rate but a significant improvement can be observed only at a very high rejection rate.

Table 4 shows the results of mass candidate count based difficult case exclusion:

Table 4. Results with mass candidate count based difficult case exclusion

Rejection Rate	MCRR	# FMs / Image
4.2 %	90.8 %	5.43
10.0 %	91.3 %	5.42
18.5 %	91.8 %	5.40
35.3 %	91.8 %	5.35
55.7 %	93.9 %	5.36

The false mark level is slightly reduced. The malignant case recognition rate increases monotonically with the rejection rate. The mass candidate count based filtering provides a better malignant case recognition than the density estimation based one at the same rejection rate. The system performs better than its previous version [7] from which the large the mass detector part was missing (Table 5).

Table 5. Some results of the previous version of the system

Rejection Rate	MCRR	# FMs / Image
0 %	89.6 %	5.6
18.7 % (density est. based filtering)	88.7 %	5.5
4.3 % (mass count based filtering)	90.3 %	5.5

7. CONCLUSION

A mammographic mass detection system and two algorithms for difficult case exclusion were presented in this paper. The system uses different methods for small and large mass detection. The first difficult case filtering algorithm is based on tissue density estimation, the second one on mass candidate count.

The mass detector is fast enough to process each pixel of mammogram in reasonable time at the resolution of 400 microns. It finds almost all malignant masses, but also returns a high number of false marks.

The malignant case recognition rate of the mass detector was improved by filtering out difficult cases. The mass count based difficult case filtering proved to be better than the tissue density estimation based method.

Further improvements could be achieved by implementing a complex post-processing for the suspicious spots returned by the fast mass detector, and by comparing the two views of the same breast.

8. REFERENCES

[1] R. Highnam, M. Brady (Editors). *Mammographic Image Analysis*, Kluwer Academic Publishers, 1999.

[2] S. Lee, P. Chung, C. Chang, C. Lo, T. Lee, G. Hsu, C. Yang. Classification of Clustered Microcalcifications Using a Shape Cognitron Neural Network. *Neural Networks* 16 (2003), pp. 121-132.

[3] M. Altrichter, Z. Ludányi, G. Horváth. Joint Analysis of Multiple Mammographic Views in CAD Systems for Breast Cancer Detection. *Proceedings of the 14th Scandinavian Conference on Image Analysis* (SCIA2005), Joensuu, Finland, 2005, pp. 760-769.

[4] American College of Radiology. *Illustrated Breast Imaging Reporting and Data System (BI-RADS)* (3rd ed). Reston, VA: American College of Radiology, 1998.

[5] M. D. Heath, K. W. Bowyer. Mass Detection by Relative Image Intensity. *Proceedings of the Fifth International Workshop on Digital Mammography* (IWDM-2000), Toronto, Canada, 2000, pp. 219-255.

[6] M. D. Heath, K. W. Bowyer, D. Kopans. Current status of the Digital Database for Screening Mammography. *Digital Mammography*, Kluwer Academic Publishers, 1998, pp. 457-460.

[7] G. Takács, B. Pataki. Fast Detection of Mammographic Masses with Difficult Case Exclusion. *Proceedings of the Third IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications* (IDAACS'2005), Sofia, Bulgaria, 2005, pp. 424-428.

Gábor Takács, *Born in Győr, Hungary, received M.Sc. degree in technical informatics from the Budapest University of Technology and Economics in 2004. He is currently a Ph.D. student at the Department of Measurement and Information Systems, Budapest University of Technology and Economics. His research interests include pattern recognition, machine learning and medical image processing.*

Béla Pataki, *Born in Budapest, Hungary, received M.Sc., Technical Doctorate and Ph.D. degrees in electrical engineering from the Budapest University of Technology and Economics, in 1978, 1994 and 1997 respectively. He is an associate professor at the Department of Measurement and Information Systems, Budapest University of Technology and Economics. His current research topics include intelligent signal analysis and processing, medical image processing, and decision support systems.*