Automatic Detection of Microcalcification Clusters Using Fuzzy Logic and Wavelet Transformation

H.D. Cheng, X.Y. Liu, X. J. Shi, and Liming Hu

Department of Computer Science
Utah State University
Logan, UT 84322-4205, USA

Abstract

In 2004, 215,990 new cases of breast cancer are expected to occur among women in the United States. Early detection of breast cancer is essential in breast cancer control. In this paper, a novel approach to microcalcification detection, based on fuzzy logic and wavelet transformation is presented. Then, the free-response operating characteristic curve (FROC) is used to evaluate the performance. Compared with existing algorithms using the same set of mammograms, the proposed algorithm achieves better results, producing both low FPs and low FNs rates.

***Keywords*:** Fuzzy logic, Wavelet transformation, Maximum entropy principle, Microcalcification.

1. Introduction

Breast cancer continues to be a significant public health problem in the world. It ranks second among cancer deaths in women (after lung cancer) [1]. Primary prevention seems impossible since the causes for this disease still remain unknown. Thus, early detection for breast cancer is critical.

Mammograms have been shown to be one of the most reliable methods for early detection of breast carcinomas. However, large volumes of mammograms need to be analyzed and compared in order to make consistent diagnosis, this is very labor intensive and cost ineffective. Further, human judgment depends on training, experience, and subjective criteria. Even the well-trained experts have an inter-observe variation rate of 65-75% [2]. Computer-aided mammography screening has received great attention because it is fast, consistent, and capable of detecting cancer early [3]. The computer aided diagnosis (CAD) methods provide objective tools to help radiologist in interpreting mammograms and detecting microcalcifications. The detection sensitivity with CAD improves to 90% [3]. Although computer-aided mammography has been studied more than two decades, and for all of its usefulness, it still has limitations: it misses some cancers, and sometimes leads to follow up of findings that are FPs.

Microcalcifications are often difficult to detect due to the following factors [4]:

1. Microcalcifications are very small. The sizes of microcalcifications are in the range of 0.1 – 1.0 mm. Some isolated ones smaller than 0.1 mm cannot be distinguished in the film-screen mammography from the high-frequency noise.
2. Microcalcifications have various sizes, shapes, and distributions, and simple template matching is impossible.
3. Microcalcifications may be low contrast so that the intensity difference between suspicious areas and their surrounding tissues can be quite slim.
4. Microcalcifications may be closely connected to surrounding tissues, and simple segmentation algorithms cannot work well.
5. In some dense tissues, and/or skin thickening, suspicious areas are almost invisible, and they may create a high false-positive (FP) rate. Most of FPs are due to film emulsion error, digitization artifacts, or anatomical structures, such as fibrous strands, breast borders, or hypertrophied lobules.

To deal with the above problems, it is very important to suppress the noise, to enhance the contrast within the region of interest (ROI), and to extract and select the features of microcalcifications effectively.

The most commonly used method, the free-response receiver operating characteristic (FROC) is employed to evaluate the performance [5].

2. Proposed Approach

The proposed approach for microcalcification detection using fuzzy logic and wavelet transformation mainly includes 5 steps: (1) image fuzzification; (2) wavelet transformation; (3) microcalcification candidate segmentation; (4) denoise by spatial influence function; (5) MCCs detection.

2.1 Fuzzy Domain Conversion

Fuzzy logic allows processing data with the membership $\in[0, 1]$ rather than $\in\{0, 1\}$ [6].

2.1.1 S-function

When fuzzifying the mammograms, the intensities of mammograms are transformed to an interval of [0, 1] using a standard S-function. The standard S-function is defined as [6]:

$$\mu_{bright}(g) = S(g;a,b,c) \qquad (1)$$

$$S(g;a,b,c) = \begin{cases} 0 & \\ \dfrac{(g-a)^2}{(b-a)(c-a)} & g \le a \\ 1 - \dfrac{(g-c)^2}{(c-b)(c-a)} & b \le g \le c \\ 1 & g \ge c \end{cases}$$

where g represents the gray level, a, b, and c are the parameters that determine the shape of the S-function. $0 \le a < b < c \le 255$.

2.1.2 Maximum Entropy Principle

According to information theory, the system with the maximum entropy will contain the most information. The maximum entropy principle can be used to determine the S-function. For a histogram, find the values of a, b, and c of the S-function that gives the maximum entropy based on the equations below:

$$P(A) = -P(A)\log_2(P(A)) - (1-P(A))\log_2(1-P(A)) \qquad (2)$$

$$P(A) = \sum_{g=0}^{L-1} \mu_{bright}(g) H(g) \qquad (3)$$

$H(g)$ represents the probability of gray level g in the image.

2.2 Application of DWT to Mammograms

Wavelets are simultaneously localized in space and frequency, and have been used in image analysis [7].

The discrete wavelet transform (DWT) is defined for sequences with the length of some power of two, and different ways of extending samples of other sizes are needed. Methods for extending the signal include zero-padding, smooth padding, periodic extension, and symmetric-padding (symmetric replication of boundary values). Here, we use symmetric-padding.

Several wavelets were tested. The SYM4 wavelet yields good result in our experiments. Six-level redundant wavelet de-composition of the fuzzified mammogram ROIs is performed. Experiments show that the first level of the detail information contains most of the noise. The second and third levels represent most of the microcalcification information. The micro-calcification information decreases in the fourth and fifth level. Basically, the sixth level contains only the mammogram background information.

2.3 Microcalcification Candidates Segmentation

The microcalcification candidates need to be segmented. A popular tool used in image segmentation is thresholding. There are a variety of thresholding techniques, including both global and local thresholding. A global thresholding technique is one that thresholds the entire image with a single threshold value, whereas a local thresholding technique is one that partitions a given image into sub-images and determines a threshold for each of sub-image. Experiment shows global thresholding either introduces too many FPs or causes too many FNs. Usually, the local thresholding will perform better, but it still cannot avoid FPs. It would be ideal to design a thresholding method using both local and global thresholding scheme. In our approach, the local thresholding is based on the local means, local variances and local means of variances of two adaptive filter windows with sizes 7×7 and 21×21, respectively.

2.3.1 Local Thresholding

The original image is divided into 7×7 sub-images. A threshold is calculated for each sub-image. Due to the complexity of the mammogram and the low contrast in some suspicious areas, one filter is not enough. We design another filter window with size 21×21 to assist the thresholding scheme.

The local variance σ_{pq} (7×7) describes the characteristics of microcalcification in some degree. We can define the following function:

$$p_{pq} = \begin{cases} 1 & if\ \sigma_{pq} > a_{pq} T_I \\ & \quad and\ g_{pq} > \mu_{pq} + a_{pq}\sigma_{pq} \\ 0 & otherwise \end{cases} \qquad (4)$$

p_{pq} represents the pixel value at the position (p, q), value 1 means this pixel is kept after segmentation whereas value 0 means this pixel is removed after segmentation. T_I is a thresholding value for image I, and defined in a 7×7 window as below:

$$T_I = f(\mu_{pq}, \sigma_{pq}) \qquad (5)$$

where μ_{pq} and σ_{pq} are the average variance value and the variance value, respectively.

Eq. 5 is nonlinear, complex, and difficult to describe explicitly. A multi-layer, feed-forward, error back-propagation neural network (BPN) is employed to find the solution for Eq. 5. Thirty-eight ROIs from 50 are selected as training set; the other 12 ROIs constitute the testing set in our algorithm.

2.3.2 Global Thresholding

Based on our experiments, the percentage q_I of the number of the pixels which belong to the MCCs

compared with that of the ROI and $q_I \in [0.0006, 0.008]$. The percentage varies according to the image size. A bigger image has a smaller percentage q_I.

Based on the above discussion, the threshold should be defined as the highest gray level which maps at least *(100-p)%* of the pixels into the objects in the thresholded image. After processing of the local thresholding, some FPs is introduced. We apply the global thresholding method to reduce FPs.

$$p_I = 0.008 - 0.0074 * (S - S_{\min}) / (S_{\max} - S_{\min}) \quad (6)$$

where, p_I is the thresholding value of image *I*, *S* is the image size, S_{min}=128×128 and S_{max}=512×1024. Coefficients 0.008 and 0.0074 are defined based on the experiments.

2.4 De-noise Using Spatial Influence Function

After thresholding, microcalcifications will be segmented from mammograms. However, noises and isolated microcalcifications are also introduced. Noises usually are small spots with one or two pixels. A simple filtering algorithm that removes small spots will not work since true microcalcifications can also removed. To complicate things even further, the shapes of isolated spots are similar to that of the microcalcifications. We must look at additional characteristics. The difference between noises and isolated microcalcifications are their spatial distribution: true microcalcifications belonging to the same cluster locate within a ground truth circle. On the other hand, noised and isolated spots locate randomly and quite far, i.e., there are no spatial relationships between them.

We use a spatial influence function to describe the mathematical model for the probability of a microcalcification to belong to a cluster. The distance *dis* between two pixels is defined as:

$$dis = \sqrt{(m - i)^2 + (n - j)^2} \quad (7)$$

where *(m, n)* and *(i, j)*are coordinates. The spatial influence value *inf* of a microcalcification is defined as:

$$inf = \begin{cases} 1 - 0.01 * dis & if \quad dis <= 100 \\ 0 & Otherwise \end{cases} \quad (8)$$

A simple thersholding method is applied in this process.

$$p_{pq} = \begin{cases} 1 & if \quad inf > T_{inf} \\ 0 & otherwise \end{cases} \quad (9)$$

Ideally, the thresholding value of the de-noising method by the influence function can be decided by the smallest value of clusters. In our experiment, we take ten cutouts of the relatively small MCCs from the ROIs to calculate the smallest value for the *inf* in the Eq. (8) as the thresholding value T_{inf} in the Eq. (9), and T_{inf} =9 here.

2.5 Microcalcification Clusters Detection

The clusters which have three or more spots within a region of $1cm^2$ are considered as clinically suspicious [8]. Automatically, the cluster search algorithm first picks up a microcalcification and then uses a recursive searching to find all the neighbor microcalcifications.

We use a rectangle mask with the size as *10mm×10mm* to identify the cluster. If no more than two spots can be found as its neighbors, this microcalcification will be removed. Otherwise, it will be combined into one cluster and labeled in the image.

3. Experiment Results and Discussion

The mammogram images were provided by (http://figment.csee.usf.edu/) [4]. The database contains 40 digitized mammogram images composed of both oblique and craniocaudal views from 21 patients. Each mammogram has one or more microcalcification clusters marked by radiologists. The total number of clusters in the database is 105. The images are digitized from films at a size of 2048× 2048 by a 12 bit CCD camera (Eikonix 1412), using a sampling aperture of *0.05mm* in diameter and *0.1mm* sampling distance. The database also has text files which record the coordinates of the centers and the diameters of the truth clusters, and a table for rescaling the images to the ones with gray levels of 256 (8 bits).

Figure 1 shows the complete process of the algorithm as applied to the mammogram. This particular mammogram is regarded as a difficult one because of its dense tissue.

The performance of the proposed algorithm is evaluated by a free-response receiver operating characteristic (FROC) [5] which is the commonly accepted method. The FROC diagram describes the fraction of TP clusters (TP fraction) as a function of the number of FP clusters per image. The thresholding value T_1 introduced in Section 2.3.1 can be used to evaluate the performance of the algorithm. The corresponding FROC curve is shown in Figure 2.

Based on the same DDSM database, 4 methods were implemented: 1. the proposed algorithm; 2. the algorithm which employs the scale-space and Laplacian-of-Gaussian filter techniques based on the fuzzy domain; 3. the algorithm which algorithm employs adaptive noise equalization (ANE) method, and noise characteristics are estimated from the images; 4. the algorithm which

employs scale space and Laplacian to detect the microcalcifications.

The proposed algorithm achieves a better result with the consistently lowest number of FPs and FNs (see Table 1).

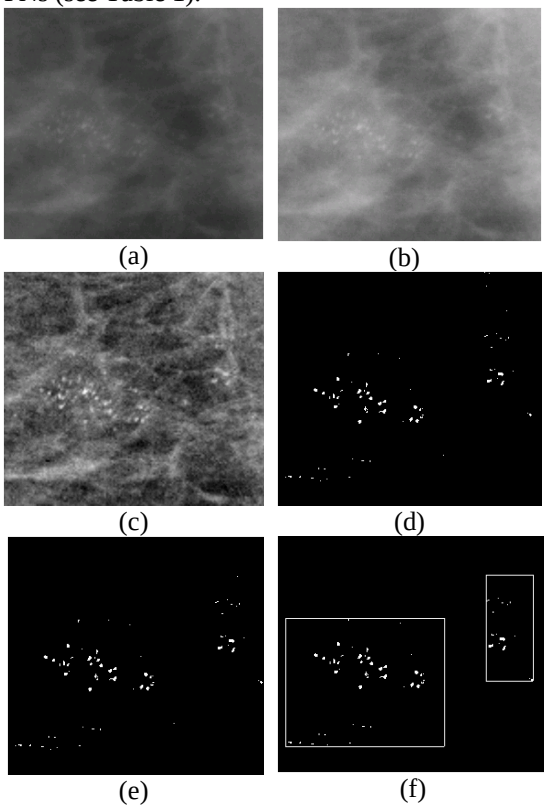

Figure 1. Complete processing process:
(a) Original image. (b) Fuzzified image.
(c) Wavelet enhancement. (d) Segmentation.
(e) De-noise by spatial influence function.
(f) Result of cluster (MCCs detection).

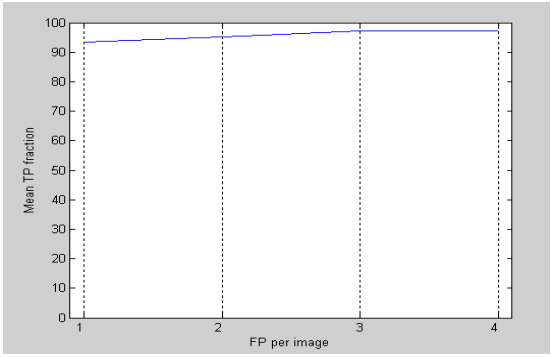

Figure 2. FROC Curve.

Table 1. Experimental data comparison of 105 ROIs.

Method	1	2	3	4
FPs	9	14	39	42
FNs	9	10	12	17

4. Conclusions

In this paper, a novel approach to micro-calcification detection based on fuzzy logic and wavelet transformation is presented. With the combination of fuzzy logic and wavelet transformation, image contrast is greatly improved and microcalcifications are better detected. We segment the image with both a global thresholding method and a local thresholding method. The local thresholding method is based on the local variances and local means of two adaptive filter windows with different sizes. The new thresholding method on the mammogram greatly improves the detection accuracy. A neural network is used in the thresholding value training. Also, in the algorithm, a new de-noising method based on spatial influence function is applied to remove isolating spots. The algorithm achieves better results. Using the same image database from DDSM, our algorithm consistently outperformed the existing algorithms with lower FPs and FNs rates.

Reference

[1] Jemal, A., Tiwari, R., Murray, T., Ghafoor, A., Samuels, A., Ward, E. Feuer, and Thun, M. Cancer statistics, 2004. *CA: A Cancer J. for Clinicians,* 54, 1 (2004), 8-29.

[2] Skaane, P., Engedal, K., and Skjennald, A. Interobserver Variation in the Interpretation of Breast Imaging. *Acta Radiologica*, 38 (1997), 497-502.

[3] Woods, K., Doss, C., Bowyer, K., Solka, J., Priebe, C., and Kegelmeyer, W. Comparative Evaluation of Pattern Recognition Techniques for Detection of Microcalcifications in Mammography. *Intl. J. Pattern Recognition Artificial Intelligence*, 7 (1993), 1417–1436.

[4] Cheng, H.D., Cai, X., Chen, X., Hu, L., and Lou, X. Computer-aided Detection and Classification of Microcalcifications in Mammograms: A Survey. *Pattern Recognition*, 36 (2003), 2967-2991.

[5] Metz, C. ROC Methodology in Radiologic Imaging. *Invest. Radiol*., 21, 9 (1986), 720–733.

[6] Pedrycz, W. and Gomide, F. *Introduction to Fuzzy Sets – Analysis and Design*. MIT Press, 1998.

[7] Pfisterer, R. and Aghdasi, F. Hexagonal Wavelets for the Detection of Masses in Digitised Mammograms. In *Proceedings of SPIE - The International Society for Optical Engineering*, 1999.

[8] Netsch, T. and Peitgen, H. Scale Space Signatures for the Detection of Clustered Microcalcifications in Digital Mammograms. *IEEE Trans. Med. Imag*., 18, 9 (1999), 774–786.