

Development of an Automated Method for Detecting Mammographic Masses With a Partial Loss of Region

Yuji Hatanaka*, Takeshi Hara, Hiroshi Fujita, Satoshi Kasai, Tokiko Endo, and Takuji Iwase

Abstract—Recently, we have been developing several automated algorithms for detecting masses on mammograms. For our algorithm, we devised an adaptive thresholding technique for detecting masses, but our system failed to detect masses with a partial loss of region that were located on the edge of the film. This is a common issue in all of the algorithms developed so far by other groups. In order to deal with this problem, we propose a new method in the present study. The partial loss masses are identified by their similarity to a sector-form model in the template matching process. To calculate the similarity, four features are applied: 1) average pixel value; 2) standard deviation of pixel values; 3) standard correlation coefficient defined by the sector-form model; and 4) concentration feature determined from the density gradient. After employing the new method to 335 digitized mammograms, the detection sensitivity for the partial loss masses jumped from 70% to 90% when the number of false positives was kept constant (0.2/image). Moreover, a combination of the existing method and the new method improved the true-positive rate up to 97%. Such results indicate that the new technique may improve the performance of our computer-aided diagnosis system for mammographic masses effectively.

Index Terms—Breast cancer, computer-aided diagnosis, mammography, mass detection, template matching.

I. INTRODUCTION

RECENTLY, in Japan, incidences of breast cancer have been increasing rapidly. It was reported that breast cancer has the highest mortality rate of any women's cancer. Mammography has been introduced as an effective way for detecting breast cancer in its early stages. However, there are few physicians experienced in breast cancer diagnosis. Therefore, it is necessary to develop a computer-aided diagnosis (CAD) system to assist physicians in interpreting mammograms more efficiently.

An automatic mass detection algorithm based on an adaptive thresholding technique has been developed in our group. In our initial study, a true-positive (TP) rate of 97% was achieved

Manuscript received June 1, 2001; revised October 25, 2001. This work was supported in part by the Ministry of Health, Labour, and Welfare under a Grant-In-Aid for Cancer Research and in part by the Ministry of Education, Culture, Sports, Science and Technology under a Grant-In-Aid for Scientific Research. *Asterisk indicates corresponding author.*

*Y. Hatanaka is with the Department of Information Science, Faculty of Engineering, Gifu University, Yanagido 1-1, Gifu-shi 501-1193, Japan (e-mail: hatanaka@fjt.info.gifu-u.ac.jp).

T. Hara and H. Fujita are with the Department of Information Science, Faculty of Engineering, Gifu University, Gifu-shi 501-1193, Japan.

S. Kasai is with MI System Group, Medical & Graphic Company, Konica Corporation, Hino-shi 191-8511, Japan.

T. Endo is with the Department of Radiology, National Hospital of Nagoya, Naka-ku, Nagoya-shi 460-0001, Japan.

T. Iwase is with the Department of Breast Surgery, Aichi Cancer Center Hospital, Nagoya-shi 464-8681, Japan.

Publisher Item Identifier S 0278-0062(01)11137-7.

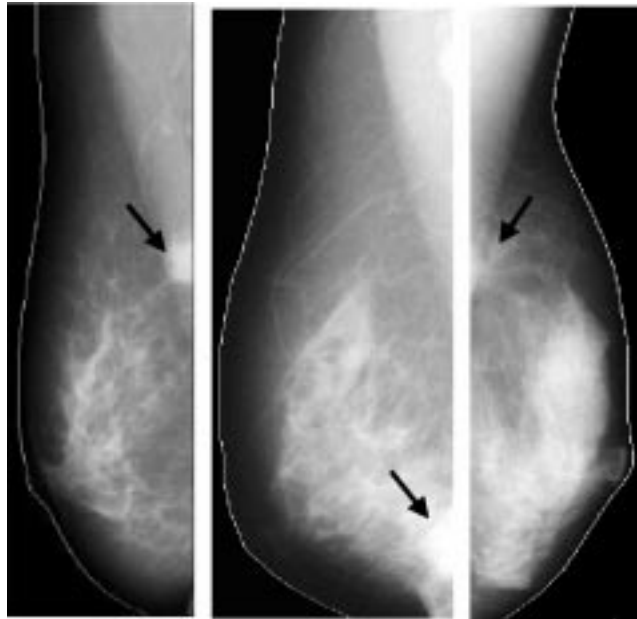


Fig. 1. Examples of masses with a partial loss of region, because of their location on the edge of the film. Three malignant masses are shown with arrows on each mammogram.

with an average number of 3.5 FPs/image [8]–[10]. Due to the high number of FPs detected by our system, we have developed some techniques to eliminate them. They include: using a second-order statistics method [11], applying 11 features for suspicious candidates [12] and comparing right and left mammograms [13]; using a prewitt filter method [14]. Using the above methods, the TP rate dropped slightly to 91%, but the number of FPs decreased to 0.82/image. Testing our CAD scheme for masses in over 6000 mammograms, it was found that the detection sensitivity became lower as the number of more difficult cases increased. Among these difficult cases, we found that in most cases the masses had a partial loss of region. In Fig. 1, three examples of partial loss masses are shown. It is difficult to detect such masses using conventional algorithms because of their lower circularity and irregularity in shape. Reports for solving this issue have never been published. In order to detect these kinds of masses efficiently, in this study, we propose a new method based on a sector-form model. Usually, breast cancer appears as masses and clustered microcalcifications in mammograms. Many studies on automated detection schemes of mass shadows have been reported recently [1]–[7]. For example, Lai *et al.* presented a method for detecting circumscribed masses [2]. In their technique, the mammograms were first enhanced by a selective median filter,

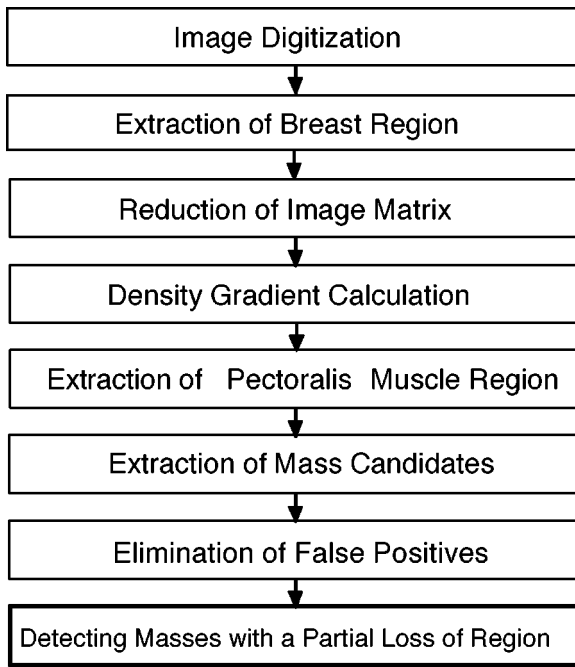


Fig. 2. Flowchart for detecting masses on mammograms. The newly developed method for mass detection is indicated by the boldfaced box.

and then template matching was used to identify suspicious areas. Li *et al.* used a technique based on adaptive threshold and segmentation using a modified Markov random field for mass detection [3]. The detected masses were classified using measures of shape complexity and the mean gradient of the region boundary with a fuzzy binary decision tree. There are some approaches based on the analysis of symmetry between both sides of mammograms. Yin *et al.* developed a nonlinear bilateral subtraction technique to identify asymmetries between the right and left breast images [4]. Different from Yin's approach, Mendez *et al.* have characterized the symmetries using only one thresholding instead of the more complex methods of linking multiple subtracted images [5]. In Japan, Matsumoto *et al.* have developed a technique using an iris filter which is based on the distribution of the concentration vector [6]. Hashimoto *et al.* proposed a subtraction technique to eliminate false-positives (FPs) with an iris filter [7].

II. METHOD

The flowchart of our overall detection scheme including the new method is shown in Fig. 2. It consists of eight stages: 1) image digitization (0.05-mm sampling pitch and 12-bit density resolution); 2) extraction of the breast region [15]; 3) reduction of the image matrix (0.05–0.4 mm sampling pitch); 4) density gradient calculation; 5) extraction of the pectoralis muscle region [16]; 6) extraction of mass candidates; 7) elimination of FP candidates [11]–[14]; and (8) detecting masses with a partial loss of region as a new stage. In this final stage, the search area is first determined, and then some features are calculated. Next, masses with a partial loss of region are detected using a template matching based on a sector-form model. Finally, the FPs are eliminated. More details are given below.

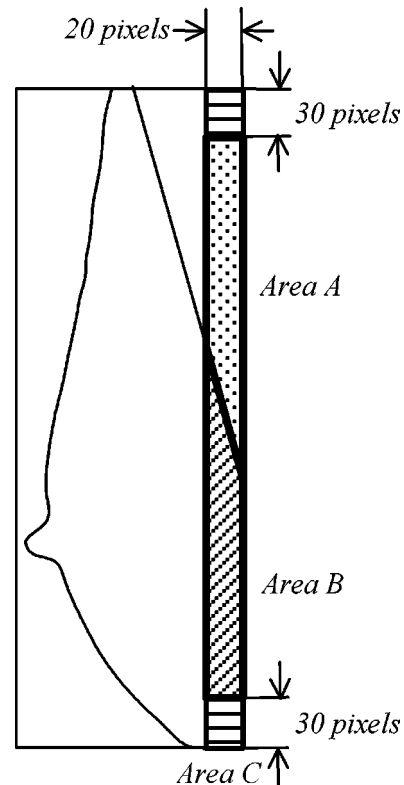


Fig. 3. Determining the search area. A bold line shows the search area. Area A and Area B are located inside and outside the pectoralis muscle, respectively. Area C was defined as outside the search area.

A. Determining the Search Area

Since the masses to be detected by the newly proposed algorithm are mainly located around the breast wall, the search area was determined to be the area within the bold lines shown in Fig. 3. The search area was divided into two parts, Area A and Area B, which are located inside and outside the pectoralis muscle regions, respectively. In this study, however, we chose to ignore Area C because some misdetections occurred in this area. While Fig. 3 shows an instance of a mammogram taken in the mediolateral oblique (MLO) direction, the new algorithm also works on mammograms taken in the cranio-caudal (CC) direction. Since the pectoralis muscle does not appear in CC-type mammograms, the search area in this case consists only of Area B, as shown in Fig. 3.

B. Calculating the Features

In the new method, we applied the following four features: 1) average pixel value; 2) standard deviation of pixel values; 3) correlation coefficient defined by the sector-form model; and 4) concentration feature determined from the density gradient. While 2) was also used as a feature in the existing method, features 1), 3), and 4) are used only in the new method.

To calculate the feature values of 1), 2), and 4), the masks of A and B shown in Fig. 4 were used. These masks were adopted based on a sector model with a diameter of 41 pixels drawn in a 120° arc. When we experimented on sector models whose diameters ranged from 31–51 pixels and arcs of 60° – 180° , we found the above values worked the best. Considering that the change

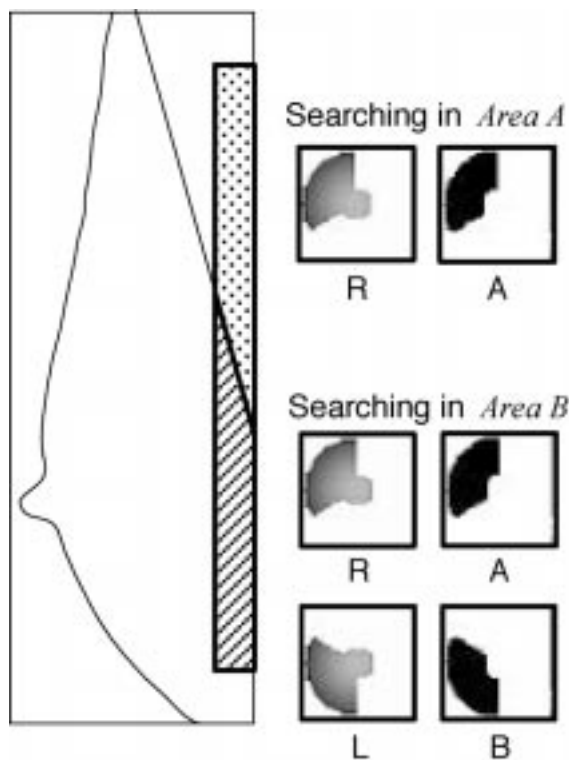


Fig. 4. A and B are mask images used to calculate three feature values. The images of R and L are used as reference patterns for template matching. For Area B, the similarity was calculated using both patterns, A and R, B, and L. For Area A, the similarity was calculated using only the patterns A and R.

of density in the central part of the sector model is not significant and that its value has no effect on the values of 1), 2), and 4), we defined the masks without the seven-pixel-diameter circle in the center of the sector model. The candidate was regarded to be at the center of the sector model.

1) *Average and Standard Deviation of Pixel Values:* The average and the standard deviation of pixel values in both of the back masks (A and B) in Fig. 4 were calculated. It is well known that masses always appear as low-photographic-density areas with a lower standard deviation of pixel values than those of the mammary gland and blood vessels. Therefore, we may disregard those candidates with either higher pixel values or higher standard deviations when calculating the values of 3) and 4). The threshold value of the average pixel value was set experimentally to 1500, and the threshold value of the standard deviation was set experimentally to 650.

2) *Correlation Coefficient:* The correlation coefficient was defined using the sector model. Assuming that the density of a mass has a Gaussian distribution, we used such a distribution to mimic a real mass in our study. The mass model was determined from the Gaussian distribution by setting an area with a diameter of 41 pixels and central angle of 120°. Since the density values in the center of a mass were low, we preferred to keep all the central circles with a diameter of seven pixels in our sector model as shown in Fig. 4. Thus, the template matching procedure was applied using the sector-shaped images like R and L in Fig. 4 as the reference images.

If we define the density value of one point in the breast image and the density value of the corresponding point in the reference

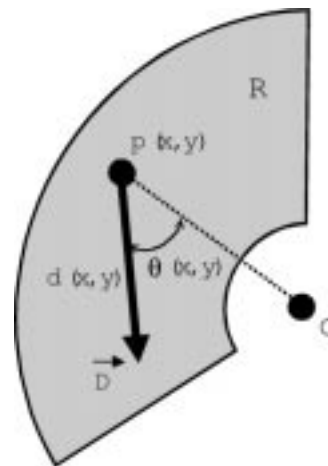


Fig. 5. Elements for calculating the concentration. $d(x, y)$ represents the intensity component of the density gradient. The value of $\theta(x, y)$ shows the angle determined by the direction from the center of mask to point (x, y) and the direction of the density gradient. \vec{D} : density gradient vector.

image as $F(i+x, j+y)$ and $T(x, y)$, respectively, we can determine $R(i, j)$ as the correlation coefficient for template matching

$$R(i, j) = \frac{\sum \sum \{F(i+x, j+y) - \bar{F}\} \{T(x, y) - \bar{T}\}}{\sqrt{\sum \sum \{F(i+x, j+y) - \bar{F}\}^2} \sqrt{\sum \sum \{T(x, y) - \bar{T}\}^2}} \quad (1)$$

where \bar{F} and \bar{T} are the means of each image. (i, j) and (x, y) correspond to mammogram and reference image coordinates, respectively. Summations were performed over the following range of values: $x = 0, 1, \dots, n$ and $y = 0, 1, \dots, n$ (n : integer).

3) *Concentration:* The concentration $G(i, j)$ was defined by the accumulative values in both regions A and B in Fig. 4, and was calculated by the density gradient from a Sobel filter

$$G(i, j) = \sum_x \sum_y d(x, y) f(x, y) \quad (2)$$

where $d(x, y)$ represents the intensity component in the density gradient as shown in Fig. 5 and $f(x, y)$ is given by the following:

$$f(x, y) = \cos[\theta(x, y) + n \sin\{2\theta(x, y) - \pi\}] \quad (3)$$

where $\theta(x, y)$ and point (x, y) are shown in Fig. 6. Since $f(x, y)$ is influenced by the direction component of the density gradient, it increases when the direction component turns to the center of the mask. From the definition above, it is noted that $d(x, y)f(x, y)$ becomes greater when the direction of the density gradient turns to the center of mask and the value of density gradient becomes larger. In (3), “ n ” is a fixed number, and the value of the equation changes by varying “ n ” as shown in Fig. 6. The sin term is added inside the cos term so that a few changes of $\theta(x, y)$ will not influence $f(x, y)$ greatly. The value of 0.2 for “ n ” was determined experimentally. Furthermore, the greater such components of the density gradient are, the greater the value of concentration $G(x, y)$ will be.

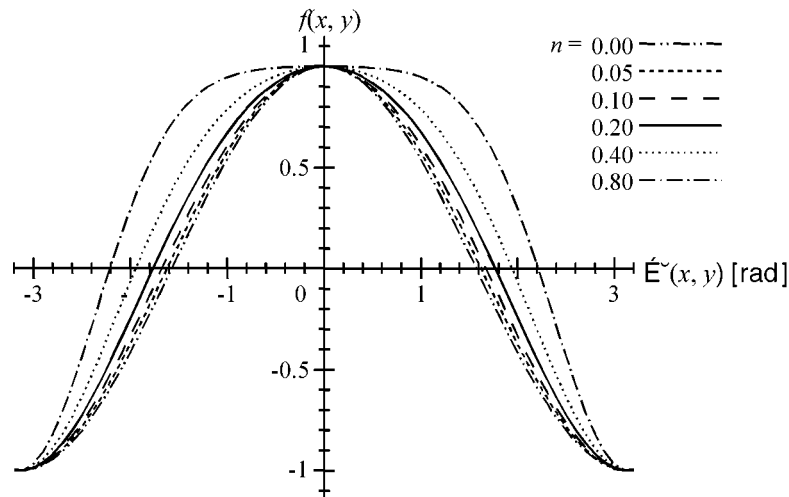


Fig. 6. The curves of $f(x, y)$ defined in (3) are plotted for six different values of n .

C. Detecting the Masses With a Partial Loss by Similarity

1) *Extraction of Suspicious Candidate Points:* We used $R(i, j)$ and $G(i, j)$ to calculate the similarity between the suspicious candidates and the sector model of a mass. Considering that when a candidate is a true mass, its similarity to the sector model should be significant, we defined the similarity as

$$S(i, j) = R(i, j)G(i, j) \quad (4)$$

The similarity was calculated in all search areas described in Section II-A. It should be noted that for the two search areas in Fig. 3, the patterns for feature calculation and template matching were different. As illustrated in Fig. 4, for *Area B*, the similarity was calculated using both patterns, *A* and *R*, *B* and *L*, in Fig. 4. But, for *Area A*, only the patterns *A* and *R* were employed to obtain the similarity since using patterns *B* and *L* caused more misdetections in the pectoralis muscle region. The reason for this is explained below. Generally speaking, the pixel values tend to increase (get darker) as one goes from the upper right portion of the mammogram (toward the pectoralis muscle) to the lower left (toward the nipple). In particular, changes in pixel value are the greatest near the boundary between the pectoralis muscle region and the breast region. To calculate the correlation coefficient $R(i, j)$, we use a 41×41 pixel area of the mammogram and the pattern *R* or *L*. Since pattern *L* is based on a Gaussian distribution, the pixel values increase as one goes from the center to the lower left corner. Likewise, if one considers certain sections of the pectoralis muscle region, the pixel values increase in a similar fashion. In addition, when one considers the mass candidate region, the same pattern emerges. Because of this, the number of false detections tends to increase when one uses pattern *L* in the pectoralis muscle region. However, the pixel values of pattern *R* increase as one goes from the center to the upper left. When one considers the same sections of the pectoralis muscle region mentioned before, none of the pixel values increase in that direction. Therefore, using pattern *R* in the pectoralis muscle region, the pixel patterns of a mass,

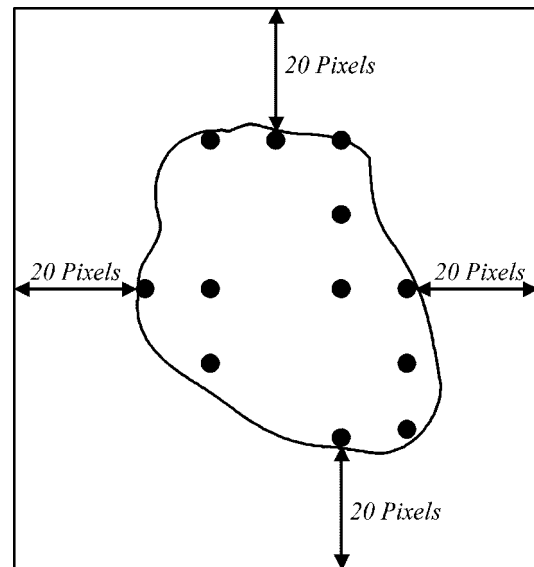


Fig. 7. The area of suspicious candidates is decided by a quadrangle. The black dots indicate the detected candidate points.

like the one in the leftmost image in Fig. 1, and those of the pectoralis muscle region appear different. As a result, it is possible to reduce the number of false mass detections.

The threshold value of $S(i, j)$ was set experimentally to 56 so that points whose similarity was greater than this value were extracted as suspicious candidate points of malignant masses.

2) *Determining the Area of Suspicious Candidates:* In addition, the area of suspicious candidates was determined using the information from the candidate points extracted above. The points whose distance to neighboring candidate points were less than 20 pixels were considered to be part of the same suspicious candidate. The quadrangle for suspicious candidates was determined by extending 20 pixels from the four farthest points among all detected suspicious points in both the x and y directions (Fig. 7). We found that all points of similarity larger than the threshold around the highest point(s) of similarity corresponded to malignant masses, so we regarded any candidate not following this pattern to be a FP.

D. Elimination of FPs

The FPs were eliminated using a second-order statistics technique [12]. There were four parameters utilized. From the gray-level co-occurrence matrix, three second-order statistics values, angular second moment (ASM), inverse difference moment (IDM), and entropy (ENT), were calculated. In addition, a contrast (CNT) was set by the matrix based on the gray-level difference method [8]. These features are defined as follows:

$$\text{ASM} = \sum_{i,j} Q(i,j) \cdot Q(i,j) \quad (5)$$

$$\text{IDM} = \sum_{i,j} \frac{Q(i,j)}{1 + (i-j)^2} \quad (6)$$

$$\text{ENT} = - \sum_{i,j} Q(i,j) \cdot \log Q(i,j) \quad (7)$$

$$\text{CNT} = \sum_{i,j} k \cdot k \cdot P(i,j) \quad (8)$$

where Q represents the co-occurrence matrix, k the difference of pixel values, and P the gray-level difference.

In (5), the ASM represents the texture's smoothness. The smoother the texture is the greater the value of the ASM becomes. In (6), the IDM is large if the difference between i and j is small in the matrix, such as in an image with flat pixel distribution. In (7), the ENT is large if the constituent values of the matrix tend to be equal. The ENT signifies the degree of overall variance in the image's pixel values. In (8), the CNT is large if the difference between neighboring pixel values is large. Since there was tendency for the ASM value to be small inside FPs, we considered those candidates having small ASM values as FPs. In addition, since the values of the IDM, ENT, and CNT were experimentally found to be large in actual masses, we eliminated those candidates having small values as FPs.

III. RESULTS AND DISCUSSION

To evaluate the new method for detecting masses with a partial loss of region, we examined 335 mammograms including 30 images with a partial loss of mass region. The other 305 mammograms were normal cases.

First, this dataset was tested by our existing method. Generally, the detection area for the existing method was defined as the whole breast area. In this study, however, to compare the existing and new methods, we limited the detection area to the *Areas* defined in Fig. 3. By adjusting the threshold values of the existing method, we obtained the free-response receiver operating characteristic (FROC) curve shown in Fig. 8. According to the results of the curve, when the fixed threshold was employed, a TP rate of 70% was achieved with 0.23 FPs/image (point "a" in Fig. 8). By changing the threshold, it was possible to get a TP rate of 90% at an expense of 0.83 FPs/image (point "b" in Fig. 8).

Then the performance of the new method was investigated. By changing the threshold value of the similarity, a FROC curve was drawn. When the threshold value of the similarity was 56, the TP rate was 90% with 0.24 FPs/image (point "c" in Fig. 8). Even when the threshold value was lowered to 40 or 30, the TP

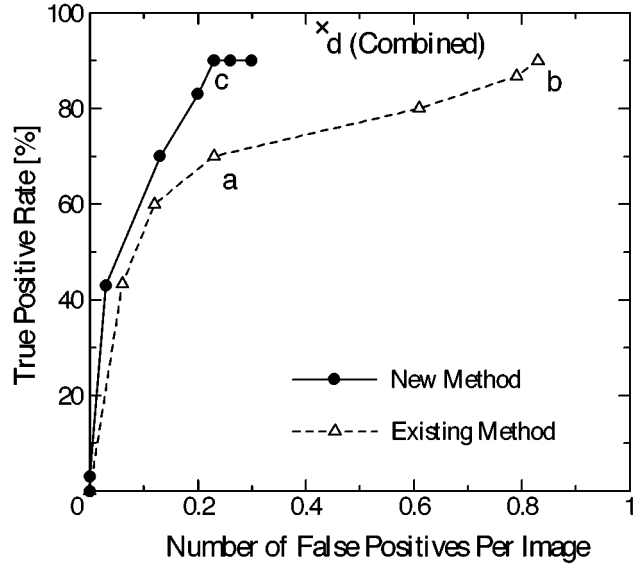


Fig. 8. FROC curves for a database of 335 Japanese mammograms using either the existing method or the new method, limited the detection area only to *Area A* and *Area B* as defined in Fig. 3.

rate remained at 90% at the expense of an increased number of FPs ranging from 0.26 to 0.30/image.

From these FROC curves, it is noted that when the number of FPs varies from 0 to 0.8, the TP rate achieved by our new method is higher than that of the existing method, with fewer FPs/image. Such results suggest that the new method may improve the performance of our CAD system for detecting masses.

The new method failed to detect three masses. The mass on one of the three images was not detected mainly because there were calcifications within the mass. The reason that the other masses were not detected was that the malignant masses appeared as a linear shape which was significantly different from our sector model for masses. However, in two of these cases, our existing method was able to detect the masses properly; the other mass was not detected because most of the mass was lost when the mammogram was taken. Therefore, a combination of both methods may improve our CAD system's performance.

The results of combining both methods are shown below. Here, we used an OR operation if either method indicated a mass. The dataset was the same one introduced above, and a TP rate of 97% with a 0.43 FPs/image (point "d" in Fig. 8) was obtained when the search area for both methods was limited to the *Areas* shown in Fig. 3. However, in actual use the existing method would be applied to the entire mammogram and the new method to *Areas A* and *B*. Considering this, we conducted another test. Our results yielded a TP rate of 97% with 1.2 FPs/image. This indicates the further necessity of improving the FP reduction method.

We also experimented on a different database composed of 1 075 clinical mammograms. The database consisted of 31 masses, of which five showed a partial loss of region. The existing method detected two of the masses with a partial loss of region but failed to detect the other three. The new method was able to detect all of the partial-loss masses with 0.29 FPs/image. For the existing method, the TP rate was 74% with 1.2 FPs/image. Combining the two methods yielded a 10%

increase in the TP rate with only a 0.2 increase in the number of FPs/image.

In conclusion, a new technique for detecting masses with a partial loss of region was proposed in this study. The partial loss mass was identified by its similarity to a sector form model in the template matching process. After applying our new method to 335 digitized mammograms, the detection sensitivity for the partial loss masses jumped from 70% to 90% when the number of FPs was kept constant (0.2/image). Moreover, the combination of these two methods improved the TP rate up to 97%. Such results indicate that the new technique may improve the overall performance of our CAD system for mammographic masses effectively.

ACKNOWLEDGMENT

The authors would like to thank J. Cole, S. Li, H. Futamura, H. Yoshimura, and A. Kano.

REFERENCES

- [1] H. Fujita, T. Endo, T. Matsubara, K. Hirako, T. Hara, H. Ueda, Y. Torisu, N. Riyahi-Alam, K. Horita, C. Kido, and T. Ishigaki, "Automated detection of masses and clustered microcalcifications on mammograms," in *Proc. SPIE*, vol. 2434, The Int. Soc. Opt. Eng., Medical Imaging 1995: Image Processing, 1995, pp. 682–692.
- [2] S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," *IEEE Trans. Med. Imag.*, vol. 8, pp. 377–386, Aug. 1989.
- [3] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, "Markov random field for tumor detection in digital mammography," *IEEE Trans. Med. Imag.*, vol. 14, pp. 565–576, Sept. 1995.
- [4] F. F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique," *Med. Phys.*, vol. 21, no. 3, pp. 445–452, 1994.
- [5] A. J. Mendez, P. G. Tahoces, M. J. Lado, M. Souto, and J. J. Vidal, "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms," *Med. Phys.*, vol. 25, no. 6, pp. 957–964, 1998.
- [6] K. Matsumoto, H. R. Jin, and H. Kobatake, "Detection method of malignant tumors in DR images -Iris filter-," *Trans. Inst. Electron. Inform. Commun. Eng. D-II*, vol. J75-D-II, no. 3, pp. 663–670, 1992.
- [7] S. Hashimoto, H. Kobatake, H. Takeo, and S. Nawano, "Reduction of false positives using qualitative feature comparison of bilateral mammograms," *Trans. Inst. Electron. Inform. Commun. Eng. D-II*, vol. J83-D-II, no. 1, pp. 415–418, 2000.
- [8] T. Matsubara, H. Fujita, S. Kasai, M. Goto, Y. Tani, T. Hara, and T. Endo, "Development of new schemes for detection and analysis of mammographic masses," in *Proc. IEEE Computer Society Int. Conf. Intelligent Information Systems (IASTED)*, H. Adeli, Ed., 1997, pp. 63–66.
- [9] T. Matsubara, H. Fujita, T. Endo, K. Horita, M. Ikeda, C. Kido, and T. Ishigaki, "Development of a high-speed processing algorithm for masse detection based on thresholding technique in mammograms," *Med. Imag. Tech.*, vol. 15, no. 1, pp. 1–13, 1997.
- [10] T. Matsubara, S. Kasai, K. Seki, H. Fujita, T. Hara, and T. Endo, "Development of a computer-aided diagnosis system for mammogram: Improvement of the method of extracting low-density regions during automated mass detection," *J. Jpn. Ass. Breast Can. Scr.*, vol. 7, pp. 87–101, 1998.
- [11] O. Ohtsuka, S. Kasai, Y. Hatanaka, H. Fujita, T. Hara, and T. Endo, "Elimination of false-positive mass candidates using second-order statistics in a mammogram CAD system," *Med. Imag. Inform. Sci.*, vol. 16, no. 1, pp. 13–19, 1999.
- [12] S. Kasai, H. Fujita, T. Hara, Y. Hatanaka, and T. Endo. (1999) Elimination of funicular-shaped false-positive candidates in an automated detection algorithm for mammographic masses. *J. Computer-Aided Diag.f Med. Images* [Online], vol (2), pp. 1–7. Available: <http://www.toriwaki.nuie.nagoya-u.ac.jp/~cadm/Journal/index.html>
- [13] —, "Elimination of false-positive candidates by comparing right and left mammograms in an automated mass detection algorithm," *Med. Imag. Tech.*, vol. 16, no. 6, pp. 655–666, 1998.
- [14] T. Hara, O. Ohtsuka, H. Fujita, and T. Endo, "Elimination of false-positive mass candidates using prewitt filter in a mammogram CAD system," *Jpn. J. Radiol. Technol.*, vol. 56, no. 3, pp. 449–454, 2000.
- [15] M. Kato, H. Fujita, T. Hara, and T. Endo, "Improvement of automated breast-region-extraction algorithm in a mammogram CAD system," *Med. Imag. Inform. Sci.*, vol. 14, no. 2, pp. 104–113, 1997.
- [16] T. Nakagawa, S. Kasai, T. Hara, H. Fujita, and T. Endo, "Automated extraction of pectoral-muscle region on mammogram," *Med. Imag. Inform. Sci.*, vol. 18, no. 1, pp. 39–45, 2001.
- [17] S. Kasai, T. Matsubara, H. Fujita, T. Hara, Y. Hatanaka, and T. Endo, "Detection algorithm for masses existing around thick mammary gland regions on mammograms," *Jpn. Sci. Med. Bio. Eng.*, vol. 38, no. 2, pp. 111–117, 2000.