Development of an automated detection system for microcalcifications on mammograms by using the higher-order autocorrelation features

Y. Ohe, N. Shinohara, T. Hara, H. Fujita, T. Endo*, T. Iwase**
Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Science, Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu, 501-1193 Japan
*Department of Radiology, National Hospital of Nagoya, 4-1-1 Sannomaru, Naka-ku, Nagoya, 460-0001, Japan
**Department of Breast Surgery, Cancer Institute Hospital, 1-37-1 Kamiikebukuro, Toshima-ku, Tokyo, 170-8455, Japan

ABSTRACT

The purpose of this work is to develop a new pattern recognition method using the higher-order autocorrelation features (HOAFs), and to apply this to our microcalcification detection system on mammographic images. Microcalcification is a typical sign of breast cancer and tends to show up as very subtle shadows. We developed a triple-ring filter for detecting microcalcifications, and the prototype detection system is nearly complete. However, our prototype system does not allow for the detection of three types of microcalcifications, two of which are amorphous and linear microcalcifications and the third is obscured microcalcifications which is often confused with the background or circumference that have almost the same density. We targeted the amorphous type of microcalcification, which has a low contrast and easily goes undetected. The various features of microcalcifications and false-positive (FP) shadows were extracted and trained using the multi-regression analysis, and unknown images were recognized as a result of this training. As a result, amorphous microcalcifications were successfully detected with no increase in the number of FPs compared with our existing detection method.

Keywords: mammogram, computer-aided detection (CAD), microcalcifications, amorphous, higher-order autocorrelation features

1. INTRODUCTION

With the recent westernization of our life styles, the incidence rate of breast cancer is fast increasing in Japan, as in the U.S. and advanced countries in Europe. Breast cancer is now the highest occurring among various types of women's cancers. To make correct diagnoses of breast cancer, tactile examination alone is insufficient, and mammographic examination is being increasingly emphasized as a primary diagnostic modality. On the other hand, this may impose a heavy burden on the physicians who make these diagnoses by observing mammographic images because it must be essentially avoided that any affected area will be missed out or the images will be misinterpreted due to excessive dependence on their own subjective judgment. To address this problem, we have made and are currently making attempts to develop a computer-assisted diagnosis (CAD) system, which assumes the role of an auxiliary imaging system for breast cancer diagnosis [1]. Breast cancer is largely classified into two categories: mass and microcalcification. We have almost completed our prototype detection system for both types of breast cancer [2-5]. Other proposed microcalcification detection systems include; those using morphological analysis by Okuno et al. [6], using the matched Filter by Nishikawa et al. [7], and using the wavelet transform by Strickland et al. [8].

When a group of fine calcifications is found on mammographic images, whether it is cancerous is determined on the basis of a comprehensive observation of its numbers, shape, size, and distribution. When a large number of calcifications clustered in a small area are observed, there is almost no doubt that the tissue in that area has turned cancerous. The density distribution value of the site of microcalcification is characterized in that it is smaller as compared to those of the surrounding tissue (pixel values). In other words, the images essentially have a common property their density gradients decrease almost linearly toward the center. We have developed a triple-ring filter [4] based on this property, which has been efficiently extracted focusing on the direction and magnitude at which the density gradient falls in the breast area. Our microcalcification detection system generally involves the following

Medical Imaging 2004: Image Processing, edited by J. Michael Fitzpatrick, Milan Sonka, Proceedings of SPIE Vol. 5370 (SPIE, Bellingham, WA, 2004)
1605-7422/04/$15 doi: 10.1117/12.536847

processing procedures. First, the triple-ring filter is used to detect all suspected microcalcifications on the mammographic images at the initial examination. Second, the candidates for microcalcifications are narrowed down to eliminate any FP cases based on conventional feature parameters such as size, contrast, and roundness. Finally, in the cluster detection process, the distance between the adjacent microcalcification candidates is measured to find the site at which the given number or more of microcalcifications aggregate in the specified area (a cluster of microcalcifications).

However, our prototype system does not allow for detection of three types of undetectable microcalcifications, two of which are amorphous and linear microcalcifications, and the third is that which goes undetected (obscured) due to confusion with the background or circumference that have almost the same density. These types are undetectable even by using the triple-ring filter at the initial examination. The main cause for the lack of detectability is the fact that they have density distributions that are different from those of the typically-shaped microcalcifications. In some cases, microcalcifications are not detected as a cluster because of the undetected ones. In this context, it is very important to detect them, because these microcalcifications are amorphous and easily go undetected.

In this paper, we focus on amorphous microcalcifications from the three types of the undetectable calcifications. Amorphous microcalcifications are indicated as those with a diameter of 3 pixels or less and a low contrast. The contrast is defined as the difference between the minimum pixel value for a signal and the average value of its surrounding pixels (i.e., a microcalcification image). For amorphous microcalcifications, the contrast is approximately 250 pixels or lower, that is, the difference in optical density is approximately 0.25 or lower. The contrast of the detectable microcalcifications on our prototype system is approximately 250 pixels or higher, and the difference in optical density is approximately 0.25 or higher. At this time, the noise level of the background (standard deviation) is approximately 45 and the CNR (contrast-to-noise ratio) levels of amorphous and detectable microcalcifications are approximately 5.5 or lower and approximately 5.5 or higher, respectively. These amorphous microcalcifications may be detected using the conventional parameter adjustment technique, although FPs are supposed to increase in number with the use of this technique. Therefore, we propose a new detection approach to the automated microcalcification detection method using HOAFs, which is widely recognized as an image recognition method [9].

2. MATERIALS AND METHOD

The detection procedure consists of two stages: training and recognition. The training stage involves the substitution of the features of various images for a multi-regression model. The recognition stage involves the determination of whether the signal indicates calcification based on the result of the training stage. The detection procedure is described in detail below.

2.1 Image digitization

The images on all of the films (screen/film mammograms) are digitized for analysis using our CAD system. The digitization process is performed using a Konica LD5500 laser digitizer with a sampling pitch of 0.05 mm and a 12-bit density resolution for the optical density range from 0.0 to 4.0. All the image data were subsampled at the suited pitch of 0.1 mm to detect microcalcifications.

2.2 Classification of calcifications

To investigate the types of undetectable calcifications using our prototype detection system, a radiologist with some experience in diagnosis using mammographic images, pointed to the spots of calcifications on 38 films with a calcification clusters. As an indication method for the physician, we had prepared both film (screen/film mammograms) and digital images on a high-intensity monitor. A radiologist checked the film for any spot(s) where a calcification exists, and plotted them, if any, on a high-intensity monitor.

The radiologist found out approximately 1,900 calcifications. Any microcalcifications undetected by our computer system were picked up by comparing the positions of calcification candidates, which were detected on the prototype system, with those indicated by the radiologist. These calcifications were classified into the following three categories at our discretion:

1. those with a low contrast type (called amorphous microcalcifications) (Fig.1)
2. those with a linear and moderately low contrast type (called linear microcalcifications)
3. those confused with (obscured by) the background or circumference that have almost the same density (called obscured microcalcifications)

The number of microcalcifications for each type (amorphous, linear, and obscured) is 120, 77, and 154, respectively. The detection rate of the prototype system is approximately 82% at the initial examination.

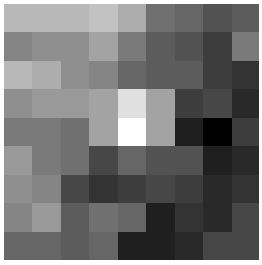


Fig. 1 An amorphous microcalcification.

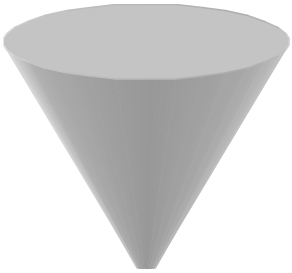


Fig. 2 Structure of microcalcification on a mammogram.

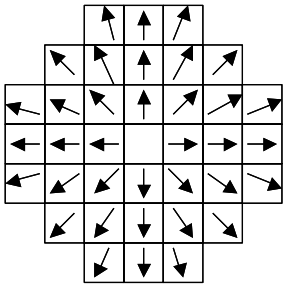


Fig. 3 Basic vector patterns in the feature extraction.

2.3 Feature extraction

2.3.1 Feature extraction using density gradients and a triple-ring filter

A typical microcalcification, as shown in Figs. 2 and 3, has a very small signal with a diameter of 8 pixels or less. The mammographic microcalcifications have various shapes, such as a circle, a comma, a V-shape, or a segment. However, the optical density is lower than that of the surroundings. This indicates that the images essentially have a common property their density gradients fall toward the center. Thus, the first-order approximation to microcalcification will result in the circular cone as shown in Fig. 2. Therefore, it is anticipated that the microcalcification pattern can be detected if the area, which contains approximated circular-cone formations, is efficiently extracted based on the direction and the magnitude at which the density gradient falls in the breast area. By calculating the triple-ring filter, the features of the direction and magnitude are extracted.

(a) Density gradients

The density gradient is calculated for each pixel from the density distribution of a digital image. Let the density at the pixel (i, j) of an original image be $f(i, j)$. Using the 8-neighbor pixel values, $\Delta_x f(i, j)$ and $\Delta_y f(i, j)$, for which the first-order derivatives are in the x and y directions, respectively, are calculated by

$$\Delta xf(i,j) \equiv f(i+1,j-1)/2 + f(i+1,j) + f(i+1,j+1)/2 - \{f(i-1,j-1)/2 + f(i-1,j) + f(i-1,j+1)/2\} \quad (1)$$

$$\Delta yf(i,j) \equiv f(i+1,j-1)/2 + f(i,j-1) + f(i-1,j-1)/2 - \{f(i+1,j+1)/2 + f(i,j+1) + f(i-1,j+1)/2\} \quad (2)$$

Then, magnitude G and direction θ of the gradient on the digital image are given by Eqs. (3) and (4), respectively,

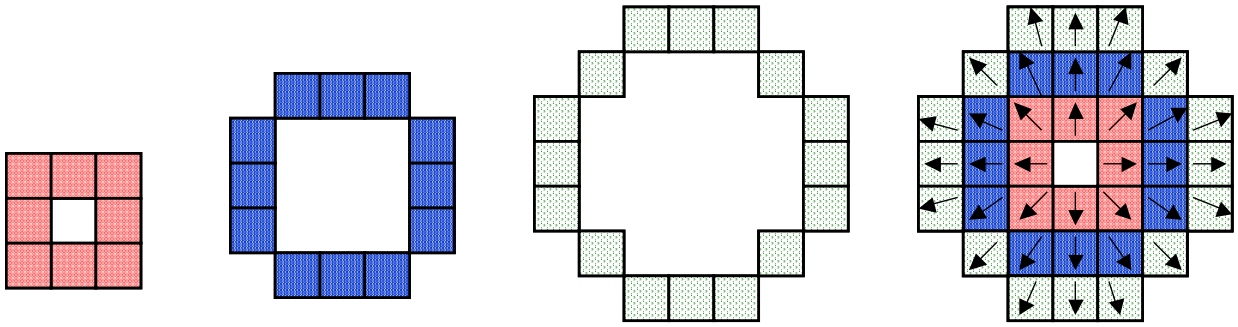
$$G = \sqrt{\{\Delta xf(i,j)\}^2 + \{\Delta yf(i,j)\}^2} \quad (3)$$

$$\theta = \tan^{-1}\{\Delta yf(i,j)/\Delta xf(i,j)\} \quad (4)$$

The gradient obtained by this calculation corresponds to Sobel's gradient.

(b) The Structure of a filter

To efficiently detect microcalcifications, the following feature filter extracts the area containing the vector pattern of an approximated circular-cone formation as shown in Fig. 3, from the two-dimensional vector distribution of the density gradient. The filter is composed of three-ring subfilters of different sizes, as shown in Figs. 4(a)-(c). These subfilters are designed to correspond to the various sizes of microcalcifications. They are labeled Filter A, Filter B, and Filter C (from the smallest to the largest); their diameters are 3, 5, and 7 pixels, respectively. Fig. 4(d) shows the three filters with the vector pattern of the approximated circular-cone formation superimposed as the basic vector pattern of the microcalcifications (called basic vector pattern). These three filters are superimposed on the candidate area. By analyzing the basic vector pattern and the vector pattern in the filter, the "directional feature" and the "magnitude feature" of the vector are calculated for each subfilter, as is described in the following paragraph. The feature extraction filter shown in Fig. 4 is called a triple-ring filter and is used in microcalcification detection.



(a) (b) (c) (d)

Fig. 4 (a)-(c): Feature extraction filter composed of three sub-ring filters with different diameters. (d): A triple-ring filter shown with basic vector patterns.

(c) The Calculation of directional features

In calculating the directional feature, it should be considered that the malignant calcification has a feature that its shape and contour is irregular. On Filter A, which is positioned closest to the center, the pattern of this calcification matches well with the basic vector pattern. In this context, on Filters B and C, which are positioned away from the center, they significantly deviate from the basic vector pattern. These properties must be considered in the design of filters for feature extraction. For this purpose, function d was introduced for calculation of a directional feature, as shown in Fig. 5. The weight for the directional feature is adjusted by setting the directional feature coefficient α to a different value for each subfilter. The coefficient was experimentally set to: α = -0.35 for Filter A, α = -0.15 for Filter B, and α = 0 for Filter C. In other words, the directional tolerance is set to a larger value for the outer subfilter than for the inner one.

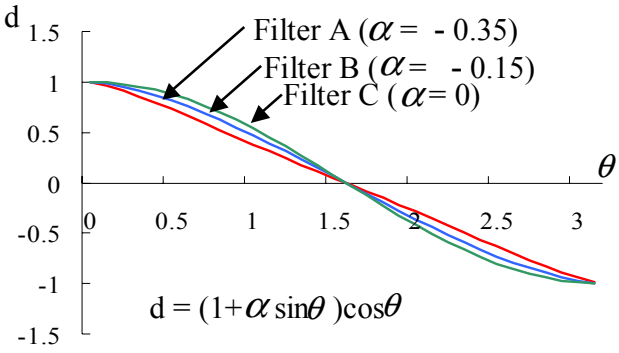


Fig. 5 Functions used for the calculations of direction feature value in each subfilter.

The feature parameter is calculated by the following procedure. The pixels are numbered based on the number of pixels in the subfilter. The candidate-side vector at the pixel with the k-th address on the subfilter is represented by $|\vec{V}|$. Let the difference in directions from the basic vector pattern be $\theta_k (0 \le \theta_k \le \pi)$. Then, the directional feature d for this pixel is determined from the directional feature function as shown in Fig. 5. The directional feature used in the extraction process is the average of those for pixels on the subfilter (hereafter, called a directional feature).

Consequently, the directional feature D for the subfilter composed of n pixels is given by

$$D = \frac{1}{n} \sum_{k=1}^{n} (1 + \alpha \sin\theta_k) \cos\theta_k \qquad (5)$$

where n = 8 for Filter A, n = 12 for Filter B, and n = 16 for Filter C. The directional feature D takes any value from -1 to 1. It approaches 1 when the candidate pattern is closer to the basic vector pattern, and it approaches -1 when it is further from the basic vector pattern. In the actual processing, $\theta(0 \le \theta \le \pi)$, representing the difference in the direction is replaced by an integer from 0 to 16, which helps to reduce both processing time and memory.

(d) The calculation of magnitude features

In the calculation of a magnitude feature, a direction parameter is also considered. When the vector falls toward the center (i.e., when the directional feature is 1), the magnitude is evaluated as high, and as low for such a pattern on a vessel image. The magnitude feature I determined for each subfilter is calculated as the average of the products of the vector magnitudes and the directional features for the pixels in the subfilters (called the magnitude feature). It is given by

$$I = \frac{1}{n} \sum_{k=1}^{n} |V_k| (1 + \alpha \sin \theta_k) \cos \theta_k \qquad (6)$$

(e) Features on a triple-ring filter

Although microcalcifications are unexceptionally confined to a limited area, some extremely fine calcifications can go undetected using Filter C (largest filter). On the other hand, larger calcifications often having a flat shape at the bottom are difficult to be detected even using Filter A, which is closest to the center. Therefore, the explanation variable of the multi-regression analysis requires a total of 8 features; the direction and magnitude features of Filter A, Filter AB, Filter BC, or Filter ABC on a triple-ring filter. The explanation variable of the multi-regression analysis is described in Section 2.4.

2.3.2 HOAFs (higher-order autocorrelation features)

The higher-order autocorrelation function means the expansion procedure using a higher-order of autocorrelation. There are millions of divisions, directional displacements, and combinations of pixels. On the other hand, local correlations between neighboring pixels are important in image recognition. In this method, we use a two-dimensional autocorrelation within an area of 3×3 pixels, for which the displacement is centered on the reference point r. When the pixel value at the reference point r of a target image is $f(r)$, the two-dimensional autocorrelation features $x2$ for the directional displacements ($a1$, $a2$) is given by Eq. (7):

$$x2(a1, a2) = \quad f(r) f(r + a1) f(f + a2) dr \qquad (7)$$

As the autocorrelation feature, a binary image procures 25 local feature's patterns from Eq. (7) excluding the same patterns produced by the translation operation. A grayscale image procures 35 local feature's patterns excluding the same patterns by translation operation. These local features are obtained by multiplying pixel values with each other in various combinations as local patterns, shown in Fig. 6. Thus, the obtained features are capable of enduring the translation of the objects, and when the number of target images is two or more, the number of features of the whole image will be the sum of the targets' features. These advantages are important in recognizing and measuring images. On the other hand, the HOAFs also have a disadvantage in that they are difficult to endure the rotation operation of objects.

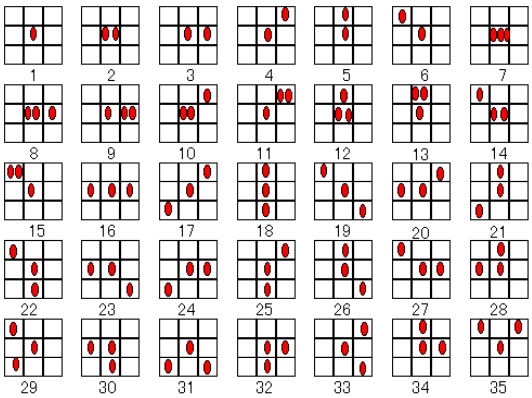


Fig. 6 Higher-order autocorrelation features.

We experimentally determined the size of ROI as 9×9 considering the sizes of calcifications on mammographic images in this method. In addition, the pixel values have major effects on pattern recognition in local areas. To address this problem, we shifted the minimum pixel value within the area to 1 prior to the calculation of the HOAFs.

2.4 Training and recognition using multi-regression analysis

2.4.1 Multi-regression analysis

The features of microcalcifications and FP shadows were extracted and trained using multi-regression analysis. Multi-regression analysis is one of the multivariate analysis techniques and is composed of criterion variables y and explanation variables $x_1, x_2, \ldots, x_p$. The expression is shown below

$$Y_N = a_1 x_{1N} + a_2 x_{2N} + \ldots + a_p x_{pN} + a_0 \qquad (8)$$

To estimate a criterion variable *y*, the coefficient matrix $a_0, a_1, \ldots, a_p$, where the sum of the squares of errors between a training signal *y* and an actual measurement *Y* as the minimum value may be determined. Assuming that the number of data is *N* and the number of criterion variables is *p*, the sum of squares of errors Q is given by

$$\begin{aligned} Q = & \{y_1 - (a_1 x_{11} + a_2 x_{21} + \ldots + a_p x_{p1} + a_0)\}^2 \\ & + \{y_2 - (a_1 x_{12} + a_2 x_{22} + \ldots + a_p x_{p2} + a_0)\}^2 \\ & + \ldots + \{y_N - (a_1 x_{1N} + a_2 x_{2N} + \ldots + a_p x_{pN} + a_0)\}^2 \end{aligned} \qquad (9)$$

The simultaneous-linear-equation (10) is obtained by the partial differential calculation. The coefficient matrix $a_0, a_1, \ldots, a_p$ can be obtained by evaluating a group of equations (10).

$$\begin{cases} \partial Q / \partial a_1 = 0 \\ \partial Q / \partial a_2 = 0 \\ \quad \vdots \\ \partial Q / \partial a_N = 0 \end{cases} \qquad (10)$$

The solution of a group of equations (10) agrees with that of a group of simultaneous-linear-equations (11) using a variance-covariance matrix. Therefore, the coefficient matrix $a_0, a_1, \ldots, a_p$ is uniquely determined by evaluating a group of equations (11).

$$\begin{cases} S_{x1^2} a_1 + S_{x1x2} a_2 + \ldots + S_{x1xp} a_p = S_{x1y} \\ S_{x1x2} a_1 + S_{x2^2} a_2 + \ldots + S_{x2xp} a_p = S_{x2y} \\ \qquad \vdots \\ S_{x1xp} a_p + S_{x2xp} a_2 + \ldots + S_{xp^2} a_p = S_{xpy} \end{cases} \qquad (11)$$

Where, $S_{x1^2}, S_{x2^2}, \ldots, S_{xp^2}$ are variances of $x_1, x_2, \ldots, x_p$, and S_{xmxn} is a covariance between x_m and x_n. S_{xy} is a covariance between $x_1, x_2, \ldots, x_p$ and *y*. a_0 is derived from $a_0 = \overline{y} - (\overline{x_1} a_1 + \overline{x_2} a_2 + \ldots + \overline{x_p} a_p)$. $\overline{y}$ and $\overline{x_m}$ are the mean values of *y* and x_m, respectively.

In this method, the unique determination of the coefficient matrix $a_0, a_1, \ldots, a_p$ is defined as training. The number of data *N* is the number of images used in training, and the number of criterion variable *p* is the number of local features (a total of 45 features) calculated from the triple-ring filter (8 features), the mean and the variance of the pixel values of the local area (2 features), and the HOAFs (35 features).

2.4.2 Training method

We employed 120 amorphous calcification images and 120 normal images, are obvious, for training. These calcifications were undetectable ones at the initial examination using the prototype system. This means that training data composed of a total of 240 images was prepared.

The number of criterion variables was 45, and a training signal would equal 1 if a calcification was formed in the local area, otherwise 0. The training procedure involves the following steps:

1. Selection of a 9×9 pixel area, which includes one amorphous calcification or none.
2. Extraction of feature values based on a triple-ring filter and the HOAFs.
3. Calculation of training using the round-robin method, since the number of training images was not too large. The number of training images changed from 218 to 239 because the number of calcifications depended on each image.
4. Determination of the coefficient matrix using the multi-regression analysis.

2.4.3 Recognition method

If the coefficient matrix $a_0, a_1, \ldots, a_p$ of the equation (8) is uniquely determined, the actual measurement *Y* can be obtained by extracting the features of an unknown image. A similarity between an unknown image and a training image is evaluated on the basis of the actual measurement *Y*. In this paper, the recognition procedure involves the following steps.

1. Selection of a 9×9 pixel area of an unknown image.
2. Extraction of feature values based on a triple-ring filter and the HOAFs.
3. Calculation of an actual measurement from the coefficient matrix using the multi-regression analysis.
4. Determination of whether calcifications are formed or not.

When the actual measurement approaches 1(0.8~1.5), it is estimated that microcalcifications have been formed in the area. All areas of the unknown image are scanned by this procedure.

3. RESULTS AND DISCUSSION

We used this method to evaluate 38 mammographic images. Figure 7 shows the result of the study, in which the vertical axis represents the true-positive (TP) rate for undetectable microcalcifications, and the horizontal axis represents the FPs per image. It is called the FROC (free-response receiver operating characteristic) curve. The new method A resulted from the training process of 45 features (a triple-ring filter, the HOAFs, and a mean and variance), and the new method B resulted from the training process of only 8 features (a triple-ring filter).

The number of FPs per image was always lower for the new method A than that for the existing method if TP was the same as that for the latter method. This demonstrated that the new method A was effective. It is considered that the restriction of the types of microcalcifications brought about better results. Then, the detection performance of the new method A was better than the new method B. It seems that the HOAFs, which reflect local features, are effective in detecting amorphous microcalcifications.

When observing a combination of the area identified using the new method A and that identified using the existing method at the initial examination, the rate of cluster detection was as follows: Sensitivity was 94% (260/276), and the number of FPs per image was 0.61 (339/556). In our prototype system, sensitivity was 92% (254/276), and the number of FPs per image was 0.50 (280/556). This demonstrated that the combination of the new method A and the existing method was not useful. This may be caused because the elimination process of FPs for the existing method was used. In other words, it is necessary to develop a method for the exclusion of any FP appropriate for detecting a certain type of microcalcification.

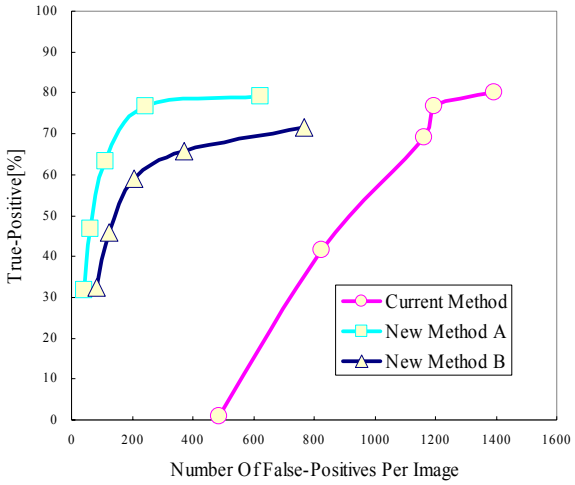


Fig. 7 Comparison of detection accuracies obtained by existing and new methods in terms of FROC curves.

4. CONCLUSION

We proposed a new pattern recognition method using HOAFs on the microcalcification detection system using mammography. The result of our experiment proved that this method could identify amorphous microcalcifications, which could not have been detected using the conventional system, and its detection performance might be improved by combining it with our conventional system.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of Health, Labor and Welfare under a Grand-In-Aid for Center Research, in part by the Ministry of Education, Culture, Sports, Science and Technology under a Grand-In-Aid for Scientific Research, Japanese Government and in part by Daiwa Securities Health Foundation.

REFERENCES

1. H. Fujita, T. Endo, T. Hara, et al., "Automated detection of masses and clustered microcalcifications on mammograms," Proc. SPIE, Image Processing, 2434, 682-692, 1995.
2. T. Matsubara, H. Fujita, S. Kasai, et al., "Development of a new algorithm for detection of mammographic masses," *in Proc. of the 4th International Workshop on Digital Mammotraphy, in DIGITAL MAMMOGRAPHY '98*, pp.139-142, Elsevier Science, 1998.
3. S. Kasai, D. Kaji, A. Kano, et al., "Mass detection algorithm for digital mammograms based on an adaptive thresholding technique utilizing multi-resolution processing"*, in Proc. of the 6th International Workshop on Digital Mammography, in DIGITAL MAMMOGRAPHY*, pp.334-338, Springer, 2002.
4. K. Hirako, H. Fujita, T. Hara, et al., "Development of detection filter for mammographic microcalcifications: A method based on density gradient and triple-ring filter analysis", IEICE, J72-D- U(9), 1334-1335, 1995.
5. K. Hirako, H. Fujita, T. Hara, et al., "Detection scheme for clustered microcalcifications with newly introduced contrast correction technique and variable-ring filter analysis", Medical Imaging Technology, 14-6, 665-679, 1996.
6. K. Okuno, H. Kobatake, S. Nawano et al., "Development of an adaptive system for microcalcification detection", Medical Imaging Technology, 14(6), 699-706, 1996.
7. RM. Nishikawa, K. Doi, ML. Giger, et al., "Computerized detection of clustered microcalcifications: Evaluation of performance on mammograms from multiple centers", Radiographics, 15, 443-452, 1995.
8. RN. Strickland and HI. Hahn: Wavelet transforms for detecting microcalcifications in mammograms, IEEE Trans. Medigal Imaging, 15, 218-229, 1996.
9. J. A. Mclaughlin and J. Raviv "Nth-order autocorrelations in pattern recognition", Information and Control, 12, 121-142, 1968.

*ooe@fjt.info.gifu-u.ac.jp; phone +81-58-293-2746; fax +81-58-230-1895; URL: http://www.fjt.info.gifu-u.ac.jp