Studies on Computerized Analysis of Clustered Microcalcifications on Mammograms for Diagnosis Aid

A DISSERTATION SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING
MIE UNIVERSITY FOR THE DEGREE OF
DOCTOR OF ENGINEERING

BY
RYOHEI NAKAYAMA

December 2010

ABSTRACT

In order to overcome the increase in breast cancer death rate, mammography which is considered the most sensitive method for detection of early breast cancers has been introduced for breast cancer screening in many advanced nations. Clustered microcalcifications which are present in 30%-50% of all cancers found in a mammography are one of the important radiographic indications on mammograms. However, it is often difficult for radiologists to detect clustered microcalcifications correctly because they are very small and obscure. It is also difficult to differentiate between benign and malignant clustered microcalcifications. With the concept of CAD (Computer-aided Diagnosis), it is expected that radiologists' performance in the diagnosis of medical images will improve by taking into account the analysis result of a lesion obtained from a computerized analysis as a "second opinion." Therefore, the purpose of this dissertation research is to develop a computerized analysis for both a detection aid and a differentiation aid of clustered microcalcifications on mammograms.

For a detection aid, we have developed a computerized detection method for indicating a potential region of clustered microcalcifications on mammograms for radiologists. In the computerized detection method, it is important to not only detect clustered microcalcifications with high sensitivity but also segment individual microcalcifications while maintaining their shapes because image features are extracted from the detected clustered microcalcifications in the computerized analysis for a differentiation aid. For the computerized detection method, therefore, we first constructed a novel filter bank with the requirements for a perfect reconstruction by introducing the concept of a Hessian matrix for classifying nodular structures and linear structures. A mammogram image is then decomposed into several subimages for a second difference at scales from 1 to 4 by this filter bank. The subimages for the nodular component (NC) and the subimages for the

nodular and linear component (NLC) are then obtained from analysis of the Hessian matrix based on those subimages for second difference. Many regions of interest (ROIs) are selected from the mammogram image. In each ROI, eight objective features are determined from each of the subimages for NC at scales from 1 to 4 and the subimages for NLC at scales from 1 to 4. A Bayes discriminant function with the eight objective features is employed for distinguishing between abnormal ROIs with clustered microcalcifications and normal ROIs without clustered microcalcifications. The region connecting the ROIs classified as abnormal ROI is considered to be a potential region of clustered microcalcifications. With the proposed detection method, sensitivity and a false positive rate was 100.0% and 0.98 per image, respectively.

For the differentiation aid, we developed a computerized classification method for providing radiologists the likelihood of histological classifications of clustered microcalcifications and a computerized retrieval method for providing radiologists images of lesions with known pathology similar to an unknown lesion. There are differences in both the image features and the growth speeds among histological classifications of clustered microcalcifications. In the computerized classification method, therefore, we extracted six objective features from clustered microcalcifications on each of the follow-up magnification mammograms (i.e. both current and previous magnification mammograms). In identifying histological classification of clustered microcalcifications, the histological classification of an unknown new case in question is assumed to be the same as that of the nearest neighbor case which has the shortest Euclidean distance in a feature-space. The feature-spaces for the nearest neighbor case consist of six objective features obtained from the previous magnification mammogram (previous features), six objective features obtained from the current magnification mammogram (current features), and the set of the six previous features and the six current features. The classification accuracies with the six current features were higher than those with the six previous features. These classification

accuracies were improved substantially by using the set of the six previous features and the six current features. With the proposed classification method, the classification accuracies were 90.9% (10 of 11) for invasive carcinoma, 89.5% (17 of 19) for noninvasive carcinoma of the comedo type, 96.0% (24 of 25) for noninvasive carcinoma of the noncomedo type, 82.6% (19 of 23) for mastopathy, and 93.3% (14 of 15) for fibroadenoma. In order to also present radiologists similar lesions as a differentiation aid, we investigated four objective similarity measures as an image-retrieval tool for selecting lesions similar to unknown lesions in terms of radiologists' visual perception. In the observer study, we confirmed that the presentation of similar images can improve radiologists' performance in the differential diagnosis of clustered microcalcifications on mammograms.

The proposed computerized analysis for both detection aid and differentiation aid for clustered microcalcifications achieves high detection performance and high classification accuracies, and would help radiologists improve the diagnosis accuracy of clustered microcalcifications at mammography in clinical practice.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor, Professor Koji Yamamoto, Suzuka University of Medical Science, for his kind guidance, valuable advice, and warmhearted encouragement through my research career. He provided me many opportunities and a great environment for me to learn professional skills to become a scientist. He also spent many hours of his valuable time in improving the quality of this research.

I am grateful to Professor Hideo Kobayashi, Professor Fumitaka Kimura, Professor Yoshikatsu Ohta, and Professor Shinji Tsuruoka, Mie University, for their time and effort in serving on my dissertation committee and for sharing their expertise. Their valuable comments were very helpful for the progress in this dissertation.

I am also grateful to Dr. Kiyoshi Namba, Hokuto Hospital, and Dr. Ryoji Watanabe, Hakuaikai Hospital, for their help and support for introducing and sharing their clinical perspective. They also provided me many opportunities to have the personal connections in this research field around the world. They always encouraged me to pursue my dream.

I would like to express my great appreciation to Professor Kunio Doi, The University of Chicago, and Professor Shigehiko Katsuragawa, Kumamoto University. They educated me carefully on the matters of exchanging ideas, expressing my thoughts, and writing scientific papers. I thank Assistant Professor Hiroyuki Abe, The University of Chicago, for his help and support by introducing and sharing his clinical perspective and by providing observer resources. I also thank Professor Junji Shiraishi, Kumamoto University, for his assistance in using the ROC software and for sharing his thoughts regarding the ROC analysis and statistical analysis.

I express my sincere appreciation to my senior colleague, Associate Professor

Yoshikazu Uchiyama, Oita National College of Technology, for his support and encouragement. I thank present and former members of my laboratory, Mr. Yosuke Mizutani, Mr. Toshihito Kawamura, Mr. Akiyoshi Hizukuri, Mr. Takanori Ogino, and Mr. Nobuo Nakako for my pleasant laboratory life.

I am also grateful to Ms. Elisabeth Lanzl, The University of Chicago, and Associate Professor Mark LaForge, Suzuka University of Medical Science, for improving this dissertation.

Finally, I would like to express my sincere gratitude to my well-beloved father, Tsuneki Nakayama, my well-beloved mother, Junko Nakayma, and especially to my dear wife, Kayoko Nakayama, and my dear son, Katsuaki Nakayama. They supported and encouraged me at all times and in any situation. I would like to dedicate this dissertation to them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Breast Cancer

Breast cancer is one of the major health problems in woman in advanced nations. Especially in the United States, it was estimated that more than 240,000 breast cancers occurred among women in 2007 and that one of eight women has breast cancer during their lives [1]. In Japan, both the disease rate and the mortality rate in breast cancer continue to increase every year. It is estimated that about 35,000 breast cancers occur among women in a year and that one of twenty women has breast cancer [2]. Breast cancer has had the highest disease rate of all the cancers for Japanese women. Early diagnosis and early treatment are very important to reduce breast cancer mortality [3]. Therefore, mammography which is considered the most sensitive method for detection of early breast cancers has been introduced to breast cancer screening in many advanced nations. Some studies [4-6] of randomized clinical trials report that that periodic mammography screening can reduce the breast cancer mortality.

Clustered microcalcifications and mass are the important radiographic indications which might relate to breast cancer on mammograms [7, 8]. Although mass can be recognized more clearly on an ultrasound image, it is difficult to obtain information on clustered microcalcifications from any modality except mammogram. Clustered microcalcifications are also present in 30%-50% of all cancers found at mammography [9, 10]. Therefore, it is very important to detect clustered microcalcifications accurately on mammograms. However, 10%-30% of cancers are missed at breast cancer screening for several reasons such as non-palpable lesions, radiologists' fatigue by repetitive tasks, and technical limitations of mammograms [11-16]. It is also difficult for radiologists to determine whether the detected lesion is benign or malignant. The positive predictive

value of mammography, i.e., the ratio of the number of breast cancers found to the number of biopsies, is typically between 15% and 30% [17-19]. Unnecessary biopsies cause patients both physical and monetary problems.

1.2 Computer-aided Diagnosis

The concept of a computer-aided diagnosis (CAD) was invented as an approach for improvement of radiologists' performance in the diagnosis of medical images [20-23]. CAD is defined as a diagnosis made by a radiologist who takes into account the analysis result of a lesion obtained from a computerized scheme as a "second opinion." It is expected to improve the diagnostic accuracy and the consistency in the radiologists' image interpretation with CAD. In generally CAD, a computerized analysis for detection aid and that for differentiation aid are used according to the purpose.

1.2.1 Computerized Analysis for Detection Aid

For the detection aid of clustered microcalcifications, many investigators have developed computerized schemes [24-39] for identifying potential regions of clustered microcalcifications on mammograms. Chan *et al.* [24, 25] developed a computerized analysis based on a difference image technique in which a signal-suppressed image was subtracted from a signal-enhanced image in order to remove structured background on the mammogram. Li *et al.* [26] proposed using fractal background modeling, taking the difference between the original image and the modeled image. Karssemeijer [27, 28] developed a computerized analysis based on the use of statistical models and the general framework of Bayesian image analysis.

Other researchers have developed a computerized scheme based on wavelet transform [29, 30], which is a robust tool for image analysis, enhancement, and pattern recognition. Wavelet transform is basically a filtering technique that represents images hierarchically on

the basis of scale or resolution. It also provides a powerful method for analyzing high-spatial-frequency phenomena localized in space, and thus can effectively extract information derived from localized high-frequency signals, such as those emitted by microcalcifications. Strickland *et al.* [31, 32] used a discrete wavelet transform with biorthogonal spline filters to detect microcalcifications. They computed four dyadic scales and two additional interpolating scales, and applied a binary threshold-operator to all six scales. The responses of the individual wavelet scales were then combined by a summing rule, and the output was used to detect microcalcifications. Yoshida *et al.* [33, 34] multiplied every scale by a weight factor and reconstructed an image by applying the inverse transform in a discrete wavelet transform. The weights were determined by supervised learning, using a set of training cases. Clarke *et al.* [35] and Qian *et al.* [36, 37] applied a denoising to the image and then took the high-pass scale of a discrete wavelet transform using spline wavelets. This resulted in a general edge detector that could locate not only microcalcifications but also several other structures, such as film artifacts or lines. Laine *et al.* [38, 39] applied several wavelet-type filter bank decompositions, such as the dyadic wavelet transform. An adaptive enhancement operator was defined on the wavelet coefficient scales. They obtained effective contrast improvement values for irregular structures such as microcalcifications. The enhancement operator was defined separately for each scale.

1.2.2 Computerized Analysis for Differentiation Aid

For differentiation aid of clustered microcalcifications, many investigators have also developed computerized schemes [40-48] for estimating the likelihood of malignancy of clustered microcalcifications on mammograms. Jiang *et al.* [40] developed a computerized method for extracting eight image features of clustered microcalcifications, and they produced an estimate of the likelihood of malignancy by using an artificial neural

network. The classification performance of their computerized scheme was slightly greater than that of radiologists. Chan *et al.* [41-43] developed a classification method using various feature classifiers with the combination of morphologic features and texture features obtained from different views of the same clustered microcalcifications. Nakayama *et al.* [44, 45] extracted some objective features from clustered microcalcifications by taking into account the image features that radiologists commonly used for describing microcalcifications in order to distinguish between benign and malignant clustered microcalcifications. Shen *et al.* [46] used 3 shape features, i.e., compactness, moments, and Fourier descriptors, to classify individual microcalcifications by use of a nearest-neighbor classifier. Kallergi *et al.* [47] used the descriptors of morphology of the individual calcifications and the distribution of the cluster. Leichter *et al.* [48] evaluated the usefulness of objective features based on the shape of individual microcalcifications and those based on the geometry of clustered microcalcifications. Although the features based on the geometry was more useful than those based on the shape in distinguishing between benign and malignant clustered microcalcifications, the combination of both of features based on the shape and geometry provided a greater classification performance.

1.3 Purpose of This Dissertation Research

The goal of this dissertation research is to develop a novel computerized analysis for detection aid and differentiation aid of clustered microcalcifications on mammograms.

In the computerized analysis for detection aid, the detection performances of computerized methods based on wavelet transform [31-39] were relatively higher than those based on the other techniques described in the section of 1.2.1. These results may indicate that the multi-resolution analysis based on wavelet transform is useful for detection of microcalcifications because microcalcifications on mammograms present various sizes.

However, these computerized detection methods did not analyze the shape of the microcalcifications in detail. Microcalcifications are nodular in structure, whereas normal tissues such as blood vessels and mammary ducts are linear in structure. Therefore, we consider that it might be possible to detect clustered microcalcifications more accurately by introducing the shape information of individual microcalcifications into the multi-resolution analysis. In chapter 2, we first construct a novel filter bank based on the concept of the Hessian Matrix for classifying nodular structures and linear structures, and then develop a computerized detection method for clustered microcalcifications based on objective features obtained from the filter bank.

In study for differentiation aid, Jiang *et al.* [49] conducted observer performance studies for classifying clustered microcalcifications as malignant or benign without and with the computer output indicating the likelihood of malignancy. The radiologists' performance was improved significantly when they used the computer output. This result indicates that radiologists are able to use computer output as a second opinion to improve their diagnostic accuracy. However, the performance level of the computerized scheme was considerably greater than that of radiologists with the computer output. This result appears to imply that it is difficult for radiologists to have complete reliance on the computer output even if the computerized scheme has a high performance level. Making clinical decisions for biopsy or follow-up on clustered microcalcifications by taking into account possible histological classifications on magnification mammograms may reduce the number of unnecessary biopsies because there are differences in the growth speeds among histological classifications of lesions [7, 50]. Therefore, the likelihood of histological classifications estimated by a computerized analysis would be useful to radiologists for their decisions on patient management. In chapter 3, we develop a computerized classification method for histological classifications of clustered microcalcifications based on differences in both the image features and the growth speeds among histological classifications on

follow-up magnification mammograms (i.e. both current and previous magnification mammograms).

Radiologists learn diagnostic skills by viewing many cases in their training and clinical practice. Based on their experience and knowledge, they make diagnostic decisions on new, unknown lesions in medical images. Therefore, it is expected that the presentation of images of lesions with known pathology similar to a new, unknown lesion would be useful for radiologists in the differential diagnosis of the unknown lesion. In chapter 4, we investigate four objective similarity measures as an image-retrieval tool for selecting lesions similar to unknown lesions in terms of radiologists' visual perception. In chapter 5, we confirm that the presentation of similar images can improve radiologists' performance in the differential diagnosis of clustered microcalcifications on mammograms.

CHAPTER 2
COMPUTERIZED DETECTION METHOD OF CLUSTERED MICROCALCIFICATIONS

In the computerized detection method, it is very important to not only detect clustered microcalcifications accurately but also segment microcalcifications while maintaining their shapes because image features such as size and shape irregularity of each microcalcification are used in the computerized analysis for differentiation aid. Therefore, we constructed a novel filter bank having three important features: i) it could enhance the nodular component in image; ii) it could enhance the nodular and linear component in image; and iii) it could reconstruct an original image from the decomposed subimages of the original image. The nodular feature (N feature) at each scale from 1 to 4 and the nodular and linear feature (NL feature) at each scale from 1 to 4 are determined by the subimages for nodular component (NC) at scales from 1 to 4 and the subimages for nodular and linear component (NLC) at scales from 1 to 4. We then investigated the effectiveness of the N features and the NL features for detecting clustered microcalcifications. Finally, we evaluated the detection performance by applying the proposed detection method to 600 mammograms.

2.1 Materials

Our database consists of 1200 standard-view (cranio-caudal view and medio-lateral-oblique view) mammograms obtained from 300 patients in the DDSM (Digital Database for Screening Mammography, University of South Florida) [51]. Six hundred and ten clustered microcalcifications (239 malignant lesions and 371 benign lesions) are included in 603 of the 1200 mammograms. The remaining 597 images are normal mammograms without clustered microcalcifications. All mammograms were digitized to a pixel size of 0.0435 mm x 0.0435 mm and a 12-bit gray scale by use of a laser scanner. In order to train

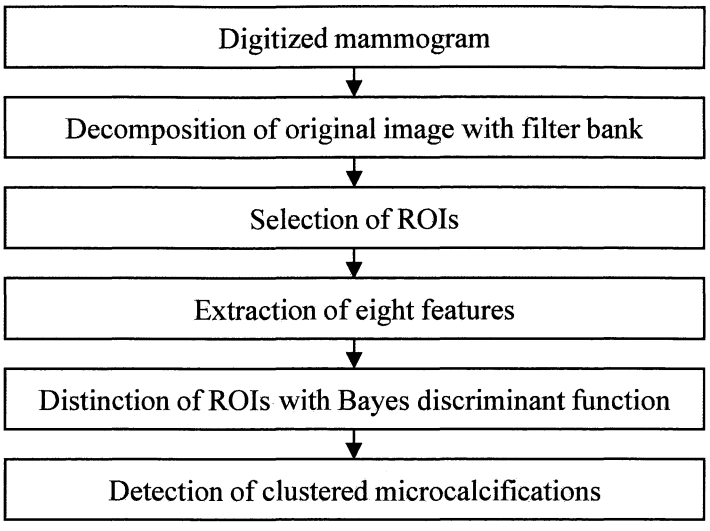


Figure 2.1: Schematic diagram of the proposed detection method for clustered microcalcifications on mammograms.

and evaluate the proposed detection method, we divided our database into a training set and a test set. Each set included 600 mammograms obtained from 150 patients. The total number of clustered microcalcifications is 300 in the training set, and 310 in the test set, respectively.

2.2 Overall Scheme for Detection of Clustered Microcalcifications

Figure 2.1 shows a schematic diagram of the proposed detection method of clustered microcalcifications on mammograms. Mammogram image is first decomposed into several subimages at different scales from 1 to 4 by a novel filter bank (section 2.3). These subimages are the horizontal subimage for second difference in the vertical direction, the vertical subimage for second difference in the horizontal direction, and the diagonal subimage for first difference in the vertical direction followed by first difference in the horizontal direction. The subimages for NC and the subimages for NLC are obtained from analysis of the Hessian matrix based on those subimages for the differences. Many regions of interest (ROIs) of 5 mm x 5 mm are then selected from the mammogram image

automatically. In each ROI, eight features are determined from the subimages for NC at scales from 1 to 4 and the subimages for NLC at scales from 1 to 4 (section 2.4). A Bayes discriminant function with these eight features is employed for distinguishing among abnormal ROIs with clustered microcalcifications and two different types of normal ROIs without clustered microcalcifications (section 2.5). The region connecting the ROIs classified as abnormal ROI is considered to be a potential region of clustered microcalcifications.

A 115 x 115 matrix (approximately 5 mm x 5 mm) is chosen as the ROI size because the clustered microcalcifications are defined as a region containing more than three microcalcifications per 5 mm x 5 mm area in clinical practice [7]. When the ROIs are selected at intervals of 5 mm so as to border on the adjacent ROIs, some clustered microcalcifications may be divided across two or more ROIs. These clustered microcalcifications might be not detected correctly because each ROI includes only the information on the divided cluster. Therefore, it is necessary to select ROIs at a shorter interval so that the center of clustered microcalcifications will be at the center of one of the ROIs. Although we must select ROIs at intervals of 1 pixel (0.0435 mm) to analyze a mammogram in detail, there are no large differences between adjacent ROIs selected at intervals of 1 pixel. It will also take much time to analyze redundant ROIs. In this study, therefore, we select the ROIs at intervals of 23 pixels (approximately 1 mm) so that one ROI would overlap with the adjacent ROIs.

2.3 Filter Bank for Detection of Nodular Components and Linear Components

2.3.1 Hessian Matrix Classifying Nodular Structures and Linear Structures

For the distinction between clustered microcalcifications and normal tissues on mammograms, it would be important to enhance both nodular components, such as microcalcifications, and linear components, such as blood vessels and mammary ducts.

The concept of the second derivative is well known as an enhancement technique for these components. The values of the second derivatives for the nodular structure in all directions become negative. On the other hand, the value of the second derivative for the linear structure becomes zero in the direction of the axis of the linear structure, whereas it becomes negative in the other directions. Therefore, the filters based on the second derivatives are often used for the detection or the enhancement of the nodular structure and the linear structure. Shimizu *et al.* [52, 53] defined a minimum directional difference filter (Min-DD Filter) based on the smallest value of the second derivatives in all directions, and a maximum directional difference filter (Max-DD Filter) based on the largest value of the second derivatives in all directions. They then applied the Min-DD Filter to detect large lung nodules in chest X-ray images.

On the other hand, the smallest value and the largest value of the second derivatives in all directions can be calculated approximately by the small eigenvalue λ_1 and the large eigenvalue λ_2 of the Hessian matrix because the second derivative of the function $f(x,y)$ in an arbitrary direction θ is given by

$$\begin{aligned} &\frac{\partial^2 f}{\partial x^2}\cos^2\theta + 2\frac{\partial^2 f}{\partial x \partial y}\cos\theta\sin\theta + \frac{\partial^2 f}{\partial y^2}\sin^2\theta \\ &= (\cos\theta \quad \sin\theta)\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{pmatrix}\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}. \end{aligned} \tag{1}$$

Details are shown in the Appendix I. Therefore, the following formulas indicate the conditions that the two eigenvalues λ_1 and λ_2 must satisfy for a nodular structure and a linear structure, respectively:

$$\text{for a nodular structure: } \lambda_1 \cong \lambda_2 < 0, \tag{2}$$

$$\text{for a linear structure: } \lambda_1 < 0, \quad \lambda_2 \cong 0. \tag{3}$$

Li *et al.* [54] and Sato *et al.* [55, 56] enhanced the nodular component and the linear component in three-dimensional medical images by analyzing the eigenvalues of the

three-dimensional Hessian matrix.

2.3.2 Filter Bank for Detection of Nodular Components and Linear Components

As we described in the previous section, the nodular component and the linear component can be enhanced by using the value of the second derivative or the eigenvalue of the Hessian matrix. Although lung nodules have various sizes, the length of the filter for the second derivative was constant in the Shimizu's method [53]. Therefore, it might be possible to enhance nodular structures and linear structures more accurately by using filters for the second derivative with various sizes. In addition, it might be necessary to properly shape the nodular structure and the linear structure using a smoothing operator, because the second derivative is usually influenced by noise. These issues are solvable with a filter bank which consists of high-pass filters and low-pass filters of various lengths. Once clustered microcalcifications are detected, the next problem is to determine whether the detected lesion is benign or malignant. Many investigators have developed computerized analysis for distinguishing between benign and malignant clustered microcalcifications [40]-[48]. In these computerized analyses, it is important to segment microcalcifications while maintaining their shapes because image features such as size and shape irregularity are used to estimate the likelihood of malignancy for clustered microcalcifications. This issue was not taken into account in Li's and Sato's methods [54-56]. This issue is solvable by use of a filter bank that satisfies the requirement for perfect reconstruction. Therefore, we introduced the concept of the Hessian Matrix into a filter bank that satisfies the requirement for perfect reconstruction.

Figure 2.2 shows a two-channel filter bank. The analysis bank on the left has a lowpass filter $H_L(z)$, a highpass filter $H_H(z)$, and a downsampling operator $(\downarrow 2)$ which removes the odd-numbered components after filtering. The synthesis bank on the right has a lowpass filter $F_L(z)$, a highpass filter $F_H(z)$, and an upsampling operator $(\uparrow 2)$ which

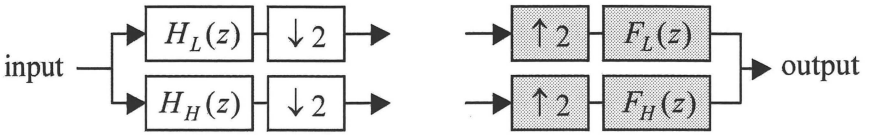


Figure 2.2: Two-channel filter bank: Separate the input frequency bands (filter and downsample). Then reassemble (upsample and filter).

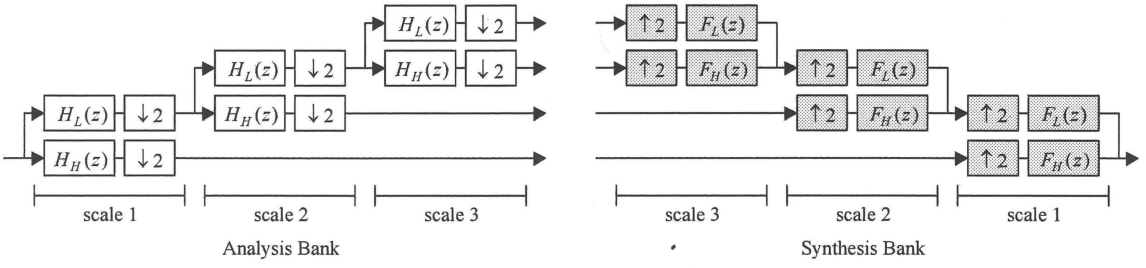


Figure 2.3: Filter bank for the one-dimensional discrete wavelet transform at scales from 1 to 3.

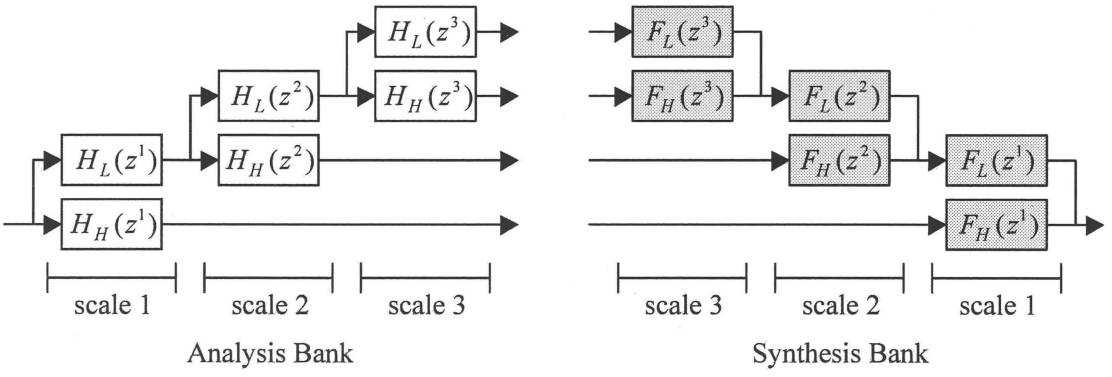


Figure 2.4: Filter bank without sampling operators at scales from 1 to 3.

inserts a zero in the odd components. The filter bank for the one-dimensional discrete wavelet transform is usually given by iterating the lowpass channel of the two-channel filter bank, as illustrated in Fig. 2.3. For the perfect reconstruction with an l-step delay, the filters of the two-channel filter bank must satisfy the following conditions [57]:

$$H_L(-z)F_L(z)+H_H(-z)F_H(z)=0, \tag{4}$$

$$H_L(z)F_L(z)+H_H(z)F_H(z)=2z^{-l}, \tag{5}$$

where $F_L(z)=H_H(-z)$ and $F_H(z)=-H_L(-z)$. This filter bank performs a discrete biorthogonal wavelet transform. For a discrete orthogonal wavelet transform, the following condition must be satisfied in addition to the above conditions [57]:

$$H_H(z)=-z^{-N}H_L(-z^{-1}), \tag{6}$$

where the overall system delay is $l=N$. Therefore, it is very difficult to give the filters for the Hessian matrix in a filter bank based on a discrete orthogonal wavelet transform with the perfect reconstruction. The analysis bank of these filter banks divides the input signal into two channels of two half-length outputs decimated by the downsampling operator. Together, these filter banks make up the maximally decimated filter bank [57]. The maximally decimated filter bank is usually used for image compression or transmission, because the length of the output signal obtained from the analysis bank is equal to the length of the input signal. However, we consider that the maximally decimated filter bank is not always useful for image analysis because details of the image are decimated by the downsampling. Therefore the filter bank without sampling operators is employed for detection of clustered microcalcifications. We can remove eq.(4) and eq.(6) from the perfect reconstruction conditions by using the filter bank without sampling operators. The condition for perfect reconstruction is given by only:

$$H_L(z)F_L(z)+H_H(z)F_H(z)=1. \tag{7}$$

Equation (5) changes to eq. (7) because the filter bank without sampling operators does not have the delay. Figure 2.4 shows the filter bank without sampling operators. Although $2j$ is usually employed for the order of z at scale j, we employ j in order to obtain details of the change of microcalcifications along the increase of scales.

Figure 2.5 shows the filter bank for the two-dimensional wavelet transform. $S_0 f$ is an original image. The smoothed subimage $S_1 f$ is obtained by successive applications

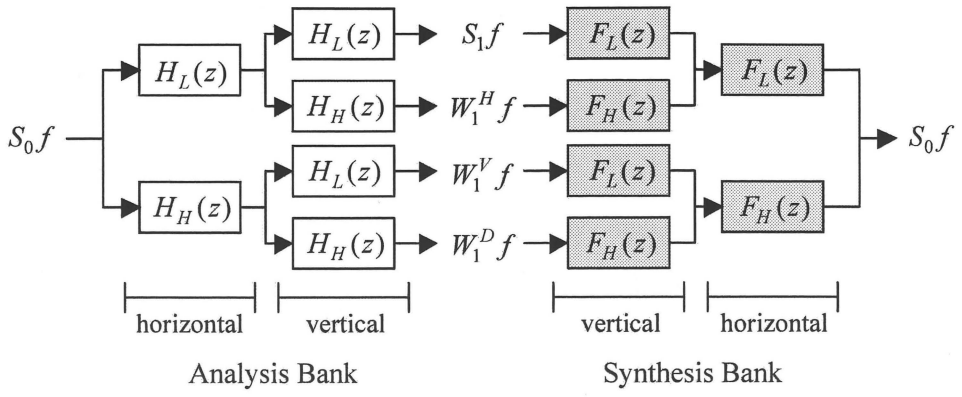


Figure 2.5: Filter bank for the two-dimensional wavelet transform.

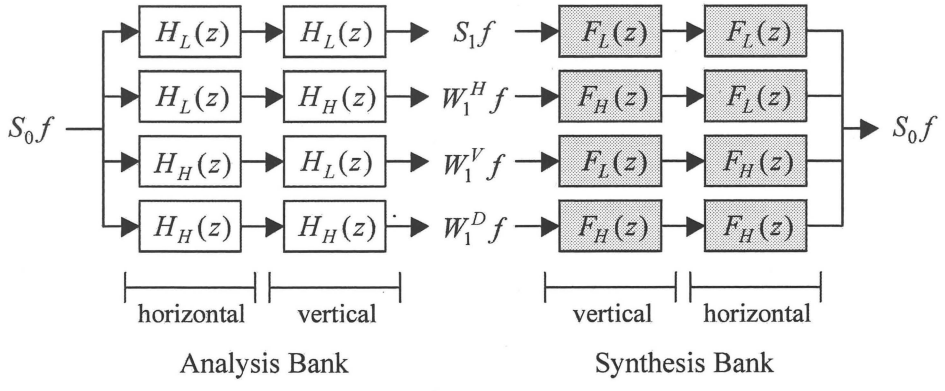


Figure 2.6: Filter bank for two-dimensional wavelet transform, which is the equivalent representation of the filter bank in Fig. 2.5.

of the vertical low-pass filter followed by the horizontal low-pass filter. The horizontal subimage $W_1^H f$ is obtained by applying the vertical high-pass filter followed by the horizontal low-pass filter. The vertical subimage $W_1^V f$ is obtained by applying the vertical low-pass filter followed by the horizontal high-pass filter. The diagonal subimage $W_1^D f$ is obtained by applying the vertical high-pass filter followed by the horizontal high-pass filter.

The filter bank in Fig. 2.6 shows the equivalent representation of the filter bank in Fig. 2.5. The filter bank satisfies the perfect reconstruction even if we move some filter in the

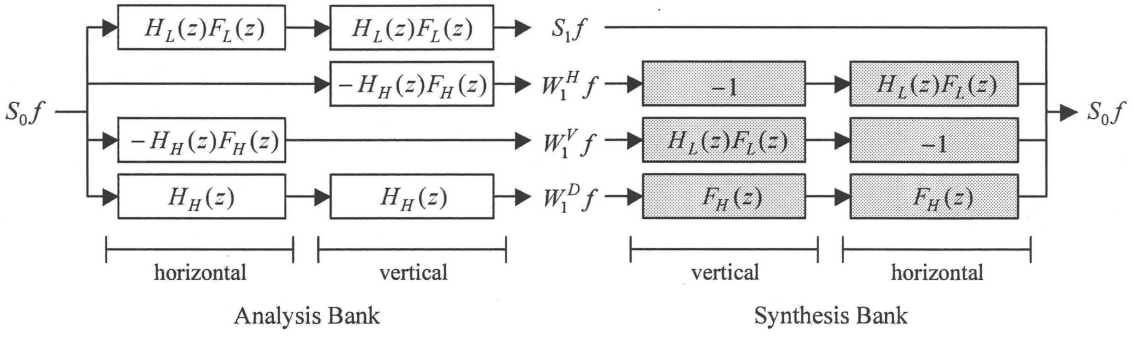


Figure 2.7: Filter bank obtained by changing the order of some filters of the filter bank in Fig. 2.6.

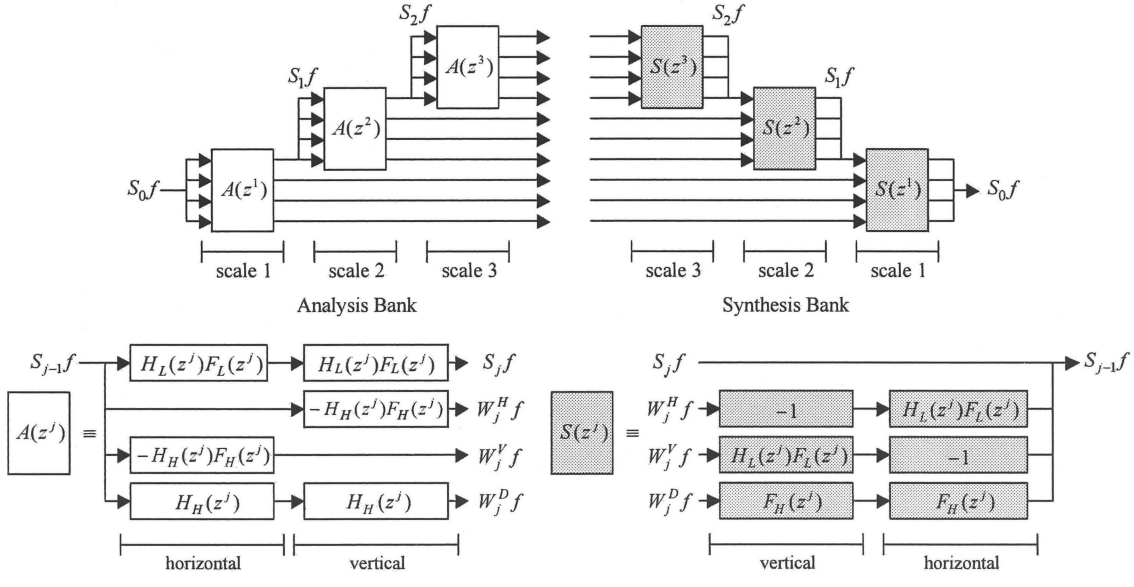


Figure 2.8: Filter bank obtained by iterating the lowpass channel of the filter bank in Fig. 2.7.

synthesis bank to the analysis bank because the filters are applied to the horizontal direction and the vertical direction independently. We obtain the filter bank in Fig. 2.7 by changing the order of some filters of the filter bank in Fig. 2.6. Note that $S_1 f$, $W_1^H f$, and $W_1^V f$ in Fig. 2.6 are not equivalent to $S_1 f$, $W_1^H f$, and $W_1^V f$ in Fig. 2.7. In order to obtain each element of the Hessian matrix from this filter bank, $H_H(z)$ and $F_H(z)$ are given by

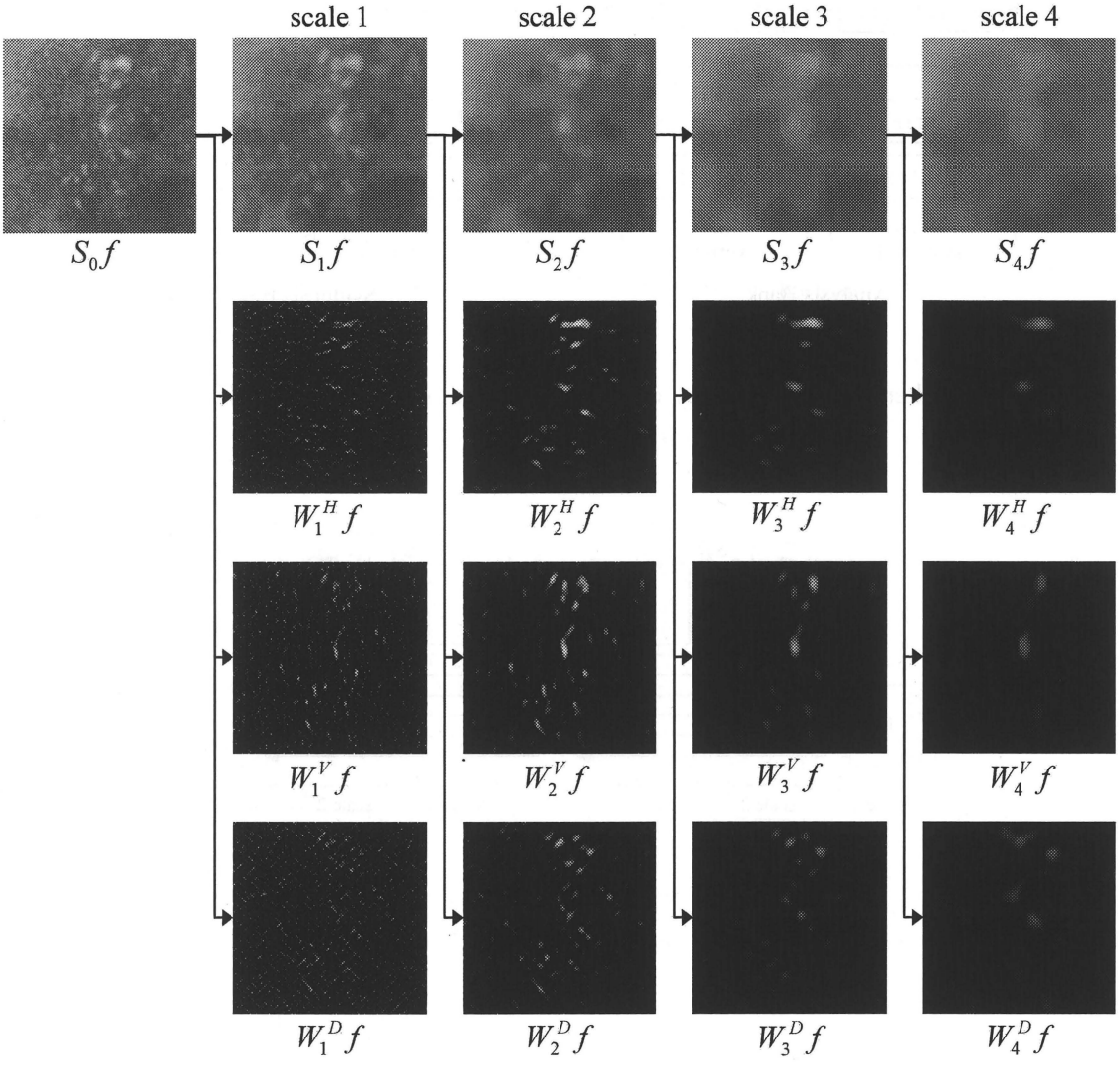


Figure 2.9: Subimages obtained from the analysis bank of the novel filter bank by input of an abnormal ROI with clustered microcalcifications.

$$H_H(z) = \frac{1}{2}(-z^1 + z^{-1}), \tag{8}$$

$$F_H(z) = \frac{1}{2}(z^1 - z^{-1}), \tag{9}$$

as new wavelet basis. $H_H(z)$ and $F_H(z)$ are the filters for the first difference. Therefore, $H_H(z)F_H(z)$ is the filter for the second difference, which is given by

$$H_H(z)F_H(z) = \frac{1}{4}(-z^2 + 2 - z^{-2}). \tag{10}$$

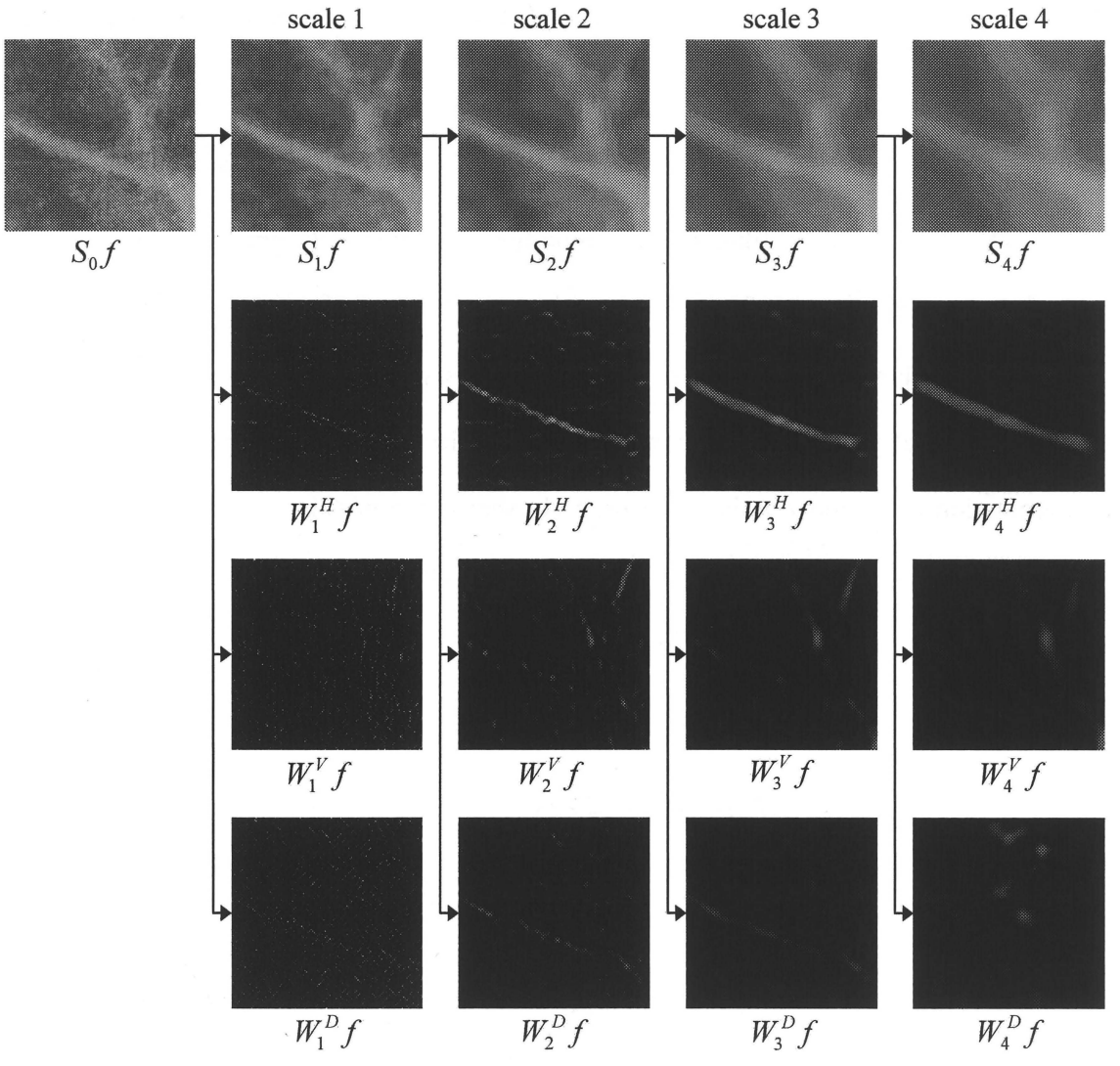


Figure 2.10: Subimages obtained from the analysis bank of the novel filter bank by input of a normal ROI with blood vessels.

The horizontal subimage $W_1^H f$ is the second difference in the vertical direction of the original image. The vertical subimage $W_1^V f$ is the second difference in the horizontal direction of the original image. The diagonal subimage $W_1^D f$ is the first difference in the vertical direction followed by the first difference in the horizontal direction of the original image. These subimages $W_1^H f$, $W_1^V f$, $W_1^D f$, correspond to the elements of the Hessian matrix. Because the condition for perfect reconstruction in the filter bank

without sampling operators is $H_L(z)F_L(z)+H_H(z)F_H(z)=1$, the other filter $H_L(z)F_L(z)$ is given by

$$H_L(z)F_L(z)=\frac{1}{4}(z^2+2+z^{-2}), \tag{11}$$

where $H_L(z)F_L(z)$ is the filter for the smoothing operator. By applying the smoothing filter to the original image $S_0 f$, the smoothed subimage $S_1 f$ at the next scale, i.e., the scale of 1, is obtained. The multi-resolution representation is obtained by iterating the lowpass channel of the filter bank, as shown in Fig. 2.8. This multi-resolution representation can remove noises and properly shapes the nodular structure and the linear structure. In this filter bank, $H_H(z)$, $F_H(z)$, $H_H(z)F_H(z)$, and $H_L(z)F_L(z)$ at scale j are given as

$$H_H(z)=\frac{1}{2}(-z^j+z^{-j}), \tag{12}$$

$$F_H(z)=\frac{1}{2}(z^j-z^{-j}), \tag{13}$$

$$H_H(z)F_H(z)=\frac{1}{4}(-z^{2j}+2-z^{-2j}), \tag{14}$$

$$H_L(z)F_L(z)=\frac{1}{4}(z^{2j}+2+z^{-2j}). \tag{15}$$

Figures 2.9 and 2.10 show the subimages obtained from the analysis bank of the novel filter bank by input of an abnormal ROI with clustered microcalcifications and a normal ROI with blood vessels, respectively. In the filter bank of Fig. 2.8, $NC_j(x,y)$ (the subimage for NC at scale j) is defined by the absolute value of the large eigenvalue λ_2 of the Hessian matrix at scale j. Here, the pixels that are $\lambda_2>0$ are given as zero because the eigenvalues for the nodular structure tend to become negative. $NLC_j(x,y)$ (the subimage for NLC at the scale j) is defined by the absolute value of the small eigenvalue λ_1 of the Hessian matrix at the scale j. Here, the pixels that are $\lambda_1>0$ are given as zero.

2.3.3 Fundamental Characteristics

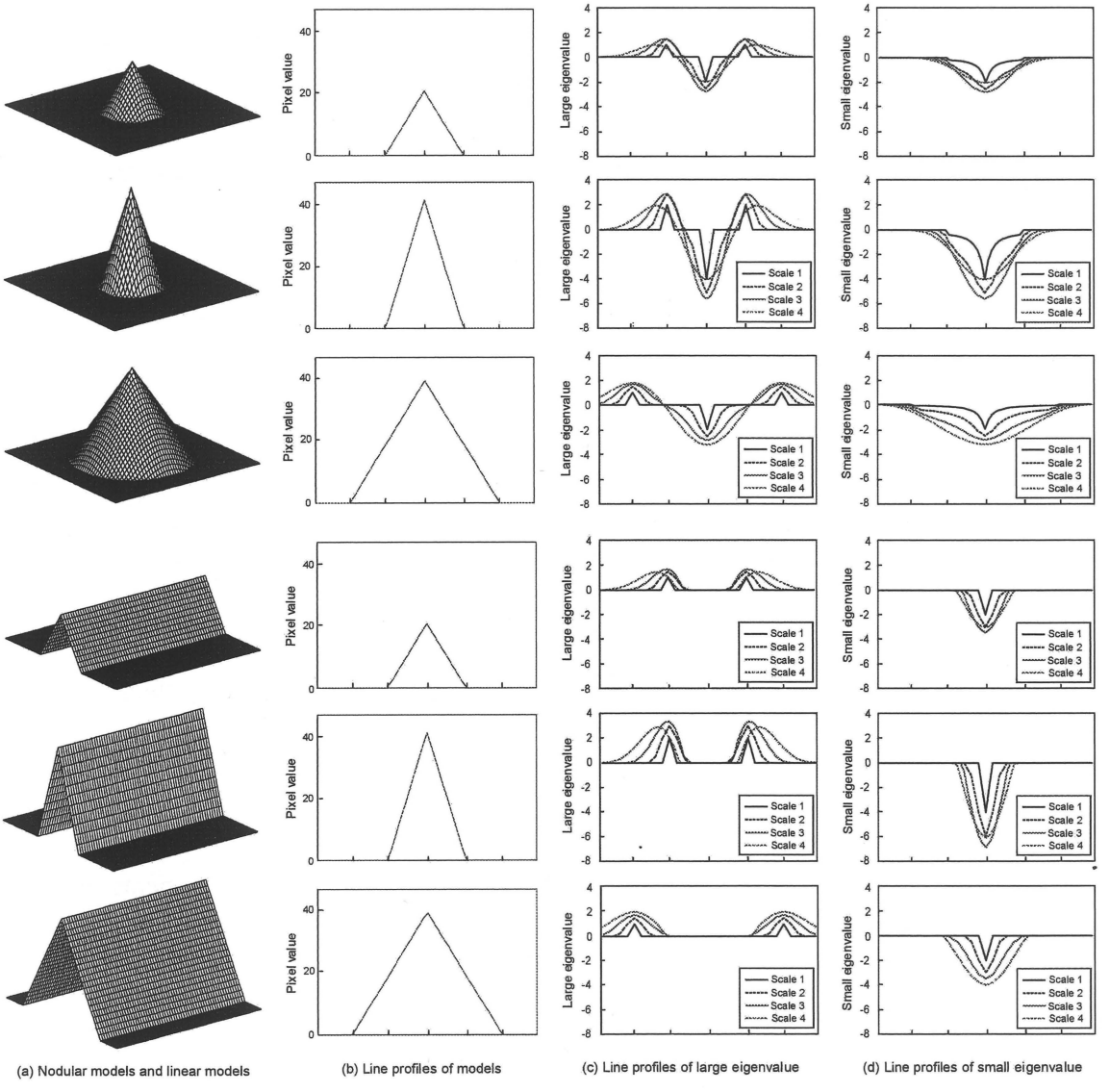


Figure 2.11: Fundamental characteristic of the absolute values of the small eigenvalue and large eigenvalue. (a) Artificial data consisting of three nodular models and three linear models with different pixel values and widths; (b) line profiles of the pixel value for artificial data; (c) line profiles of the large eigenvalue for artificial data; and (d) line profiles of the small eigenvalue for artificial data.

In order to investigate the fundamental characteristics of the subimages for NC and the subimages for NLC, we performed the simulations using two sets of artificial data. We used three nodular models and three linear models in Fig. 2.11(a). The pixel values at

each center of the nodular models are 20, 40, and 40, whereas those of the linear models are 20, 40, and 40. The diameters of the nodular models are 20 pixels, 20 pixels and 40 pixels, whereas the widths of the linear models with the length of 100 pixels are 20 pixels, 20 pixels, and 40 pixels. Figure 2.11(b) shows the line profiles on each center of the nodular models and linear models.

Figure 2.11(c) shows the line profiles of the large eigenvalue of the Hessian matrix for these models. Although the large eigenvalues near the edges of the nodular models and linear models were positive, most of the large eigenvalues for nodular models were negative. However, the large eigenvalues for linear models were not less than zero. When the pixel value at the center of the model doubles, the absolute values of the large eigenvalue also double. This is because these eigenvalues correspond to the values of second derivatives. The absolute values of the large eigenvalues at a coarser scale were larger than those at a finer scale when the diameter of the nodular model was increased. This indicates that the size of the nodular model can be estimated by comparing the absolute values of the large eigenvalues of each scale. Therefore, we can detect nodular components such as microcalcifications by using the absolute values of large eigenvalues at each scale. Here, the pixels that had large eigenvalues $\lambda_2 > 0$ were given as zero.

Fig. 2.11(d) shows the line profiles of small eigenvalues of the Hessian matrix for nodular models and linear models. Although the small eigenvalues near the edges of the nodular models and linear models were zero, most of the small eigenvalues for the nodular models and linear models were negative. The fundamental characteristics which depended on the pixel value and the diameter of models were the same as the characteristics of the large eigenvalues. Therefore, we can detect both nodular components, such as microcalcifications, and linear components, such as blood vessels and mammary ducts, by using the absolute value of the small eigenvalues at each scale. Here, the pixels that had small eigenvalues $\lambda_1 > 0$ were given as zero.

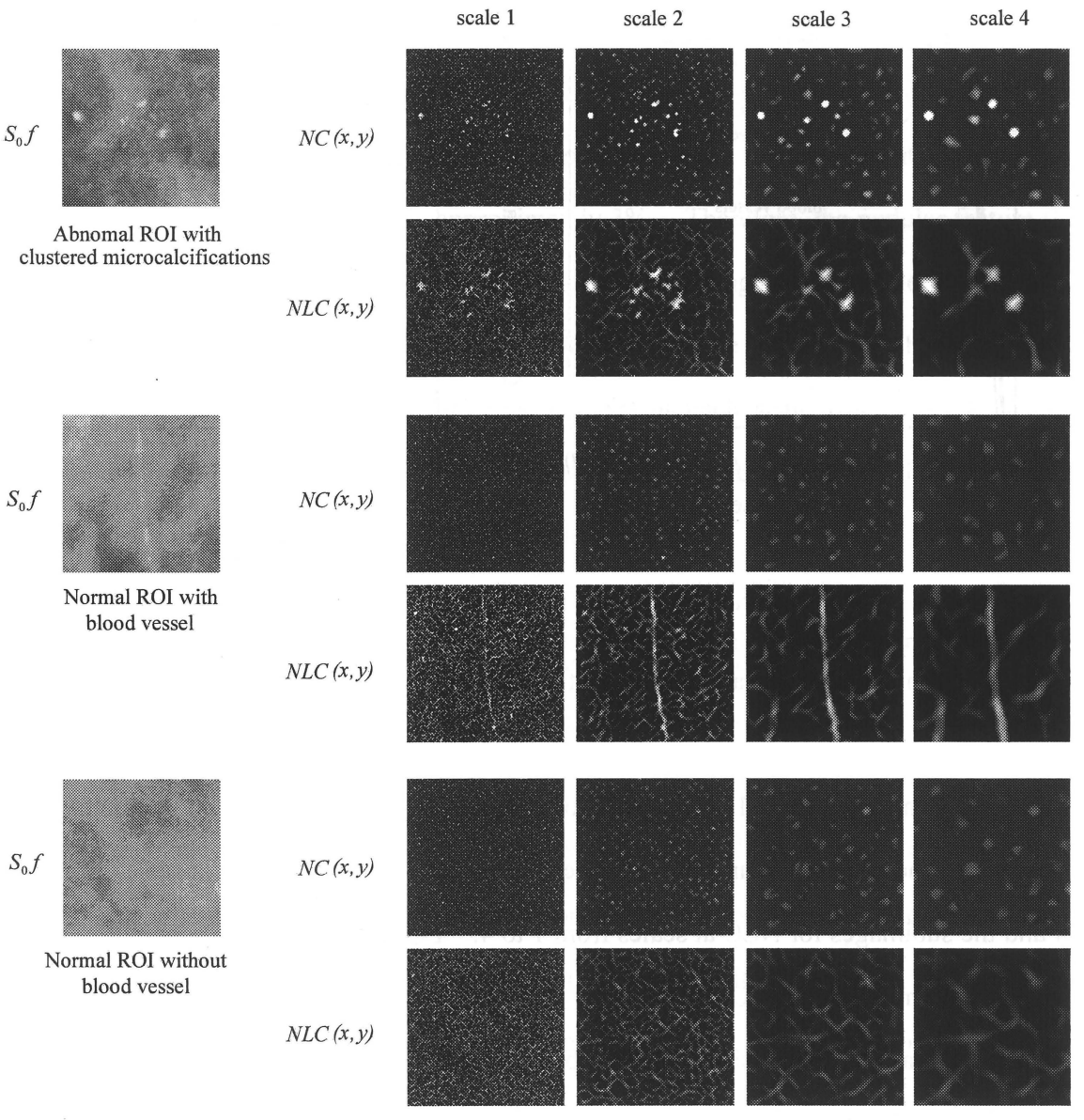


Figure 2.12: Subimages for NC and for NLC, which were obtained from an abnormal ROI with clustered microcalcifications, a normal ROI with blood vessels, and a normal ROI without blood vessels.

2.4 Extraction of Features for Detection of Clustered Microcalcifications

We determined eight features for distinguishing among abnormal ROIs with clustered microcalcifications and two different types of normal ROIs without clustered

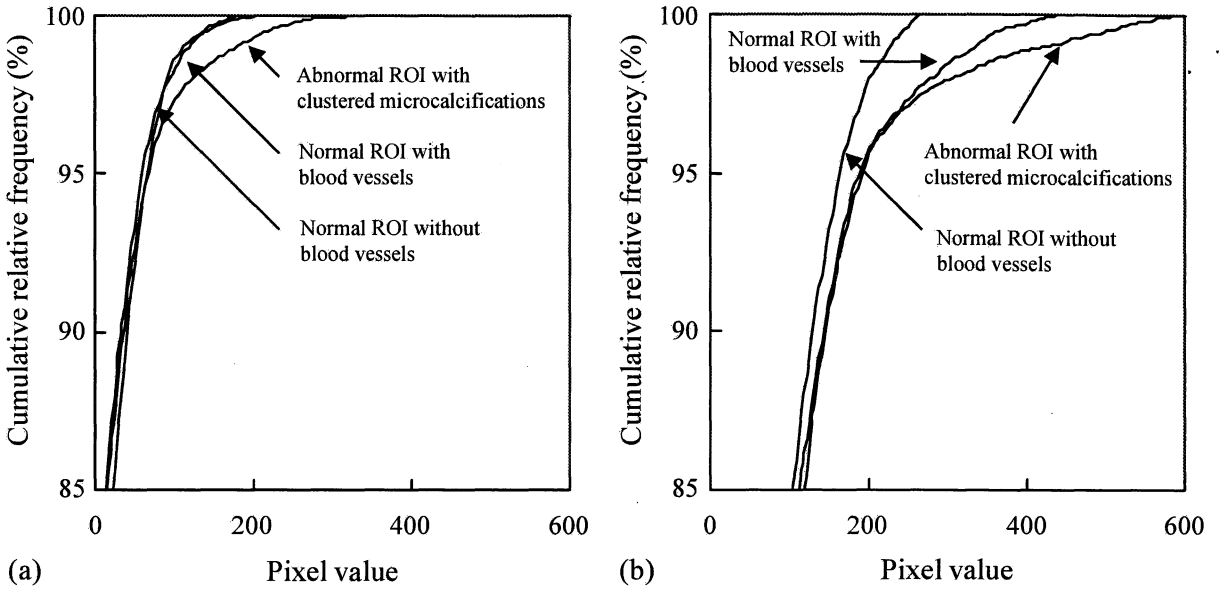


Figure 2.13: Cumulative histograms for three types of ROI. (a) Cumulative histogram obtained from the subimage adding NCs at scales from 1 to 4 and (b) cumulative histogram obtained from the subimage adding NLCs at scales from 1 to 4.

microcalcifications (normal ROIs with blood vessels and normal ROIs without blood vessels). These eight features are extracted from the subimages for NC at scales from 1 to 4 and the subimages for NLC at scales from 1 to 4. Fig. 2.12 shows the subimages for NC and the subimages for NLC, which were obtained from an abnormal ROI with clustered microcalcifications, a normal ROI with blood vessels, and a normal ROI without blood vessels. As shown in the subimages for NC, some pixel values for the abnormal ROI were higher than those for the two normal ROIs. As shown in the subimages for NLC, some pixel values for the normal ROI without blood vessels were lower than those for the abnormal ROI and the normal ROI with blood vessels. Figure 2.13 shows the cumulative histograms of the subimages for NC and the subimages for NLC. These cumulative histograms were obtained from the subimage adding NCs at scales from 1 to 4, and the subimage adding NLCs at scales from 1 to 4. Each of these cumulative histograms was also an average of the cumulative histograms of 30 abnormal ROIs with clustered

microcalcifications, 30 normal ROIs with blood vessels, and 30 normal ROIs without blood vessels which were randomly selected from the database. Large differences appeared among the three types of ROI in the pixel value higher than 97% of these cumulative histograms. This implies that the ratio of the total area of all microcalcifications included in an ROI to the area of an ROI was approximately 3%. Therefore, the nodular feature (N feature) at each scale from 1 to 4 is determined by the average value of the pixel values higher than 97% of the cumulative histograms of the subimage for NC at each scale from 1 to 4. The nodular and linear feature (NL feature) at each scale from 1 to 4 is also determined by the average value of the pixel values higher than 97% of the cumulative histograms of the subimage for NLC at each scale from 1 to 4.

2.5 Evaluation of Detection Performance

In order to detect clustered microcalcifications, we employed the Bayes discriminant function [58] for distinguishing among three classes $\omega_i \ (i = 1,2,3)$. Classes ω_1, ω_2, and ω_3 correspond to the abnormal ROI with clustered microcalcifications, the normal ROI with blood vessels, and the normal ROI without blood vessels, respectively. We first trained the Bayes discriminant function by using three different types of ROI selected from the training set. These ROIs are 300 abnormal ROIs with clustered microcalcifications, 300 normal ROIs with blood vessels, and 300 normal ROIs without blood vessels. Abnormal ROIs are selected so that the centers of clustered microcalcifications would be coincident with the centers of the ROIs. Normal ROIs are randomly selected from normal mammograms that do not include clustered microcalcifications. In each of these three classes, N features and NL features at scales from 1 to 4 which are determined from each of these ROIs are used for calculating the mean vector m_i and the covariance matrix $\sum_i$. The mean vector m_i and the covariance matrix $\sum_i$ are defined as

$$m_i = \frac{1}{n_i} \sum_{x \in \chi_i} x, \tag{16}$$

$$\Sigma_i = \frac{1}{n_i} \sum_{x \in \chi_i} (x - m_i)(x - m_i)^t. \tag{17}$$

Here, n_i and χ_i are the number of samples and the sample set in class ω_i, respectively. The probability density function for each of the three classes is assumed to be in normal distribution, whereas the prior probabilities are assumed to be equal. This is because the mammograms are clinical data. Therefore, the Bayes discriminant function for distinguishing among the three classes ω_i $(i = 1,2,3)$ is given by

$$g_i(x) = -\frac{1}{2}(x - m_i)^t \Sigma_i^{-1}(x - m_i) - \frac{1}{2}\log|\Sigma_i|, \tag{18}$$

where $|\Sigma_i|$ is the determinant. We then selected the ROIs at intervals of approximately 1 mm in the test set. In order to distinguish among the three types of ROI, eight features determined from selected ROIs are inputted to the Bayes discriminant function as the feature vector x. The Bayes discriminant function $g_i(x)$ outputs three values indicating the likelihood of each class. The class yielding the largest output value was considered to be the result of the distinction among the three types of ROI. Regions connecting the ROIs which are classified as abnormal are considered to be potential regions of clustered microcalcifications.

The free-response receiver operating characteristic (FROC) curve [59] is usually used to summarize quantitatively the detection performance of the computerized scheme. An FROC curve is a plot of the true-positive fraction (TP) achieved by a computerized detection method versus the average number of false positives (FPs) per image varied over the continuum of the decision threshold. An FROC curve provides a comprehensive summary of the trade-off between detection sensitivity and specificity. However, it is not easy to calculate an FROC curve in the present context because the Bayes discriminant function outputted three values indicating the likelihood of each class. Therefore, we first

multiply the output value $g_1(x)$, indicating the likelihood of an ROI with a clustered microcalcifications, by various coefficients before comparing $g_1(x)$, $g_2(x)$, and $g_3(x)$. The ROI is then considered as abnormal when $g_1(x)$ is the highest value of the three output values. When a center of the region connecting the ROIs which are classified as abnormal is within a true cluster determined by an experienced radiologist, this region is considered to have been "truly" detected. When a center of the connecting region is not within a true cluster, this region is considered a false positive. In this study, the coefficient is varied from 0.5 to 1.5 by the unit of 0.02. This range of the coefficient is determined empirically to detect all clustered microcalcifications in the test set. By using this method, we obtain the relationship between TP and FPs per image varied over the continuum of the coefficient.

2.6 Results and Discussion

2.6.1 Usefulness of Eight Features

In order to investigate the usefulness of the used features for distinguishing among three classes, we showed the relationship between N features and NL features at scales from 1 to 4 as shown in Fig. 2.14. These N features and NL features were determined from 300 abnormal ROIs with clustered microcalcifications, 300 normal ROIs with blood vessels, and 300 normal ROIs without blood vessels which were used for training the Bayes discriminant function in the section 2.5. The N features for abnormal ROIs with clustered microcalcifications at all scales tended to be larger than those for normal ROIs with and without blood vessels because individual microcalcifications were generally nodular in structure. The NL features for normal ROIs without blood vessels at all scales tended to be smaller than those for abnormal ROIs with clustered microcalcifications and normal ROIs with blood vessels. Although N features and NL features for abnormal ROIs with clustered microcalcifications decreased greatly at scales between 2 and 3, those for normal

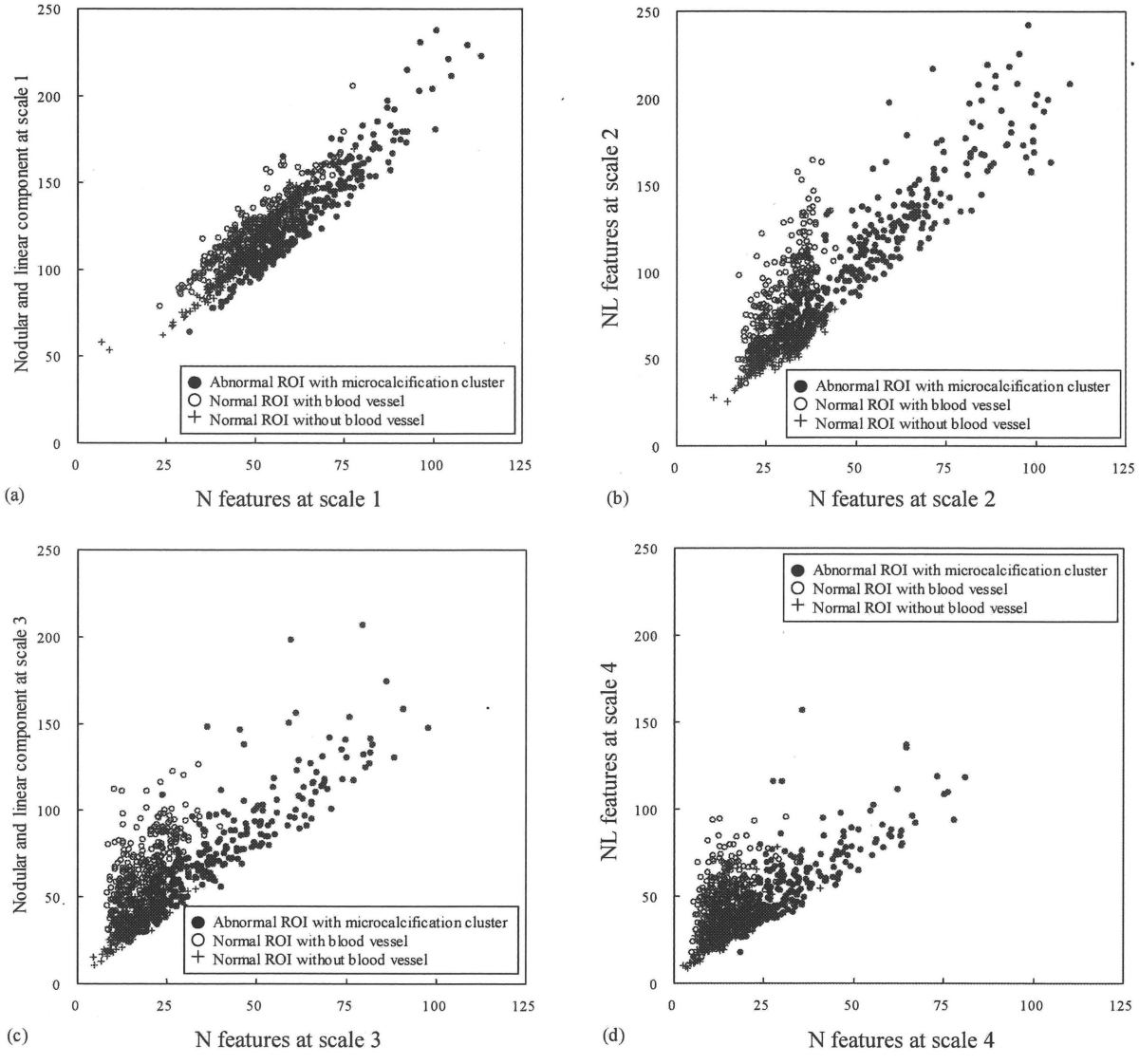


Figure 2.14: Relationship between N features and NL features (a) at scale 1, (b) at scale 2, (c) at scale 3, and (d) at scale 4.

ROIs with and without blood vessels did not decrease greatly at any of the scales. This appears to imply that most of the microcalcifications are of a size corresponding to the width of the filter at scale 2. The difference among the three types of ROIs was small in N features and NL features at scale 4. This implies that N features and NL features at scale 5 might not be useful for distinguishing among the three types of ROI. For detecting clustered microcalcifications, many investigators have developed computerized schemes

Table 2.1: Tests for univariate equality means.

	Wilk's lambda	F value	p value
N feature at scale 1	0.708	185.20	$p < .001$
N feature at scale 2	0.521	412.78	$p < .001$
N feature at scale 3	0.584	319.56	$p < .001$
N feature at scale 4	0.602	282.23	$p < .001$
NL feature at scale 1	0.832	135.65	$p < .001$
NL feature at scale 2	0.575	372.22	$p < .001$
NL feature at scale 3	0.665	258.09	$p < .001$
NL feature at scale 4	0.727	206.51	$p < .001$

based on the distinction between clustered microcalcifications and normal tissue. However, in the relationship between N features and NL features, large differences appeared between normal ROIs with blood vessels and normal ROIs without blood vessels. Therefore, we consider that it might be possible to detect clustered microcalcifications more accurately by distinguishing among the abnormal ROIs with clustered microcalcifications, the normal ROIs with blood vessels, and the normal ROIs without blood vessels. In addition, the detection performance of the computerized scheme using N features and NL features at different scales might be higher than those using an N feature and an NL feature at one scale, because the differences appeared among the three types of ROI at each scale.

Table 2.1 shows the results of tests for univariate equality of group means. These results were calculated by using the objective features in Fig. 2.14. Wilk's lambda [60] for the N feature at scale 2 was smaller than that for any other feature. The F-value [60] for the N feature at scale 2 was also larger than that for any other feature. This result would indicate that the N feature at scale 2 made a larger contribution for distinguishing among the

Table 2.2: AUCs for each individual feature in the discriminant of abnormal and normal ROIs.

	AUC
N feature at scale 1	0.841
N feature at scale 2	0.919
N feature at scale 3	0.916
N feature at scale 4	0.897
NL feature at scale 1	0.687
NL feature at scale 2	0.850
NL feature at scale 3	0.728
NL feature at scale 4	0.722

three types of ROIs. The NL feature at scale 1 made some contributions to the classification. This was because the NL feature at scale 1 was influenced strongly by noise. However, the contribution of the NL feature at scale 1 reached the level of statistical significance ($p < .001$). Therefore, the eight features were statistically significant for distinguishing among the three types of ROIs.

Table 2.2 shows the areas under the receivers operating characteristic curve (AUCs) [59] of individual features for distinguishing between abnormal ROIs and normal ROIs. This AUC was calculated by use of the features in Fig. 2.14. The AUC for the N feature at scale 2 was larger than that for any other feature. This result would indicate that the N feature at scale 2 made a larger contribution for distinguishing between the two types of ROIs. The AUCs for NL features were smaller than those for N features because both abnormal ROIs with clustered microcalcifications and normal ROIs with blood vessels included linear components. Therefore, NL features might be more effective when they are used for distinguishing the three types of ROIs.

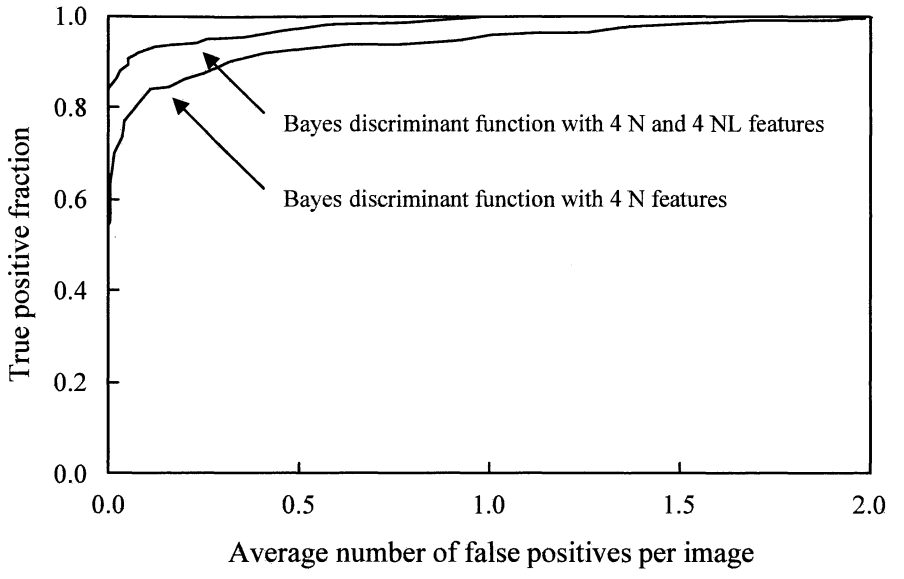


Figure 2.15: Comparison of the relationships between TP and FPs obtained by the Bayes discriminant functions with eight features and with four features.

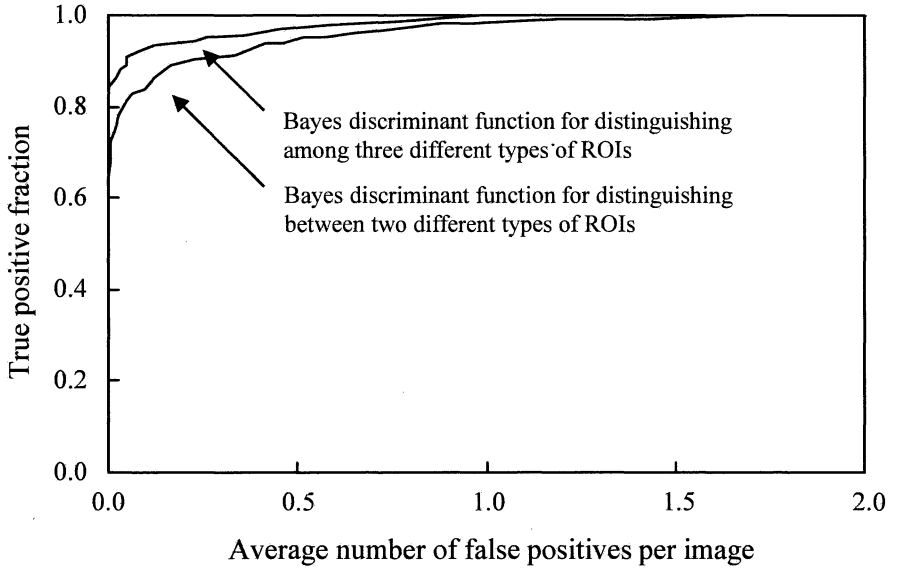


Figure 2.16: Comparison of the relationships between TP and FPs obtained by the Bayes discriminant functions for distinguishing among three different types of ROI and between two different types of ROI.

2.6.2 Performance of Detection

Figure 2.15 shows the relationship between TP and FPs obtained by applying the Bayes

discriminant function with eight features to 600 mammograms in the test set. For detecting clustered microcalcifications, many investigators have developed computerized schemes using only the objective features related to the nodular structure. Therefore, in order to investigate the usefulness of NL features, the relationship between TP and FPs for the Bayes discriminant function with four features, i.e., all features except the NL feature, is also shown in Fig. 2.15. The detection performance of the Bayes discriminant function with eight features was much higher than that of the Bayes discriminant function with the four N features. The points at which the blood vessels intersect tended to become nodular in structure. Therefore, they were detected as FP candidates in the other computerized schemes for the detection of clustered microcalcifications. This result would indicate that the number of these FP candidates was reduced by identifying the ROI with blood vessels. Many investigators have also developed computerized schemes based on the distinction between clustered microcalcifications and normal tissue. Therefore, in order to investigate the usefulness of distinguishing among three different types of ROIs, the relationship between TP and FPs for the Bayes discriminant function for distinguishing between abnormal ROIs and normal ROIs is also shown in Fig. 2.16. The detection performance of the Bayes discriminant function for distinguishing among three types of ROIs was higher than that of the Bayes discriminant function for distinguishing between two types of ROIs. This result indicates that the detection performance was improved by distinguishing among three different types of ROIs.

The proposed detection method based on the Bayes discriminant function with eight features for distinguishing among three types of ROIs identified 310 of the 310 clustered microcalcifications in the test set, yielding a sensitivity of 100.0% and a FP rate of 0.98 per mammogram.

CHAPTER 3
COMPUTERIZED CLASSIFICATION METHOD FOR IDENTIFYING HISTOLOGICAL CLASSIFICATIONS OF CLUSTERED MICROCALCIFICATIONS

Making clinical decisions for biopsy or follow-up on clustered microcalcifications by taking into account possible histological classifications on magnification mammograms would reduce the number of unnecessary biopsies [7, 50]. For example, patients with clustered microcalcifications associated with invasive carcinoma that may metastasize to other organs must undergo biopsy immediately. Patients associated with noninvasive carcinoma of the comedo type that grows rapidly must undergo biopsy or immediate follow-up at a very short interval of one month. Patients associated with noninvasive carcinoma of the noncomedo type with lower risk than the comedo type should also have follow-up at a short interval of three months. Patients associated with mastopathy and fibroadenoma of benign breast lesions should have follow-up at a relatively long interval of six months. Therefore, the computerized analysis for estimating the likelihood of histological classifications on clustered microcalcifications would be helpful to radiologists for their decisions on patient management.

In this chapter, we develop a computerized classification method for histological classification of clustered microcalcifications in order to assist radiologists' interpretation as a "second opinion." There are differences in both the image features and the growth speed among histological classifications of clustered microcalcifications. In the computerized classification method, therefore, we extract six objective features from clustered microcalcifications on each of follow-up magnification mammograms (i.e. both current and previous magnification mammograms). We show that the differences in growth speed among histological classifications are reflected in the six objective features used in the

proposed classification method. We then evaluate the potential contribution of the follow-up mammograms in the computerized classification method for estimating the likelihood of histological classifications on clustered microcalcifications.

3.1 Materials

Our database consists of current and previous magnification mammograms obtained from 93 patients before and after three-month follow-up examination at the Breastopia Namba Hospital, Miyazaki, Japan. It includes 55 malignant clustered microcalcifications (11 invasive carcinomas, 19 noninvasive carcinomas of the comedo type, and 25 noninvasive carcinomas of the noncomedo type) and 38 benign clustered microcalcifications (23 mastopathies and 15 fibroadenomas). The histological classification of all clustered microcalcifications was proven by stereotaxic core needle biopsy after a three-month follow-up examination. Informed consent was obtained for the research use of each patient's mammograms. Figure 3.1 shows a current magnification mammogram and the corresponding previous magnification mammogram in each of the five histological classifications. It should be noted that the change in visual image features of the clustered microcalcifications in invasive carcinoma and noninvasive carcinoma of the comedo type are larger than those in the other lesions.

The magnification mammograms were acquired with a Kodak MinR-2000 screen/film system. The magnification factor of the magnification mammograms was 1.8. The mammographic x-ray system included an x-ray tube with a 0.1 mm focal spot and a molybdenum anode, a 0.03-mm-thick molybdenum filter, and a 5:1 reciprocating grid. These mammograms were digitized to a 512x512 matrix size with a 0.0275 mm pixel size and a 12-bit gray scale by use of an EPSON ES-8000 digitizer (optical resolution 800x1600 dpi, optical density range 0.0-3.3D).

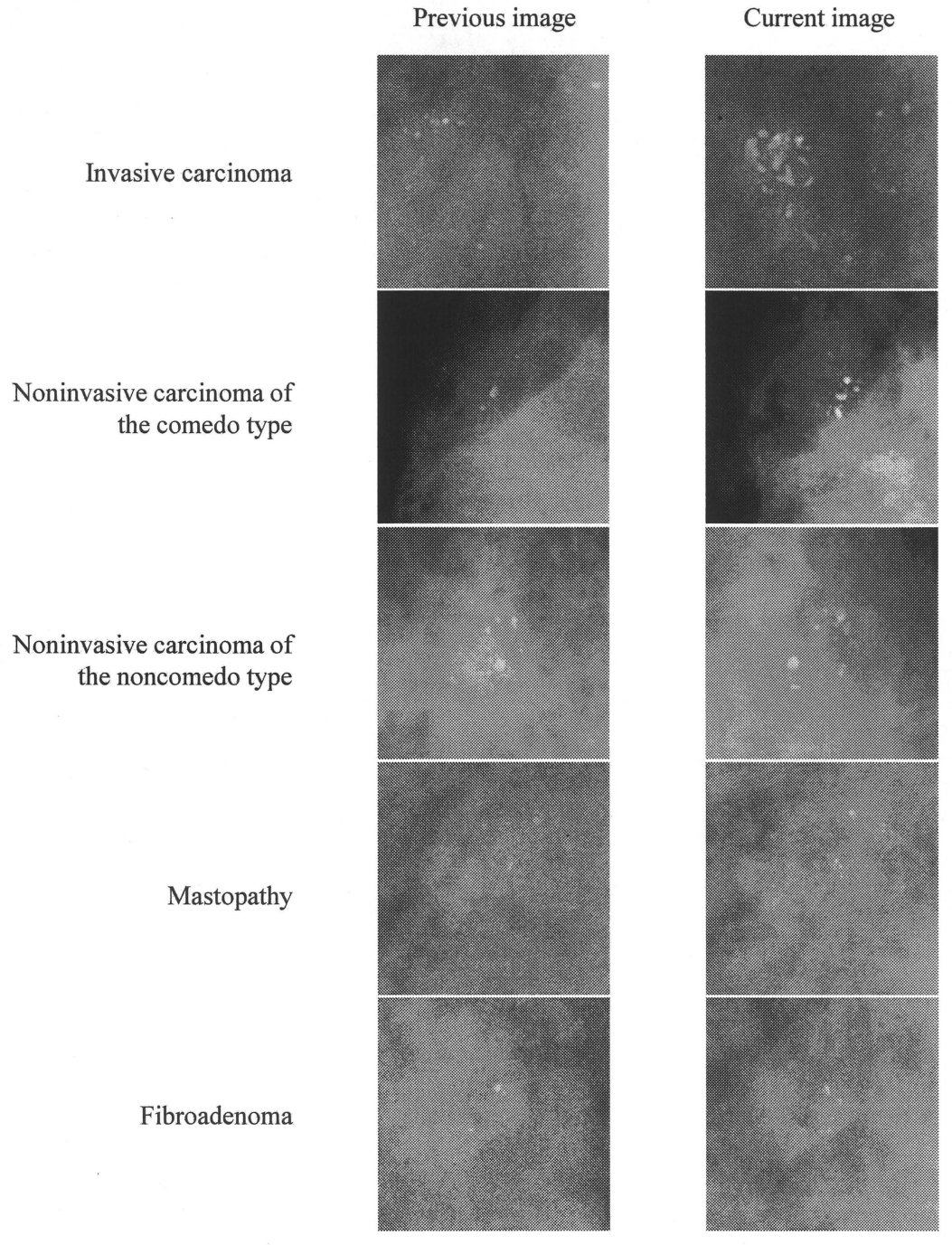
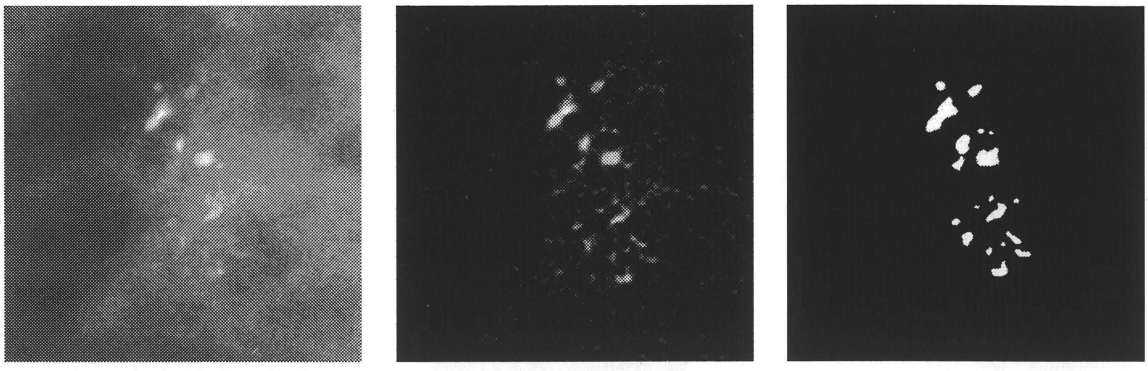


Figure 3.1: Example of current images and the corresponding previous image in each histological classification.

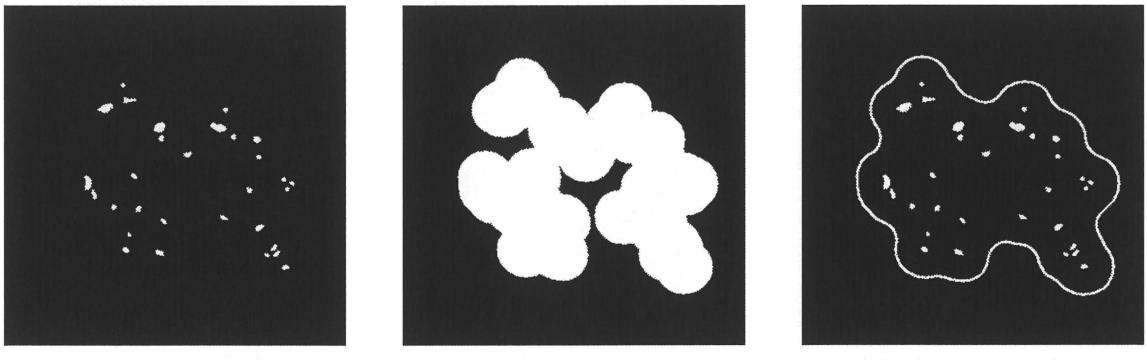
3.2 Segmentation of Microcalcifications and Definition of Cluster Margin

For segmentation of individual microcalcifications within a cluster on magnification



(a) Original image (b) Enhanced image for calcifications (c) Segmented image for calcifications

Figure 3.2: Illustration of microcalcification segmentation by a novel filter bank and a threshold technique. (a) Original image, (b) enhanced image for microcalcifications, and (c) segmented image for microcalcifications.



(a) Segmented calcifications (b) Candidate for clustered microcalcifications (c) Cluster margin

Figure 3.3: Illustration of definition of cluster margin, (a) segmented microcalcifications, (b) candidate for the cluster margin, and (c) cluster margin.

mammograms, we first enhance the microcalcifications by use of the novel filter bank in section 2.3. The microcalcifications are enhanced while maintaining their shape by use of this filter bank, as shown in Fig. 3.2(b). A gray-level thresholding technique [61] is then applied to the enhanced image, as shown in Fig. 3.2(c). The binary image for microcalcifications is obtained at a threshold level of a 600 pixel value which is determined empirically to segment all microcalcifications in 186 magnification mammograms. All

detected regions at the threshold level of 600 are considered to be microcalcifications.

In order to obtain information about the shape of the distribution of clustered microcalcifications, the margin of a cluster is automatically determined by using a computer. We first draw circles with a diameter of 20 pixels at the center of gravity of each microcalcification, as shown in Fig. 3.3(b). The center of gravity of each microcalcification is calculated from the region of the segmented microcalcifications. The region connecting the circles is considered to be a candidate for a cluster margin. However, some circles for each microcalcification within a cluster were not connected, when the diameter of the circles was too small. Therefore, the diameter is increased from 20 to 60 pixels until all circles within a cluster are connected, thus yielding a candidate for a cluster margin. The shape of the candidate for the cluster margin strongly depends on the locations of individual microcalcifications. The shape of the margin cannot be estimated accurately because there are some indentations in the candidate for the cluster margin, as shown in Fig. 3.3(b). We apply a binary morphologic closing operator [62] to the candidate for smoothing the shape of the cluster margin. The structure element for the binary morphologic closing operator is given by the circle with half the diameter of the circle which is used for determining the candidate for the cluster margin. The edge of the smoothed binary image is determined as the cluster margin.

Although, in this study, we use only magnification mammograms acquired with MLO positioning, there would be a small position variation at each acquirement of the magnification mammograms. Therefore, even if each image contains the same clustered microcalcification, visual image features may be slightly different in each magnification mammogram. This is an important issue in the consistency between objective features of clustered microcalcifications extracted from the current and the previous magnification mammograms. However, in our previous study, we confirmed that the objective features of clustered microcalcification extracted from MLO magnification mammogram were

nearly similar to those extracted from CC magnification mammogram [63]. Therefore, we consider that the extracted features are little influenced from this position variation.

3.3 Extraction of Six Objective Features

Clustered microcalcifications associated with invasive carcinoma tend to be very heterogeneous in terms of microcalcifications' sizes and pixel values [7, 8, 50]. Their shape is generally a rodlike/branching pattern. The edges of their distribution are commonly irregular, because invasive carcinoma invades interstitial tissue. Clustered microcalcifications associated with noninvasive carcinoma also tend to be heterogeneous. However, there are some differences in the shape between the comedo type and the noncomedo type of noninvasive carcinoma. Microcalcifications for the comedo type tend to be in a rodlike/branching pattern, whereas those for the noncomedo type tend to be in a linear/branching and granular/punctate pattern [7, 64, 65]. The distribution of microcalcifications of the comedo type is likely to become linear and branching pattern in the direction toward the nipple, because they are guided by the course of the duct [7, 8, 50]. Clustered microcalcifications associated with a benign lesion are commonly uniform in microcalcifications' size and pixel values; they tend to have round patterns.

We selected six objective features on clustered microcalcifications to distinguish among five different types of histological classifications. These objective features were: (1) the variation in the sizes of microcalcifications within a cluster, (2) the variation in pixel values of microcalcifications within a cluster, (3) the shape irregularity of microcalcifications within a cluster, (4) the extent of linear and branching distribution of microcalcifications, (5) the distribution of microcalcifications in the direction toward the nipple, (6) the number of microcalcifications within a cluster. These features are frequently used for describing microcalcifications.

3.3.1 Variation in Sizes of Microcalcifications

The variation in the sizes of microcalcifications within a cluster is determined by the relative standard deviation in the areas of microcalcifications within a cluster. The area of each microcalcification is determined by the number of pixels within the segmented microcalcification. Although standard deviation is frequently used for quantifying the variation in sizes, standard deviation becomes large when the sizes of the microcalcifications within a cluster are large. Because the size information of microcalcifications alone is not directly related to the likelihood of malignancy [7, 8, 50], we use the relative standard deviation to quantify the variation in the sizes of microcalcifications within a cluster. In order to demonstrate the usefulness of the relative standard deviation, we compared the relative standard deviation and the standard deviation in determining the variation in the sizes of microcalcifications, as shown in Fig. 3.4, where these values were normalized by using the mean value and the standard deviation of all cases in the database. When the standard deviation was used, the distributions for malignant cases extensively overlapped with those for benign cases. Thus, it is very difficult to distinguish between them. However, when the relative standard deviation was used, the variations in the sizes for malignant cases tended to be greater than those for benign cases.

3.3.2 Variation in Pixel Values of Microcalcifications

The variation in pixel values of microcalcifications within a cluster is determined by the standard deviation for the pixel values of microcalcifications. The pixel value of each microcalcification is defined by the mean value of the five largest pixel values on the segmented microcalcification in the original image. Because the pixel values of microcalcifications is likely to be related to the likelihood of malignancy [7, 8, 50], we use the standard deviation to quantify the variation in pixel values of microcalcifications.

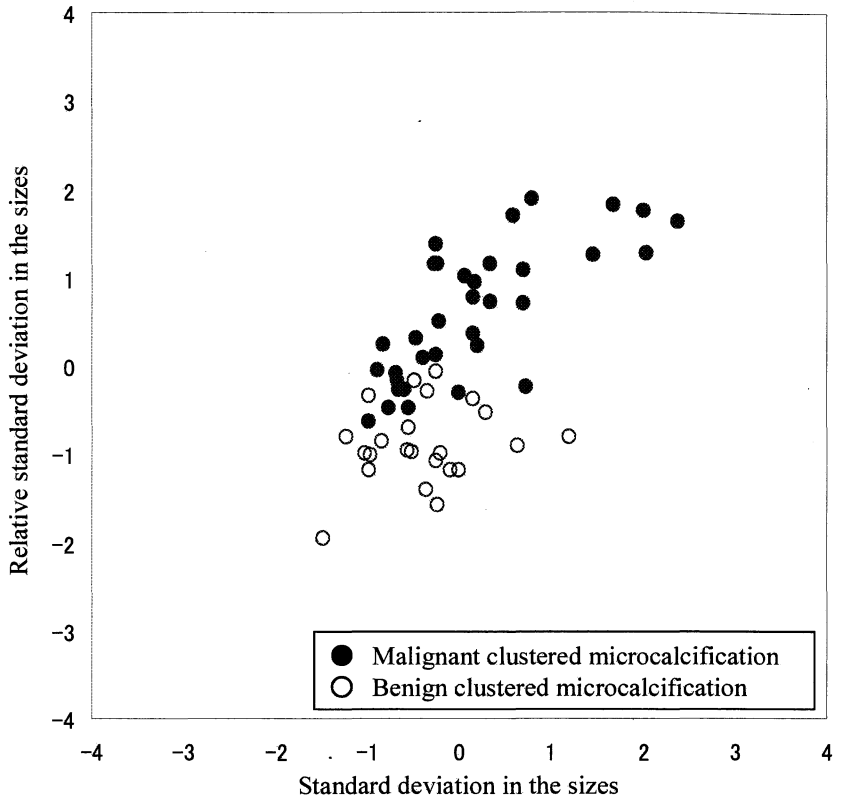


Figure 3.4: Comparison between the relative standard deviation and the standard deviation in determining the variation in the sizes of microcalcifications.

Figure 3.5 shows a comparison between the standard deviation and the relative standard deviation in determining the variation in the pixel values of microcalcifications. When the standard deviation was used, the variations in the pixel values for malignant cases tended to be larger than those for benign cases.

3.3.3 Shape Irregularity of Microcalcifications

In order to define the shape irregularity of microcalcifications within a cluster, we employed two kinds of irregularity indices for each microcalcification. One irregularity index for each microcalcification is defined by the standard deviation of the 16 shape factors, as shown in Fig. 3.6. Sixteen shape factors consist of 8 minimum distances and 8 maximum

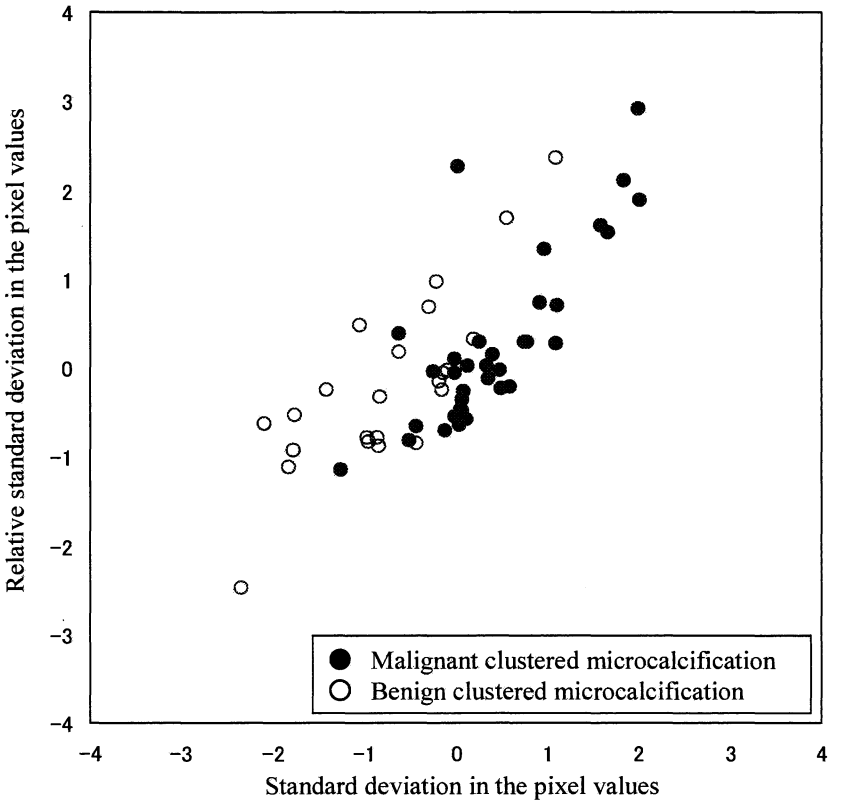
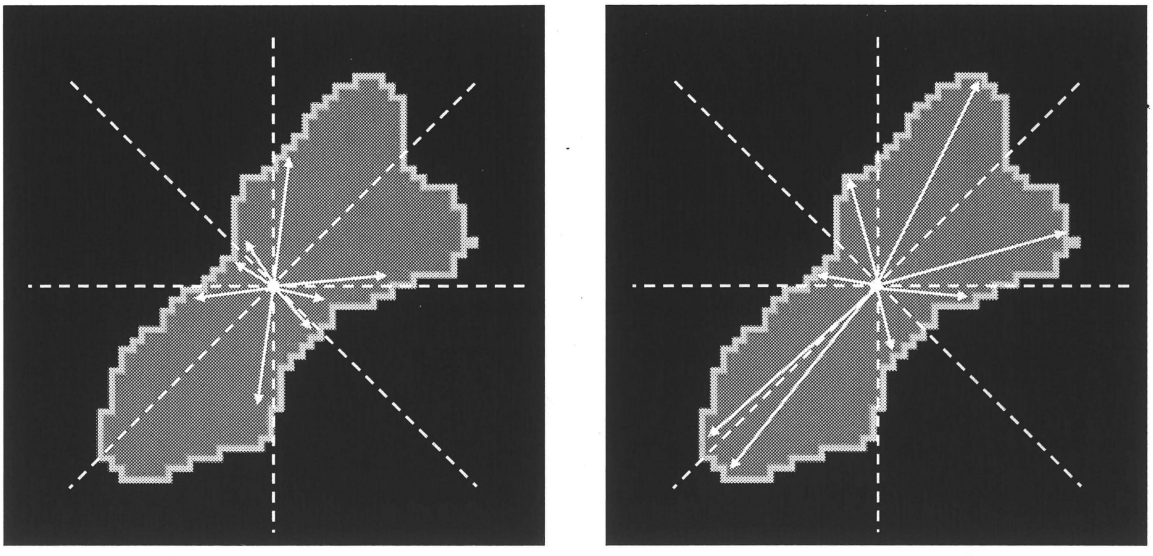


Figure 3.5: Comparison between the standard deviation and the relative standard deviation in determining the variation in pixel values of microcalcifications.

distances between the center of the segmented microcalcification and the edges of the segmented microcalcification. The minimum distance and the maximum distance are obtained in each of 8 regions located at intervals of 45 degrees. The standard deviation of these 16 shape factors is used for identifying irregular microcalcifications. For a round microcalcification, all 16 shape factors would have similar values, and thus the standard deviation would be small. For an irregular (linear and branching pattern) microcalcification, some of the 16 shape factors have large values, whereas others have small values; therefore, the standard deviation would be large. Another irregularity index of each microcalcification was evaluated by use of the degree of irregularity (1-P/N; P = perimeter of the circle with the same area as the microcalcification, N = length of the



(a) Minimum distances (b) Maximum distances

Figure 3.6: Illustrations of the definition of the shape factors for individual microcalcifications. (a) Eight shape factors are the minimum lengths of the distance, and (b) eight shape factors are the maximum lengths of the distance between the center-of-microcalcification pixel and the edges of microcalcification in eight regions separated by forty-five degrees.

microcalcification outline) which is generally used for quantifying the irregularity in the shape [66].

The shape irregularity of microcalcifications within a cluster is defined by the mean value of the five largest irregularity indices of individual microcalcifications within a cluster. Figure 3.7 shows a comparison of two shape irregularities of microcalcifications within a cluster, which was obtained from the standard deviation of the 16 shape factors and the degree of irregularity. When the standard deviation of the 16 shape factors was used as a basis for the irregularity index, the shape irregularities for malignant cases tended to be larger than those for benign cases. Because the length of the microcalcification outline could not be determined accurately for small microcalcifications, the reliability for the degree of irregularity is somewhat uncertain. Therefore, we use only the standard

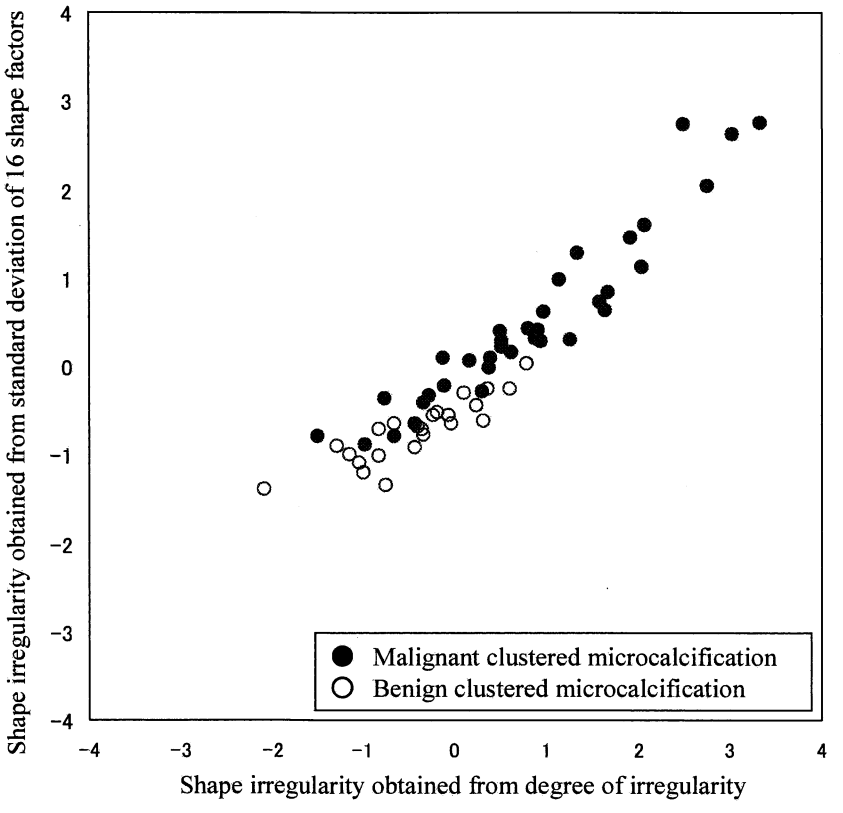
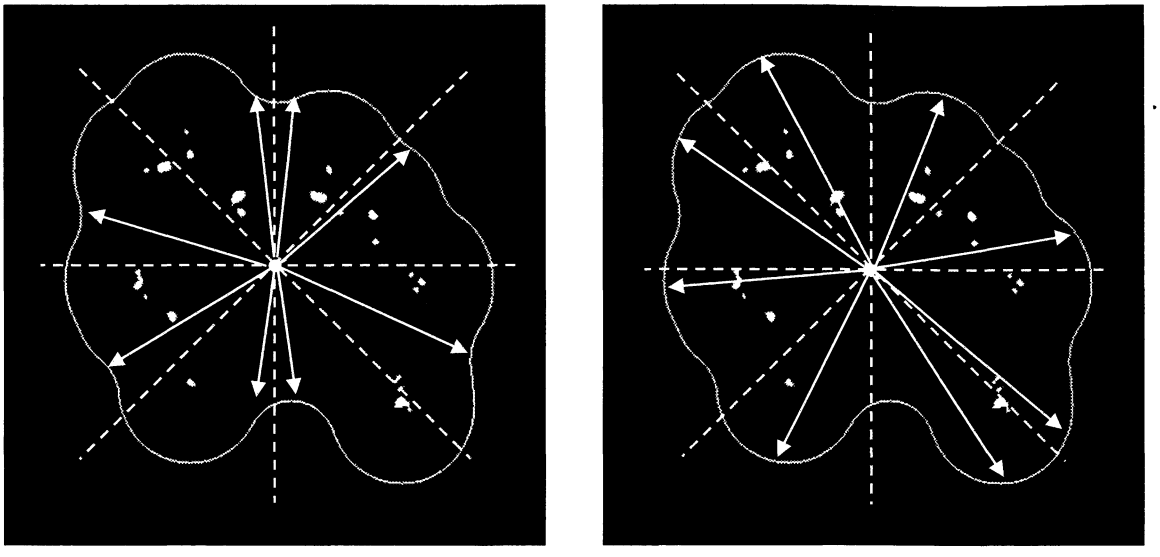


Figure 3.7: Comparison of two shape irregularities of microcalcifications within a cluster.

deviation of the 16 shape factors for evaluation of the irregularity index of each microcalcification and the shape irregularity of clustered microcalcifications in the following sections.

3.3.4 Extent of Linear and Branching Distribution of Microcalcifications

The extent of the linear and branching distribution of microcalcifications cluster is evaluated by use of the standard deviation of the 16 shape factors. Note that the 16 shape factors are applied to the cluster margin defined in section 3.2, instead of individual microcalcifications as described in section 3.3.3. The 16 shape factors consist of 8 minimum distances and 8 maximum distances between the center of a cluster and the edges of a cluster, as shown in Fig. 3.8. The minimum distance and the maximum distance are



(a) Minimum distances (b) Maximum distances

Figure 3.8: Illustrations of the definition of the shape factors for a cluster margin. (a) Eight shape factors are the minimum lengths of the distance, and (b) eight shape factors are the maximum lengths of the distance between the center of the cluster and the edges of the cluster in eight regions separated by forty-five degrees.

obtained from each of 8 regions located at intervals of 45 degrees in the cluster margin.

3.3.5 Distribution of Microcalcifications in Direction toward Nipple

To determine the measure for the distribution of microcalcifications in the direction toward the nipple, we define the average distances to a main straight line and a sub-straight line as shown in Fig. 3.9. The main straight line is drawn from the center of the nipple to the center of a cluster. The average distance of microcalcifications to the main line is given by the average distance from each center of microcalcifications to the main straight line. Note that the unit of this distance is the pixel of the digitized mammogram. The sub-straight line is a line perpendicular to the main straight line at the center-of-cluster pixel. The average distance of microcalcifications to the sub-straight line is given by the average distance from each center of microcalcifications to the sub-straight line. The measure for

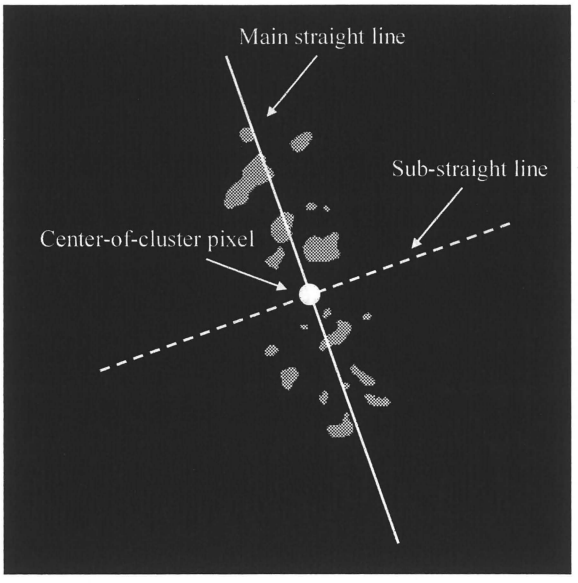


Figure 3.9: Illustration of a main straight line and a sub-straight line for determining the distribution of microcalcifications in the direction toward the nipple. The main straight line is drawn from center of nipple to center of cluster. The sub-straight line is perpendicular to the main straight line at the center-of-cluster pixel.

the distribution of microcalcifications in the direction toward the nipple is determined by the ratio of the average distance for the sub-straight line to the average distance for the main straight line. If the distribution of microcalcifications is extended toward the nipple, the average distance for the main straight line would have a small value, whereas the average distance for the sub-straight line would have a large value. Therefore, the ratio of the average distance for the sub-straight line to the average distance for the main straight line would be large.

3.3.6 Number of Microcalcifications

The number of microcalcifications within a cluster is determined by the number of the segmented microcalcifications within a cluster in section 3.2.

3.4 Determination of Histological Classification

For determining the histological classifications of clustered microcalcifications, the nearest neighbor case is identified by the Euclidean distance in the previous and current feature-space which consisted of six objective features obtained from the previous magnification mammogram (previous features) and six objective features obtained from the current magnification mammogram (current features). The histological classification of an unknown new case in question is assumed to be the same as that of the nearest neighbor case which has the shortest Euclidean distance in our database: the validity of this assumption is examined in this study.

3.5 Results and Discussion

3.5.1 Usefulness of Six Objective Features

Figure 3.10 shows comparisons between the current features and the previous features in each of six objective features. The six objective features in malignant lesions tended to increase over the three-month follow-up examination, whereas those in benign lesions tended to be almost constant or decrease. These results were consistent with the differences in growth speed among histological classifications in clinical experience [64, 65]. The findings of the five objective features excluding the number of microcalcifications also corresponded to the radiologic finding [7, 50] of clustered microcalcifications in each histological classification. Although, on the other hand, clustered microcalcifications associated with malignant lesions tend to have many microcalcifications, this radiologic finding did not appear in this study. This cause might be because we used only clustered microcalcifications in which the extent of the distribution was small (i.e. the extent was less than 1.5 cm x 1.5 cm).

Table 3.1 shows the results of tests for univariate equality of group means for each objective feature in each of the current features and the previous features. The Wilk's

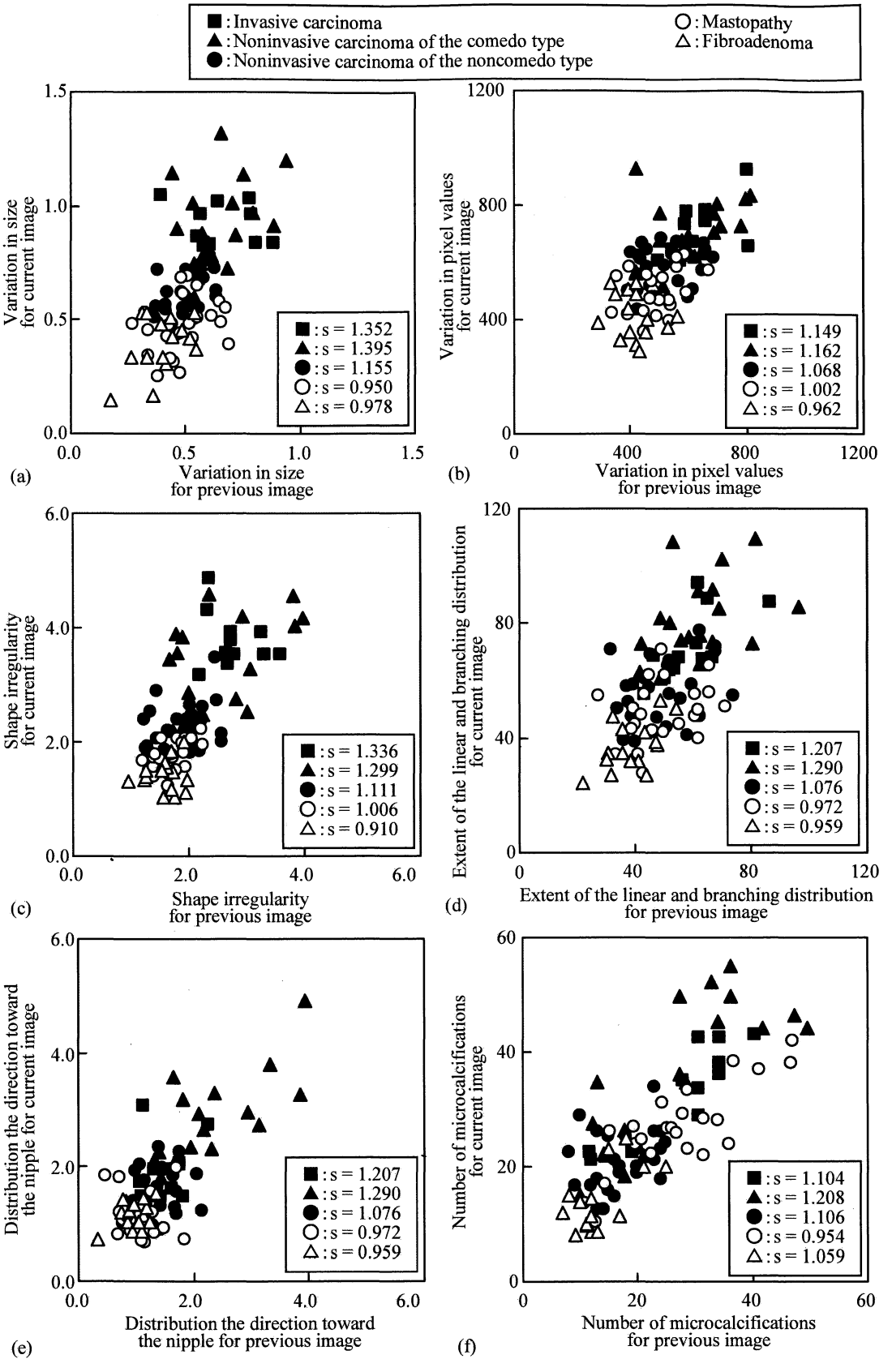


Figure 3.10: Comparisons between the current features and the previous features in (a) variation in size, (b) variation in pixel values, (c) the shape irregularity, (d) extent of linear and branching distribution, (e) distribution in direction toward the nipple. and (f) number of microcalcifications. "s" presents the slope of the approximation straight line for each of histological classifications.

Table 3.1: Tests for univariate equality of group means for each objective feature in each of the current features and the previous features.

Objective features	Previous features (n = 93)			Current features (n = 93)			Correlation coefficient
	Wilk's lambda	F value	*p* value	Wilk's lambda	F value	*p* value	
Variation in size	0.68	8.26	< 0.001	0.52	21.03	< 0.001	0.34
Variation in pixel values	0.70	7.16	< 0.001	0.60	14.39	< 0.001	0.32
Shape irregularity	0.60	14.68	< 0.001	0.48	24.75	< 0.001	0.35
Extent of linear and branching distribution	0.73	5.29	< 0.001	0.62	12.86	< 0.001	0.40
Distribution in direction toward the nipple	0.65	10.54	< 0.001	0.59	15.17	< 0.001	0.51
Number of microcalcifications	0.74	5.23	< 0.001	0.69	7.64	< 0.001	0.75

lambdas [60] for the current features were smaller than those for the previous features, whereas the F values [60] for the current features were larger than those for the previous features. These results would indicate that the difference in each objective feature among the five histological classifications became larger over the three-month follow-up examination. For the number of microcalcifications, the Wilk's lambdas were larger than any other objective features, and the F values were smaller than the other objective features. The correlation coefficient for the number of microcalcifications was also high between the current feature and the previous feature. These results indicated that the number of microcalcifications did not have properties which would be useful to be applied for a classifier like a Linear Discriminant Analysis [67]. However, the p values for all objective features reached the level of statistical significance. Therefore, these twelve objective features were statistically significant for determining the histological classifications of clustered microcalcifications.

Table 3.2: Comparisons of the classification accuracies obtained by the Nearest neighbor criterion and the Modified Bayes discriminant function with the six current features, the six previous features, and the set of the six previous features and the six current features..

	Classification accuracy					
	With six previous features including the number of microcalcifications		With six current features including the number of microcalcifications		With six previous features and six current features	
Pathological diagnosis	MBDF	NNC	MBDF	NNC	MBDF	NNC
Invasive carcinoma (11)	6 (54.5%)	7 (63.6%)	7 (63.6%)	8 (72.7%)	10 (90.9%)	10 (90.9%)
Noninvasive carcinoma of the comedo type (19)	10 (52.6%)	11 (57.9%)	13 (68.4%)	13 (68.4%)	16 (84.2%)	17 (89.5%)
Noninvasive carcinoma of the noncomedo type (25)	16 (64.0%)	17 (68.0%)	18 (72.0%)	18 (72.0%)	21 (84.0%)	24 (96.0%)
Mastopathy (23)	15 (65.2%)	15 (65.2%)	16 (69.6%)	17 (73.9%)	17 (73.9%)	19 (82.6%)
Fibroadenoma (15)	11 (73.3%)	10 (66.7%)	11 (73.3%)	11 (73.3%)	13 (86.7%)	14 (93.3%)

3.5.2 Classification Performance

Table 3.2 shows the classification accuracies obtained by the nearest neighbor criterion (NNC) and the Modified Bayes discriminant function (MBDF: see Appendix II) [68-70] based on a leave-one-out testing method [67]. In this method, the training was carried out for all except one case in the database; the case not used for training was used for testing with the trained MBDF. This procedure was repeated until every case in our database had been used once. The MBDF can reduce the estimation error of higher-order eigenvectors in the Bayes discriminant function. For the construction of the NNCs and the MBDFs, we used the six previous features, the six current features, and the set of the six previous features and the six current features. The classification accuracies with the six current features were higher than those with the six previous features in both the NNCs and the MBDFs. These classification accuracies were improved substantially by using the set of

Table 3.3: Comparisons of the classification accuracies obtained by the Nearest neighbor criterion and the Modified Bayes discriminant function with the five current features, the five previous features, and the set of the five previous features and the five current features.

	Classification accuracy					
	With five previous features excluding the number of microcalcifications		With five current features excluding the number of microcalcifications		With five previous features and five current features	
Pathological diagnosis	MBDF	NNC	MBDF	NNC	MBDF	NNC
Invasive carcinoma (11)	6 (54.5%)	6 (54.5%)	8 (72.7%)	8 (72.7%)	9 (81.8%)	9 (81.8%)
Noninvasive carcinoma of the comedo type (19)	11 (57.9%)	11 (57.9%)	13 (68.4%)	13 (68.4%)	16 (84.2%)	17 (89.5%)
Noninvasive carcinoma of the noncomedo type (25)	15 (60.0%)	16 (64.0%)	17 (68.0%)	17 (68.0%)	19 (76.0%)	21 (84.0%)
Mastopathy (23)	15 (65.2%)	15 (65.2%)	15 (65.2%)	16 (69.6%)	17 (73.9%)	18 (78.3%)
Fibroadenoma (15)	11 (73.3%)	10 (66.7%)	11 (73.3%)	11 (73.3%)	13 (86.7%)	13 (86.7%)

the six previous features and the six current features. The classification accuracies obtained with the NNCs were higher than those obtained with the MBDF. This reason would be that the MBDF was not trained optimally because the number of samples was small compared to the number of the used features.

In order to investigate the usefulness of the number of microcalcifications, Table 3.3 shows also the classification accuracies obtained by the NNCs and the MBDFs with the five current features excluding the number of microcalcifications, the five previous features excluding the number of microcalcifications, and the set of the five previous features and the five current features. The classification accuracies with the NNCs were the same or higher as those with the MBDFs. Although both the classification accuracies with the NNCs and those with the MBDFs were also improved by adding the number of

Table 3.4: Classification results obtained by the Nearest neighbor criterion with the set of the six previous features and the six current features.

Pathological diagnosis	Computer output				
	Invasive carcinoma	Noninvasive carcinoma of comedo type	Noninvasive carcinoma of noncomedo type	Mastopathy	Fibroadenoma
Invasive carcinoma (11)	10 (90.9%)	1 (9.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Noninvasive carcinoma of comedo type (19)	1 (5.3%)	17 (89.5%)	1 (5.3%)	0 (0.0%)	0 (0.0%)
Noninvasive carcinoma of noncomedo type (25)	0 (0.0%)	1 (4.0%)	24 (96.0%)	0 (0.0%)	0 (0.0%)
Mastopathy (23)	0 (0.0%)	0 (0.0%)	2 (8.7%)	19 (82.6%)	2 (8.7%)
Fibroadenoma (15)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (6.7%)	14 (93.3%)

microcalcifications, the improvement in the NNCs tended to be larger than that in the MBDFs.

Table 3.4 shows the classification results obtained by the NNC with the set of the six previous features and the six current features. The classification accuracies of histological classifications were 90.9% (10/11) for invasive carcinoma, 89.5% (17/19) for noninvasive carcinoma of the comedo type, 96.0% (24/25) for noninvasive carcinoma of the noncomedo type, 82.6% (19/23) for mastopathy, and 93.3% (14/15) for fibroadenoma.

Figure 3.11 shows an example of identification results of a nearest neighbor case for each of the five histological classifications as shown in Fig. 3.1. It might be that the nearest neighbor cases are similar to cases in Fig. 3.1.

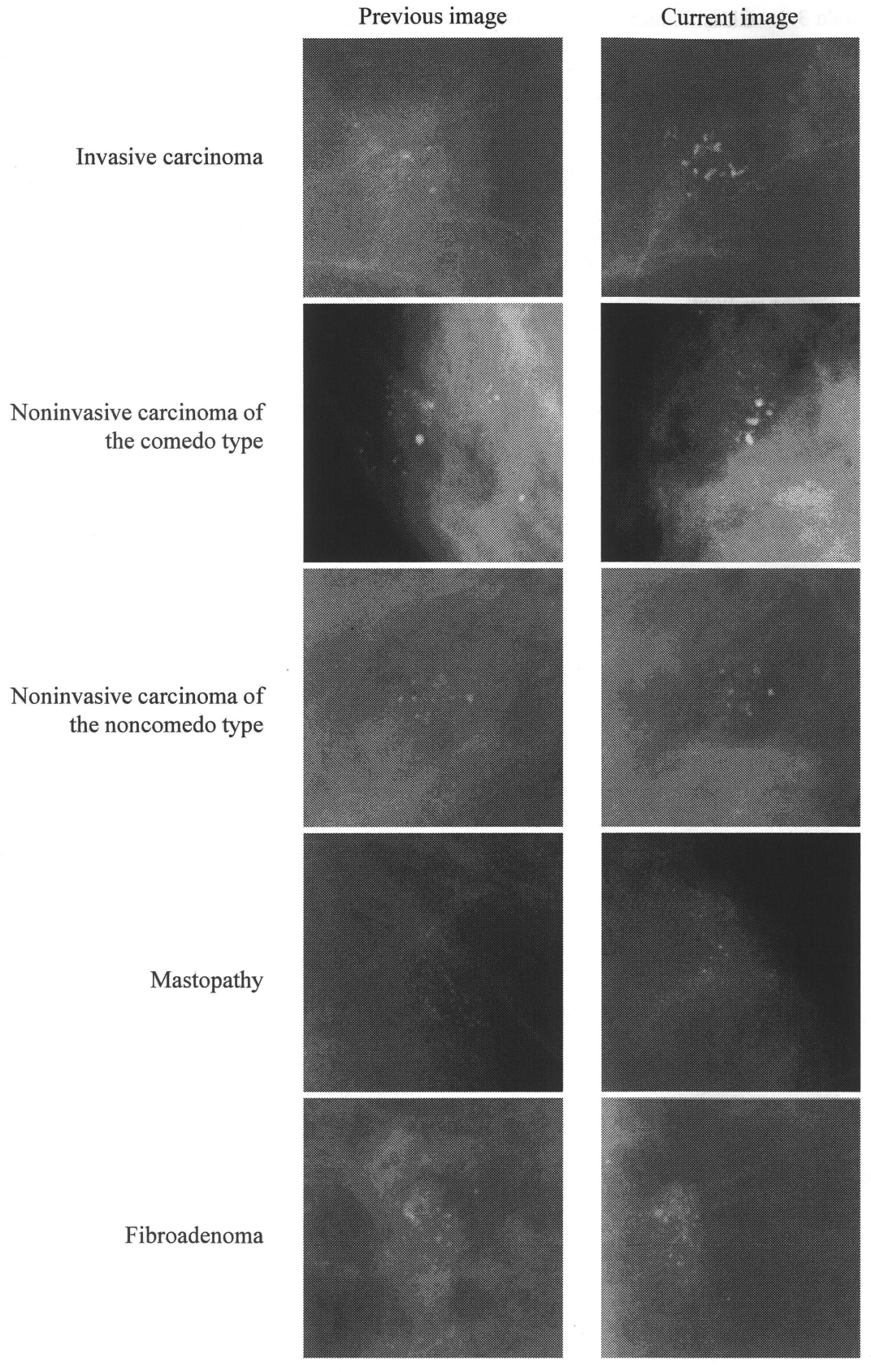


Figure 3.11: Example of identification results of a nearest neighbor case for each of the five histological classifications as shown in Fig. 3.1.

CHAPTER 4
COMPUTERIZED RETRIEVAL METHOD FOR SIMILAR IMAGES OF CLUSTERED MICROCALCIFICATIONS

In the interpretation of medical images, radiologists attempt to make diagnostic decisions based on the medical knowledge derived from viewing many clinical images over the years through education, training, and clinical practice. It is commonly known that, when a radiologist encounters a new, unknown case in daily clinical work, he/she may occasionally search for clinical images with known pathology similar to that of the unknown case by reviewing images in previous clinical cases, teaching files, and textbooks. Therefore, the presentation of similar images would be useful and would have the potential to improve radiologists' performance in the differential diagnosis of lesions in clinical images [72-75].

In order to develop a useful tool for selecting similar images to be used as a diagnostic aid, many investigators have studied content-based or feature-based image-retrieval methods [76-88]. However, these retrieval methods do not take into account radiologists' subjective impression of similarity when two images are compared. If retrieved images were not really similar to an unknown lesion visually for clinical purposes, they would not be useful for radiologists in the differential diagnosis of the unknown lesion. Therefore, Li *et al.* [89] and Muramatsu *et al.* [90-94] proposed a psychophysical similarity measure, as an image retrieval tool, which was determined by use of an artificial neural network (ANN) for learning the relationship between radiologists' subjective similarity ratings and the objective features of lesions. They showed that the correlation coefficients (r = 0.72, 0.74, and 0.71 for nodules on low-dose CT, and masses and clustered microcalcifications on mammograms, respectively) between radiologists' subjective similarity ratings and psychophysical similarity measures were greater than those (r = 0.60, 0.60, and 0.58 for nodules on low-dose CT, and masses and clustered microcalcifications on mammograms,

respectively) between radiologists' subjective similarity ratings and objective similarity measures based on the Euclidean distance in feature space that was frequently used in many studies. Their results indicated that similar images selected based on the psychophysical similarity measures would be more similar in terms of radiologists' visual perception than those selected based on the Euclidean distance in feature space. However, it appears that the psychophysical similarity measures were not highly accurate as a reliable objective similarity measure for selecting similar images, because the correlation coefficients were less than 0.80, i.e., they were not extremely high.

In this study, therefore, we investigate new objective similarity measures based on both the Euclidean distance in feature space and the psychophysical similarity measure. In order to evaluate the usefulness of these measures, we select pairs of masses and pairs of clustered microcalcifications on mammograms by using four different measures. We conduct two observer studies based on a two-alternative forced-choice (2AFC) method [95] for mass pairs and for calcification pairs, for comparison of subjective similarities in terms of radiologists' visual perception on pairs of images selected by use of different measures.

4.1 Materials and Methods

The use of the following database and the participation of radiologists in the observer study were approved by the Institutional Review Board at the University of Chicago. Informed consent for this observer study was obtained from all observers.

In this study, we investigate four objective similarity measures (A, B, C, and D) based on the Euclidean distance in feature space and the psychophysical similarity measure determined by the ANN. In Muramatsu's studies [90, 93, 94], fifty images including 25 benign and 25 malignant lesions were first selected as representative lesions for both mass and calcification studies by an attending breast radiologist to include various sizes and types of lesions. Three hundred pairs were created by the combination of each representative

lesion and six images (3 benign and 3 malignant lesions) selected subjectively by consensus of three investigators to include pairs with a wide range of similarities. Ten breast radiologists provided their subjective similarity ratings for the 300 mass pairs and the 300 calcification pairs. For specific image features considered in both the Euclidean distance and the ANN, we employ the combination of six and seven objective features for masses and clustered microcalcifications, respectively, which provide the highest correlation coefficients between the average subjective similarity ratings and psychophysical similarity measures [90, 93, 94]. The six features for masses include the degree of irregularity, the full width at half maximum of a cumulative modified radial gradient histogram, the radial gradient index, the minor-to-major-axis ratio of an ellipse fitted to the outline of the mass, the edge contrast, and the standard deviation of pixel values [90, 94]. On the other hand, the seven features for clustered microcalcifications include the circularity of the cluster, the number of microcalcifications per unit area, the mean effective diameter of microcalcifications, the standard deviation of the effective diameters of microcalcifications, the mean contrast of microcalcifications, the standard deviation of contrasts of microcalcifications, and the standard deviation of the shape irregularities of microcalcifications [93, 94]. Measures A and B are based on the Euclidean distance in feature space and the psychophysical similarity measure, respectively. Measure C is the sequential combination of B and A, which is derived first based on the psychophysical similarity measure and then the Euclidean distance in feature space, whereas measure D is the sequential combination of A and B, which is derived based on the Euclidean distance in feature space and then the psychophysical similarity measure.

4.1.1 Databases

To compare the usefulness of four different measures as an image-retrieval tool, we use pairs of masses and pairs of clustered microcalcifications on mammograms which were

obtained from the Digital Database for Screening Mammography (DDSM) developed by the University of South Florida [51]. Our database for masses consists of 1,568 regions of interest (ROIs), including 840 benign and 728 malignant masses [94]. The size of the ROI is 5 cm by 5 cm (pixel size 100 μm), centered at each mass. On the other hand, our database for clustered microcalcifications consists of 1,101 ROIs, including 644 benign and 457 malignant clustered microcalcifications [94]. The size of the ROI is 3 cm by 3 cm (pixel size 50 μm), centered at each clustered microcalcification. All lesions were proved by biopsy. The contrast and the density level in each ROI were manually adjusted to an appropriate level by an attending breast radiologist.

4.1.2 Selection of Pairs of Images

The pairs of images for masses and those for clustered microcalcifications are selected for each of the observer studies by use of the method described below. We first remove 300 ROIs used for training the ANN [90, 93, 94], which is then applied to the determination of psychophysical similarity measures for all of pairs of images used in this study. One hundred ROIs are selected randomly from the remaining ROIs (1,268 and 801 for masses and clustered microcalcifications, respectively) such that only one ROI is selected from the same patient. For the selected 100 ROIs, 4,950 pairs are created by all possible combinations of two different ROIs. Pairs of ROIs with the highest similarity measures are then selected for an observer study by use of four different measures. For measure A, five pairs with the five highest similarity measures based on the Euclidean distances in feature space are selected from the 4,950 pairs. For measure B, five pairs with the five highest psychophysical similarity measures are selected from the 4,950 pairs. For measure C, a pair with the highest psychophysical similarity measure is pre-selected in 99 pairs created by the combinations of one ROI and the other 99 ROIs. This procedure is repeated for all of the selected 100 ROIs. Subsequently, five pairs with the five highest similarity

measures based on the Euclidean distances are selected from the pre-selected 100 pairs. For measure D, a pair with the highest similarity measure based on the Euclidean distance is pre-selected in 99 pairs created by the combinations of one ROI and the other 99 ROIs. This procedure is repeated for all of the selected 100 ROIs. Subsequently, five pairs with the five highest psychophysical similarity measures are selected from the pre-selected 100 pairs. Here, five pairs for each measure are selected such that the same ROI is not selected again as another ROI in different pairs obtained with the same measure.

4.1.3 Observer Study

We conduct two observer studies for 20 mass pairs and for 20 calcification pairs, for comparison of subjective similarities in terms of radiologists' visual perception on pairs of ROIs selected by use of the 4 different measures. The 2AFC method, known as a paired comparison method, is employed in the observer study because it is a sensitive method for the distinction of a small difference in the comparison of two similar patterns [94]. In the observer study, two pairs of lesions are displayed on a high-resolution liquid-crystal-display monitor (ME511L/P4, 21.3 in., 2048 by 2560 pixels, 410 cd/m2 luminance; Totoku Electric Co., Ltd.) with one pair above and another pair below, as shown in Fig. 4.1. The observer is asked to compare the similarity of the two pairs and to select the pair considered more similar than the other pair. During the observer study, each pair is compared to all of the other 19 pairs one by one. The frequency with which a pair is selected as the more similar pair is considered as the subjective similarity ranking score for the pair; the maximum and the minimum score would be 19 and zero, respectively. The subjective similarity ranking scores indicate the relative rankings of similarities among the 20 pairs selected by four different measures.

Six observers, including three attending breast radiologists and three breast-imaging fellows, participate independently in the observer study. The instructions to the observers

Figure 4.1: Observer interface for obtaining subjective similarity ranking scores based on the 2AFC method.

included: 1) The purpose of this study is to obtain experimental data for subjective impression of similarity for pairs of masses (and pairs of clustered microcalcifications in the second session) on mammograms selected by four computerized methods. 2) Two pairs of images are displayed on a monitor. You are asked to compare the similarity of one pair above with that of another pair below, regarding the overall impression for diagnosis. Click on the one pair that is more similar than the other. 3) A training session including two comparisons of pairs of lesions is provided at the beginning of the study. 4) There is no time limit.

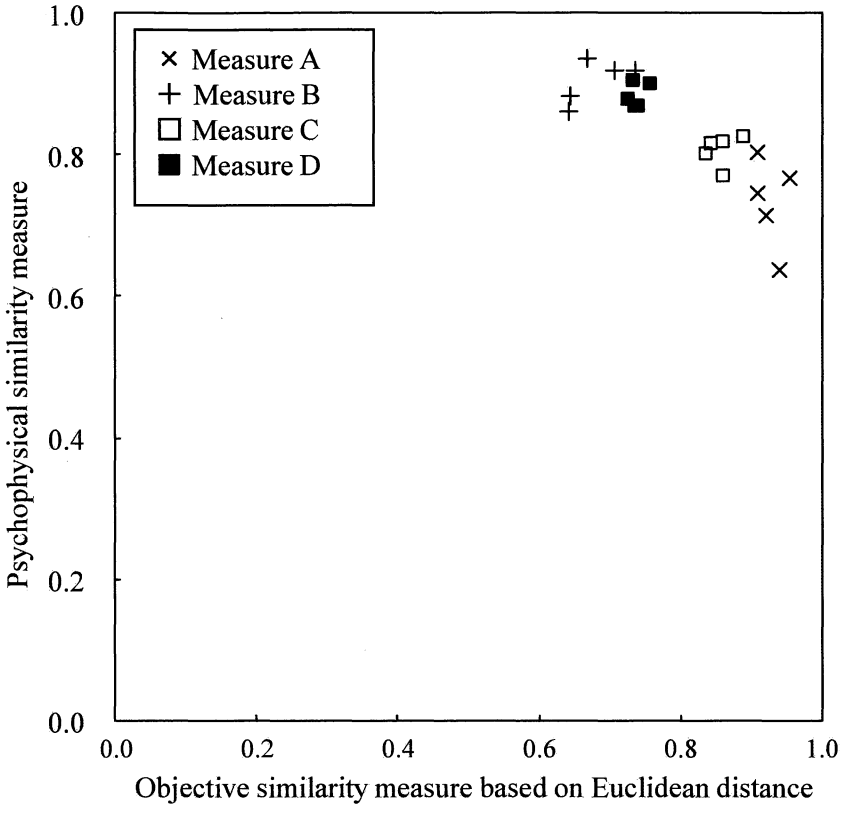
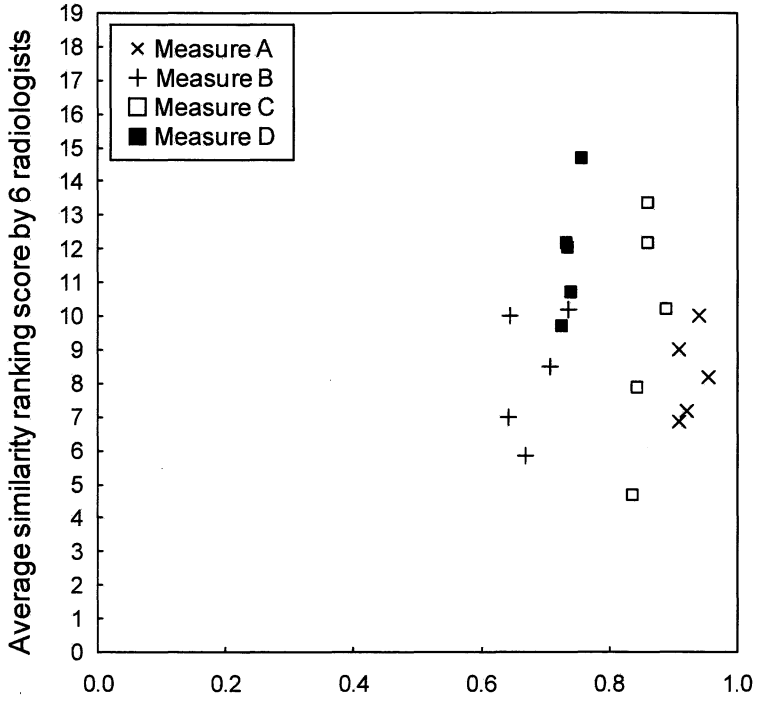


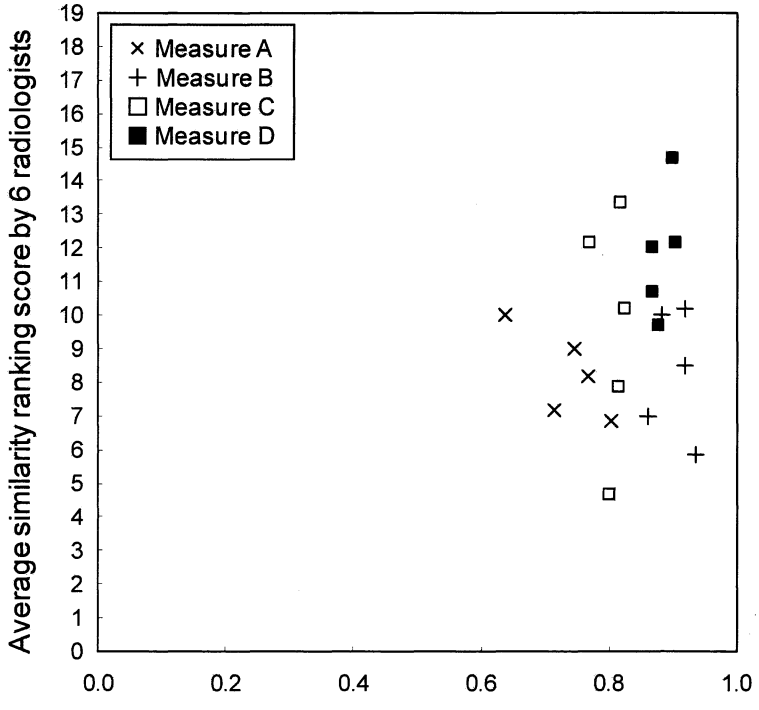
Figure 4.2: Relationship between objective similarity measure based on the Euclidean distance in feature space and psychophysical similarity measure for 20 mass pairs selected by 4 different measures.

4.2 Results

Figure 4.2 shows the relationship between the objective similarity measures based on the Euclidean distance and the psychophysical similarity measures for 20 mass pairs selected by the 4 different measures. The mass pairs selected by use of measure A tended to have high objective similarity measures based on the Euclidean distances and relatively low psychophysical similarity measures, whereas those by measure B tended to have relatively low objective similarity measures based on the Euclidean distances, but high psychophysical similarity measures. The pairs selected by use of measures C and D were distributed between the pairs for measures A and B. The pairs for measure C were distributed near the pairs for measure A, whereas the pairs for measure D were distributed



a. Objective similarity measure based on Euclidean distance



b. Psychophysical similarity measure

Figure 4.3: (a) Relationship for average similarity ranking score of each mass pair by 6 radiologists with objective similarity measure based on the Euclidean distance. (b) Relationship for average similarity ranking score with psychophysical similarity measure.

Table 4.1: Mean values and standard deviations of average subjective similarity ranking scores of mass pairs by six radiologists for each measure.

	Mean ± SD
Measure A	8.23 ± 1.30
Measure B	8.30 ± 1.88
Measure C	9.63 ± 3.47
Measure D	11.83 ± 1.89

Table 4.2: Relationship of statistical significances between computerized Measures based on average similarity ranking score of each mass pair.

	Measure A	Measure B	Measure C
Measure B	< .950		
Measure C	< .423	< .472	
Measure D	< .008	< .018	< . 248

near the pairs for measure B. It should be noted that there is a noticeable difference among 4 groups of pairs of masses selected as "most similar" based on the four different methods. Figures 4.3 (a) and (b) show the relationships for the average subjective similarity ranking score of mass pairs by 6 radiologists with the objective similarity measure based on the Euclidean distance, and also with the psychophysical similarity measure, respectively. Table 4.1 shows the mean values and the standard deviations of the average subjective similarity ranking scores for four groups of mass pairs selected by use of different measures. Although there was a large variation in the average similarity ranking scores for each measure, the mean value of the average similarity ranking scores for measure D was greater than those for the three other measures. On the other hand, the mean value of the average

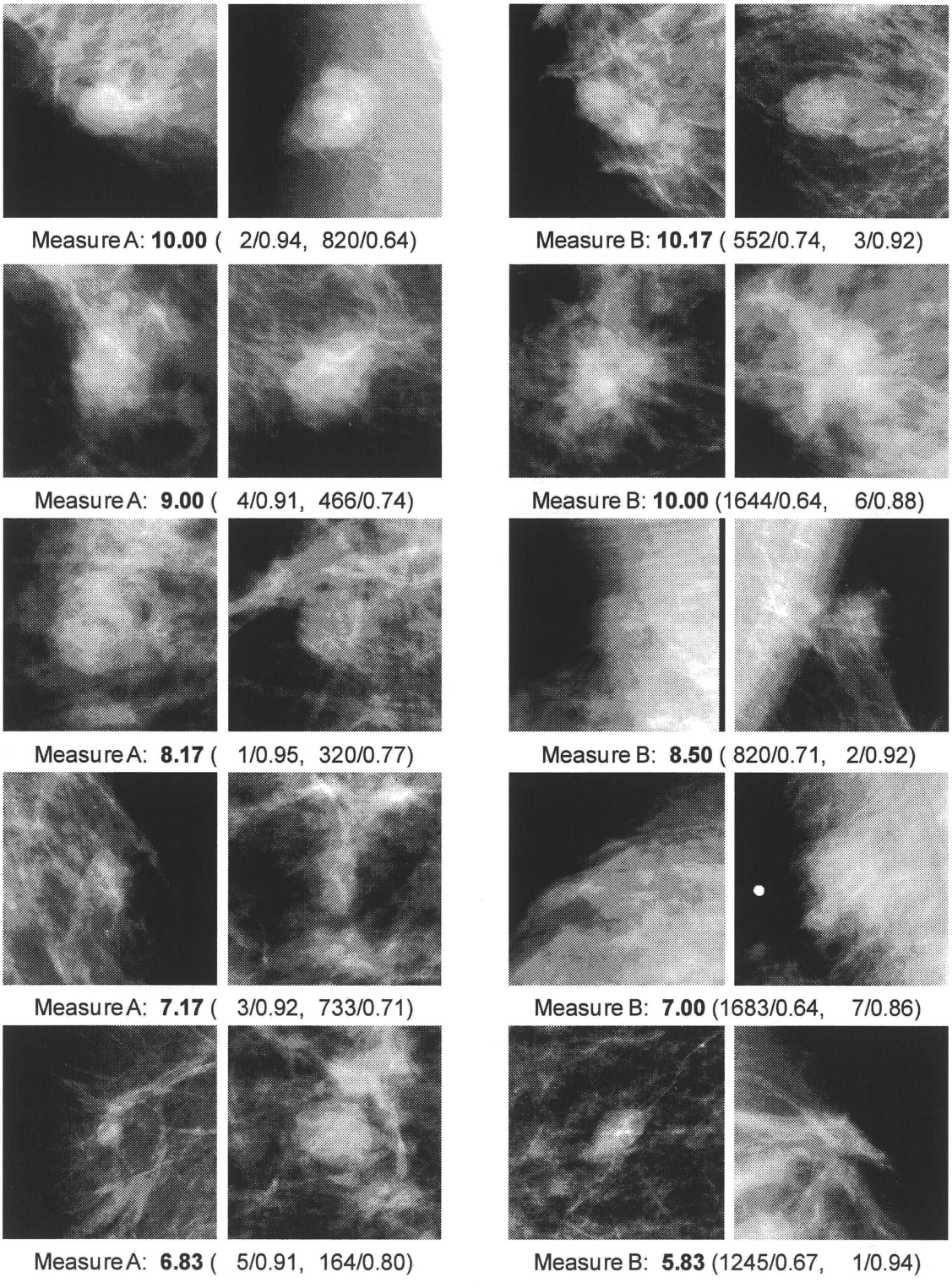


Figure 4.4: Mass pairs for measures A and B, and the average subjective similarity ranking score in bold (ranking on objective similarity measures based on the Euclidean distance in 4950 pairs / objective similarity measure, ranking on psychophysical similarity measures in 4950 pairs / psychophysical similarity measure) for each pair.

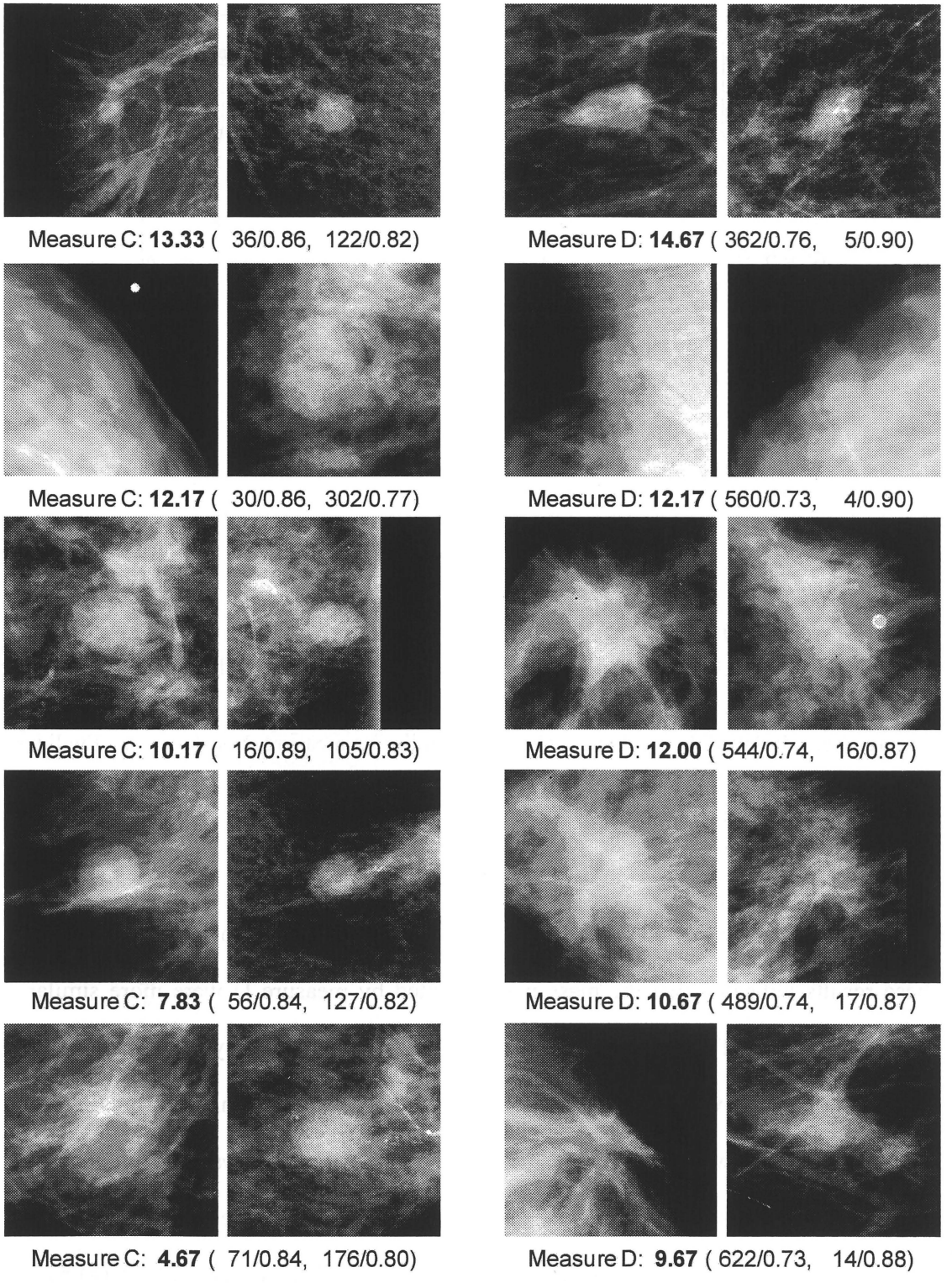


Figure 4.5: Mass pairs for measures C and D, and the average subjective similarity ranking score in bold (ranking on objective similarity measures based on the Euclidean distance in 4950 pairs / objective similarity measure, ranking on psychophysical similarity measures in 4950 pairs / psychophysical similarity measure) for each pair.

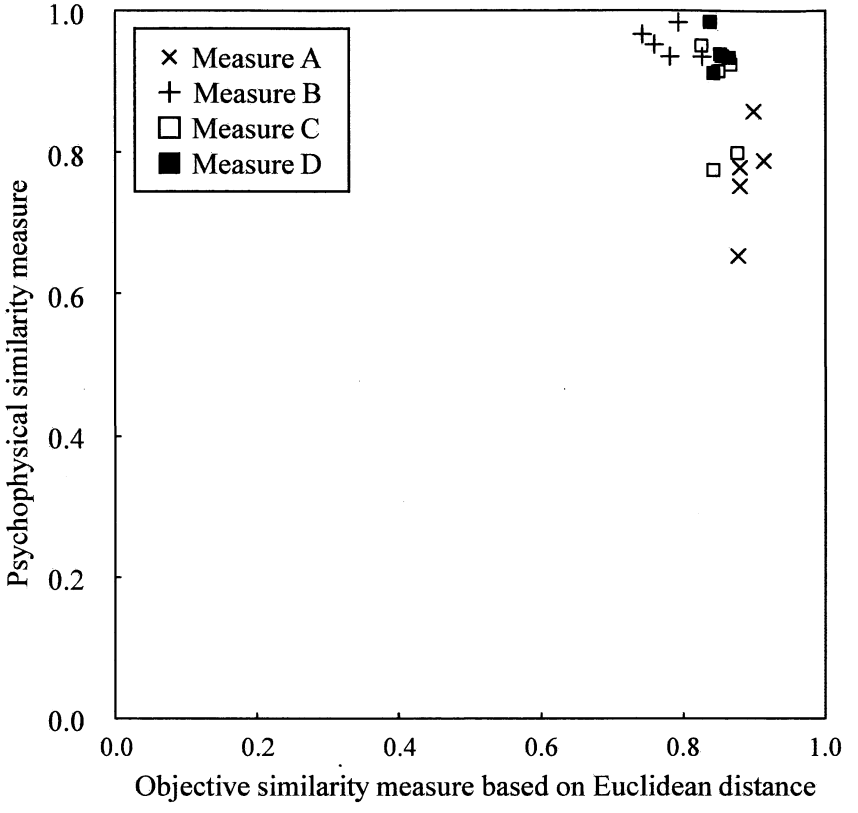
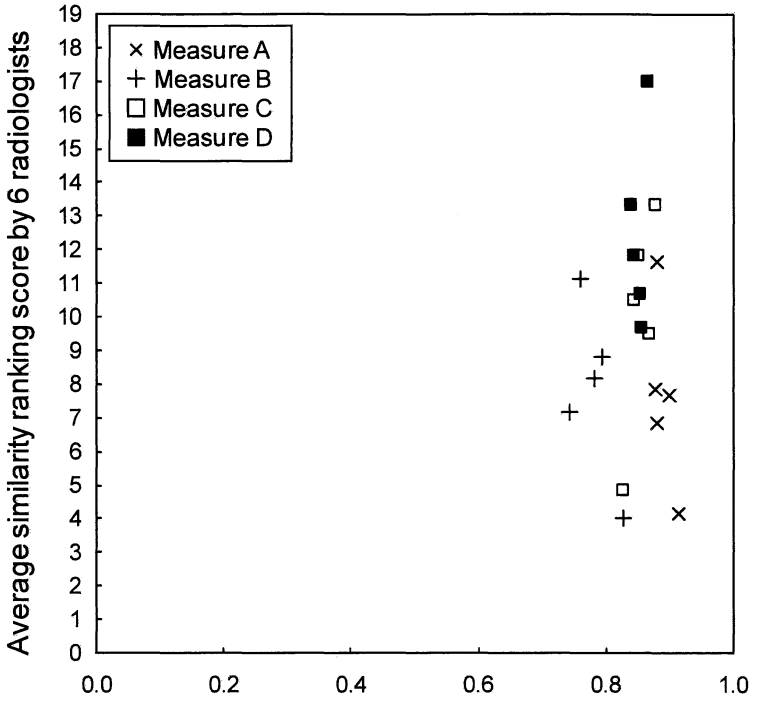
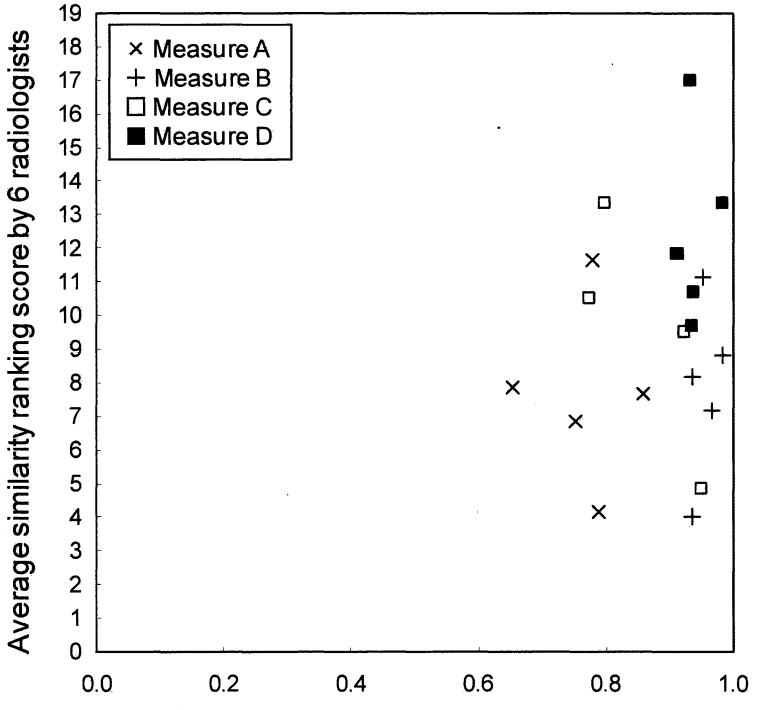


Figure 4.6: Relationship between objective similarity measure based on the Euclidean distance in feature space and psychophysical similarity measure for 20 calcification pairs selected by 4 different measures.

similarity ranking scores for measure A was lower than those for the other measures. These results indicated that the mass pairs selected by measure D were more similar, on average, in terms of radiologists' visual perception, than those by the other measures. Table 4.2 shows *P*-values for the difference in the average similarity ranking scores obtained by use of two different measures. A statistical analysis was performed with use of Student's t test based on the average similarity ranking score for each pair obtained by 6 radiologists. The difference ($P = .008$) between measures D and A and ($P = .018$) between measures D and B were statistically significant. Figures 4.4 and 4.5 show the 20 mass pairs obtained by use of the 4 different measures, together with the average subjective similarity ranking score in bold (ranking on objective similarity measures based on the



a. Objective similarity measure based on Euclidean distance



b. Psychophysical similarity measure

Figure 4.7: (a) Relationship for average similarity ranking score of each calcification pair by 6 radiologists with objective similarity measure based on the Euclidean distance. (b) Relationship for average similarity ranking score with psychophysical similarity measure.

Table 4.3: Mean values and standard deviations of average similarity ranking scores of calcification pairs for each computerized Measure.

	Mean ± SD
Measure A	7.63 ± 2.69
Measure B	7.87 ± 2.62
Measure C	10.00 ± 3.23
Measure D	12.50 ± 2.86

Table 4.4: Relationship of statistical significances between computerized Measures based on average similarity ranking score of each mass pair.

	Measure A	Measure B	Measure C
Measure B	< .893		
Measure C	< .244	< .285	
Measure D	< .024	< .028	< .231

Euclidean distance in 4950 pairs / objective similarity measure, and also ranking on psychophysical similarity measures in 4950 pairs / psychophysical similarity measure) for each pair. The first pair for measure D in Fig. 4.5 had the highest average similarity ranking score, whereas the fifth pair for measure C had the lowest average similarity ranking score.

Figure 4.6 shows the relationship between the objective similarity measure based on the Euclidean distance and the psychophysical similarity measure for 20 calcification pairs selected by the 4 measures. Although there was a small overlap in the distributions of calcification pairs among the 4 measures, the calcification pairs for each of the measures tended to be distributed in a way similar to those for the mass pairs in Fig. 4.2. Figures 4.7

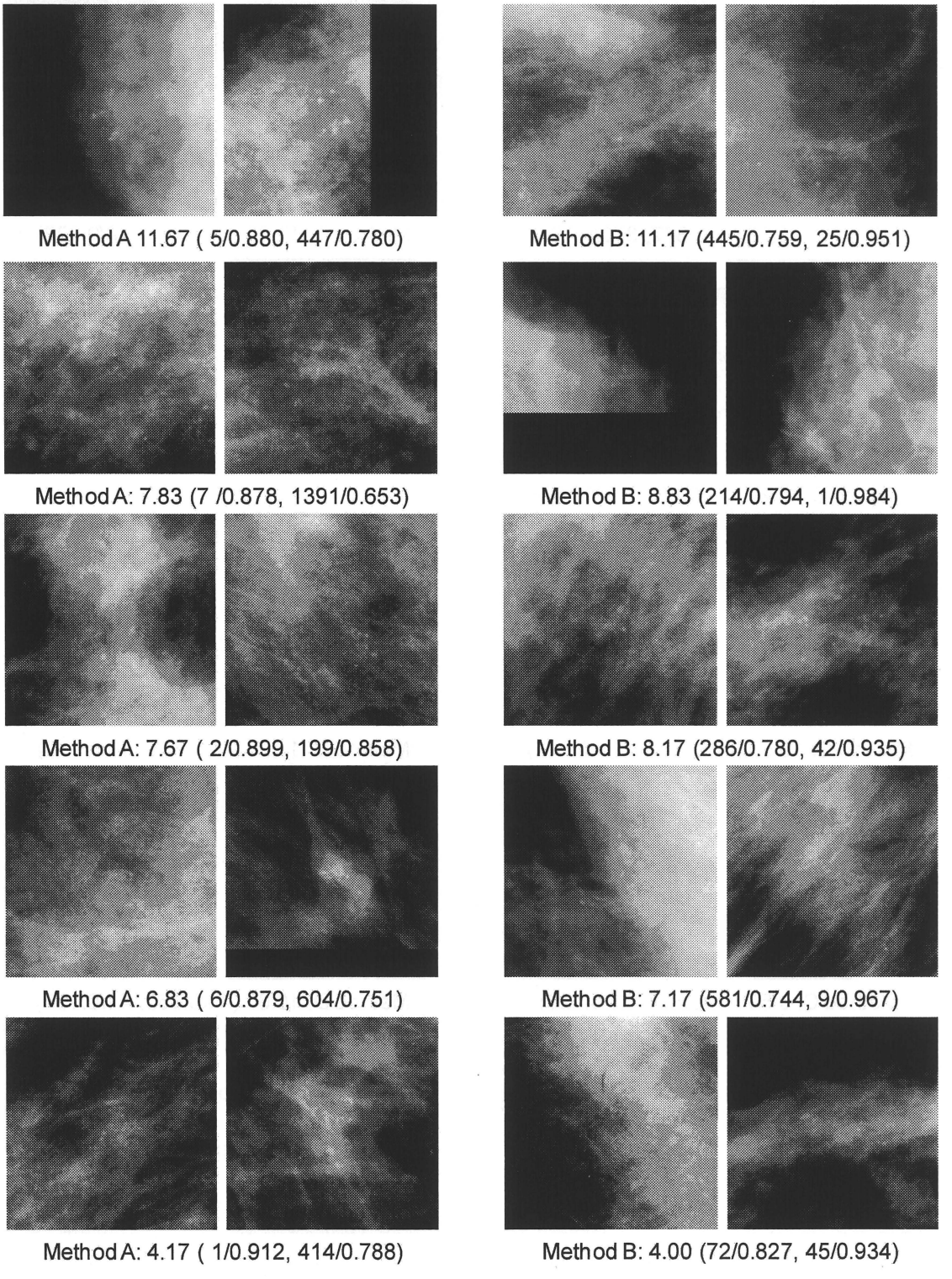
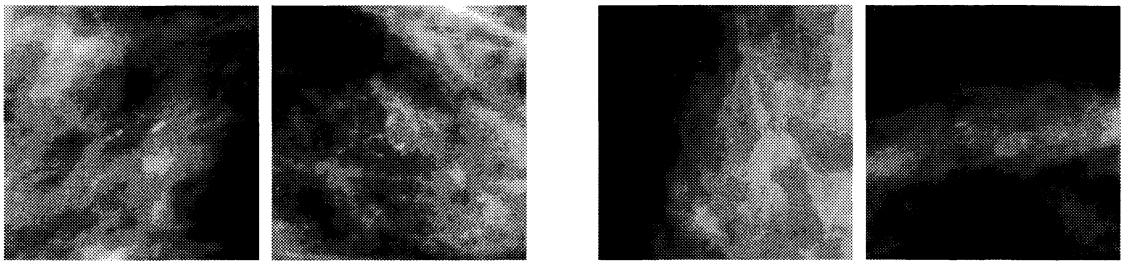
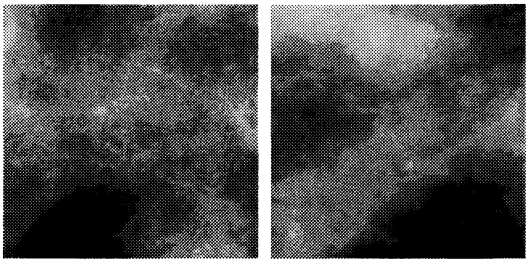


Figure 4.8: Calcification pairs for measures A and B, and the average subjective similarity ranking score in bold (ranking on objective similarity measures based on the Euclidean distance in 4950 pairs / objective similarity measure, ranking on psychophysical similarity measures in 4950 pairs / psychophysical similarity measure) for each pair.

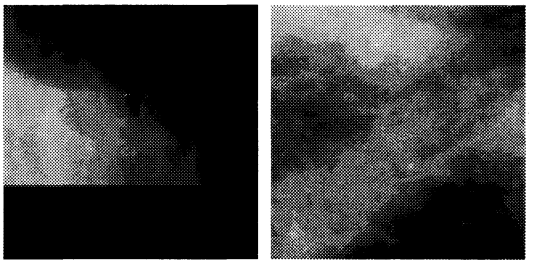


Method C: 13.33 (7/0.877, 370/0.797)

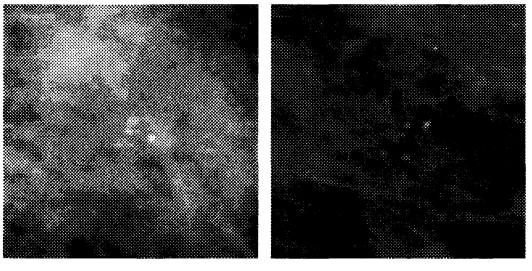
Method D: 17.00 (13/0.866, 48/0.932)



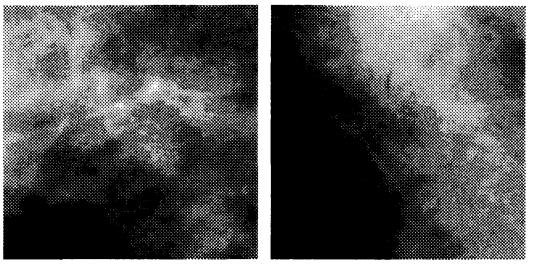
Method C: 11.83 (32/0.851, 76/0.912)



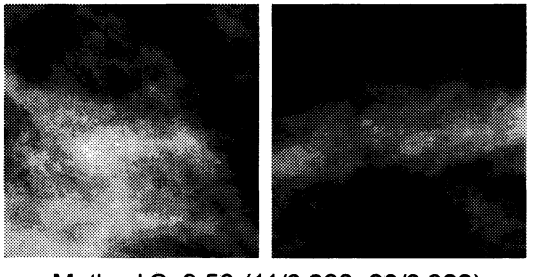
Method D: 13.33 (52/0.839, 2/0.983)



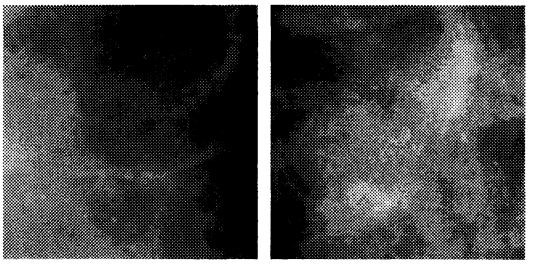
Method C: 10.50 (44/0.844, 475/0.774)



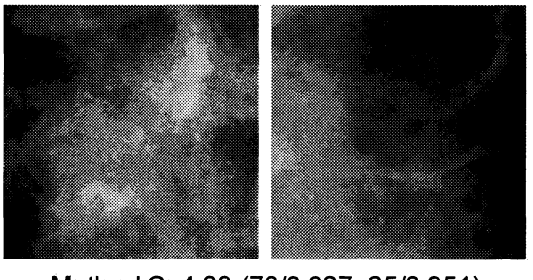
Method D: 11.83 (45/0.843, 80/0.910)



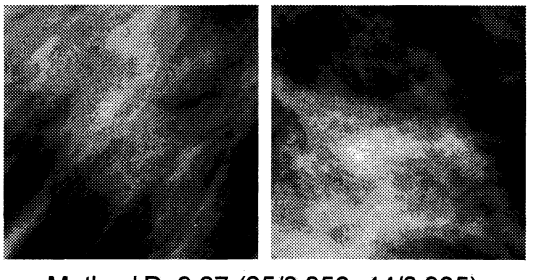
Method C: 9.50 (11/0.868, 63/0.922)



Method D: 10.67 (30/0.853, 41/0.937)



Method C: 4.83 (73/0.827, 25/0.951)



Method D: 9.67 (25/0.856, 44/0.935)

Figure 4.9: Calcification pairs for measures C and D, and the average subjective similarity ranking score in bold (ranking on objective similarity measures based on the Euclidean distance in 4950 pairs / objective similarity measure, ranking on psychophysical similarity measures in 4950 pairs / psychophysical similarity measure) for each pair.

(a) and (b) show the relationships for the average subjective similarity ranking score of calcification pairs to the objective similarity measure based on the Euclidean distance, and to the psychophysical similarity measure, respectively. Table 4.3 shows the mean values and the standard deviations of the average subjective similarity ranking scores of calcification pairs for each measure. The calcifications pairs for measure D had the highest average subjective similarity ranking scores, whereas those for measure A had the lowest average similarity ranking scores; these results were the same as those for masses. Table 4.4 shows P-values for the difference in the average similarity ranking scores obtained by use of two different measures. The difference ($P = .024$) between measures D and A and that ($P = .028$) between measures D and B were statistically significant. Figures 4.8 and 4.9 show the 20 calcification pairs obtained by use of the 4 different measures, together with the average subjective similarity ranking score in bold (ranking on objective similarity measures based on the Euclidean distance in 4950 pairs / objective similarity measure, and also ranking on psychophysical similarity measures in 4950 pairs / psychophysical similarity measure) for each pair. The pairs with very high objective similarity measures both for the Euclidean distance and the ANN tended to have high average subjective similarity ranking scores in measures C and D.

4.3 Discussion

In both observer studies for mass pairs and calcification pairs, the mean values of the average subjective similarity ranking scores for measure B were greater than those for measure A, although the difference between measures A and B was not statistically significant in this study. This result tended to be consistent with the results presented by Li et al. [89] and Muramatsu et al. [90, 93, 94], where the correlation coefficient of radiologists' subjective similarity ratings with psychophysical similarity measures was greater than that with objective similarity measures based on the Euclidean distance.

These results may indicate that the psychophysical similarity measure is a better tool in retrieving similar images than is the objective similarity measure based on the Euclidean distance.

The mean values of the average similarity ranking scores for measures C and D were greater than those for measures A and B. The mean value of the average similarity ranking scores for measure D was greater than that for C. For measure D, the pairs with comparable physical characteristics were first pre-selected by use of an objective similarity measure based on the Euclidean distance, and thus the subsequent selection of pairs with high psychophysical similarity measures would be more reliable, because inadequate pairs which may not be similar due to a large difference in physical characteristics were removed initially. Therefore, we believe that the pairs selected by measure D would be more similar in terms of radiologists' visual perception than those by measure B, because measure B was improved substantially by the sequential combination with measure A. With measure C, on the other hand, the pairs were first pre-selected by use of a psychophysical similarity measure, and thus some pairs with high objective similarity measures, which would be located closely in feature space, would have been removed, and the subsequent selection of pairs may provide pairs with different physical characteristics. Therefore, we believe that the pairs for measure D would be more similar subjectively than those for measure C.

The implementation of selecting similar images by use of measure D in clinical situations can be illustrated in the example described below. When a radiologist encounters a new, unknown case in daily clinical practice at a breast clinic, a search engine would determine first the objective similarity measures based on the Euclidean distance in feature space for all of the combinations for the unknown case with all of the known benign/malignant cases in the database available in the clinic, which may include a large number of cases such as 1,000 benign cases and 1,000 malignant cases stored in a Picture

Archiving and Communication System. The search engine would then select a certain pre-selected number of cases such as the top 100 pairs, each for benign and malignant cases, with higher objective similarity measures; these pairs would be subjected to determination of the psychophysical similarity measures by use of the trained ANN. Finally, the radiologist may indicate a desired number of similar cases to be presented as an aid to his/her diagnosis, such as five cases each for benign/malignant cases. The search engine then could retrieve those cases with the five highest psychophysical similarity measures in each category to be presented as similar cases. It is likely that the cases selected would look more similar to the unknown case in question for radiologists in making their diagnostic decision than other cases which might be selected by the three other measures, A, B, or C.

There are some limitations in this study. One limitation is that the number of pairs for each objective similarity measure was small in the observer study because the time required for a radiologist has to be limited to an hour in one session. Another limitation is that four of six breast radiologists who participated in the observer study provided their subjective similarity ratings for the 300 mass pairs and the 300 calcification pairs in our previous studies. However, we believe that the bias due to this overlap would be minimal, because for training the ANN, the average subjective similarity ratings were obtained by ten breast radiologists.

CHAPTER 5
POTENTIAL USEFULNESS OF SIMILAR IMAGES IN THE DIFFERENTIAL DIAGNOSIS

Radiologists commonly learn diagnostic skills by viewing many cases in their training and clinical practice. Based on their experience and knowledge, they may make a diagnostic decision on a new, unknown lesion that appears in medical images. Therefore, it is expected that the presentation of images of lesions with known pathology similar to a new, unknown lesion would be useful for radiologists in the differential diagnosis of the unknown lesion [72-75, 78-82, 84, 86, 87, 89-93, 96]. Several investigators have attempted to use similar images as a diagnostic aid for chest lesions [78, 84, 89] and breast lesions [79-80, 82, 86, 87, 91-93, 96]. However, if similar images were not really similar to an unknown lesion in terms of radiologists' visual perception, those similar images would not be useful in assisting radiologists in the differential diagnosis of the unknown lesion. In selecting similar images from a database, investigators [89-93] have developed a computerized scheme for automatically selecting similar images based on a psychophysical similarity measure which is obtained by use of an artificial neural network for learning the relationship between radiologists' subjective ratings of similarity and the objective features of lesions. However, these studies have not demonstrated clearly and convincingly that radiologists' performance in the differential diagnosis of lesions would be improved by the presentation of similar images [89, 90].

Recently, Muramatsu [94] has conducted an observer study for evaluating the usefulness of similar images selected based on a psychophysical similarity measure in the distinction between benign and malignant masses on mammograms. The results indicated that there was little difference between radiologists' performance in the differential diagnosis without and with the presentation of the similar images. From a detailed

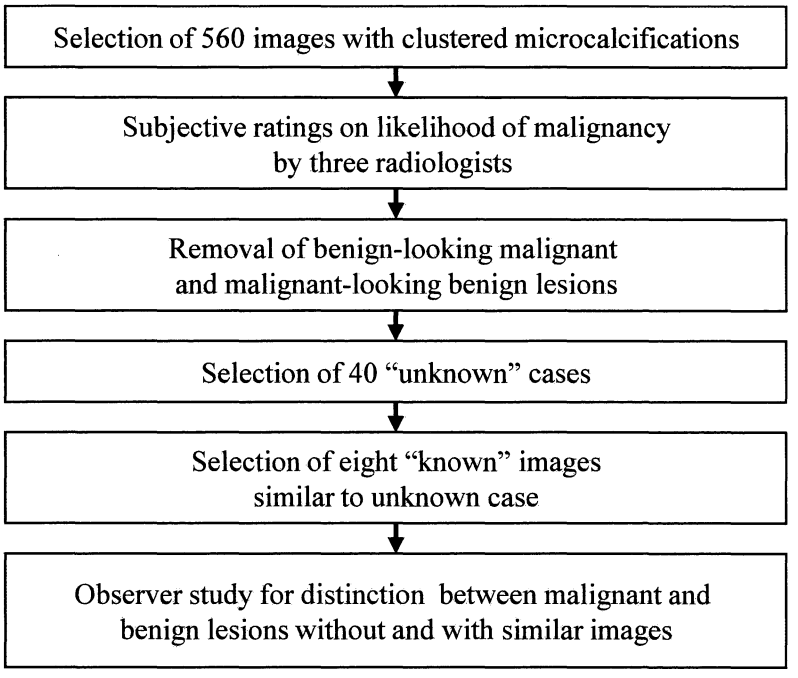


Figure 5.1: Overall scheme of this study

analysis of her results, she found that the cases used in the observer study included unusual and inadequate cases, i.e., "malignant-looking" benign and "benign-looking" malignant lesions. Thus, radiologists appeared to be confused in the differential diagnosis of lesions when these malignant-looking benign and benign-looking malignant lesions were presented as similar images. However, in the practical setting for the application of similar images, we can eliminate such unusual or inadequate cases in advance. In this study, therefore, we evaluated the usefulness of the presentation of similar images in the distinction between benign and malignant clustered microcalcifications on mammograms by removing these unusual cases from our database.

5.1 Materials and Methods

The use of the following database and the participation of radiologists in the observer study were approved by the Institutional Review Board at the University of Chicago. Informed

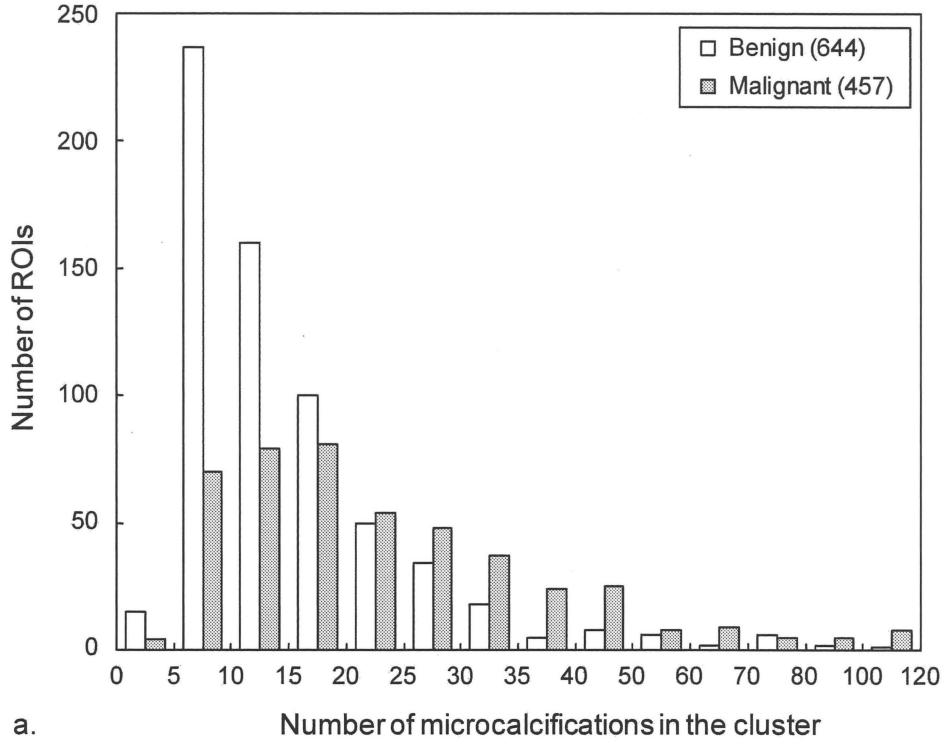
consent for this observer study was obtained from all observers.

Figure 5.1 shows the overall scheme of this study. First, we select 560 regions of interest (ROIs) with clustered microcalcifications, including both benign and malignant cases. Three attending breast radiologists independently provide subjective ratings on the likelihood of malignancy for each of the 560 clustered microcalcifications. Based on the average subjective ratings, one hundred ROIs are removed as malignant-looking benign or benign-looking malignant lesions. Forty lesions are then selected as "unknown" cases for an observer study by use of a stratified randomization method. For each unknown case, four "known" malignant and four "known" benign lesions similar to the unknown case in terms of radiologists' visual perception are selected as the similar images. An observer study is conducted for evaluating the usefulness of the presentation of the similar images in distinguishing between benign and malignant clustered microcalcifications.

5.1.1 Case Selection

Our database initially consists of 1,101 ROIs, including 644 benign and 457 malignant clustered microcalcifications which were obtained from a publicly available database, the Digital Database for Screening Mammography (DDSM) developed by the University of South Florida [51]. All microcalcification lesions were proved by biopsy. The size of the ROI is 3 cm by 3 cm, centered at each lesion. Figure 5.2(a) and (b) shows the distribution of the number of microcalcifications in the clusters and the distribution of the size of clusters in our database, respectively.

To remove malignant-looking benign and benign-looking malignant lesions from our study, it is necessary to obtain radiologists' subjective ratings on the likelihood of malignancy for all of the clustered microcalcifications. In order to minimize this task by radiologists, we remove 146 ROIs if the number of microcalcifications is less than 5, or greater than 35, and if the cluster size is greater than 25 mm. It would be difficult to select



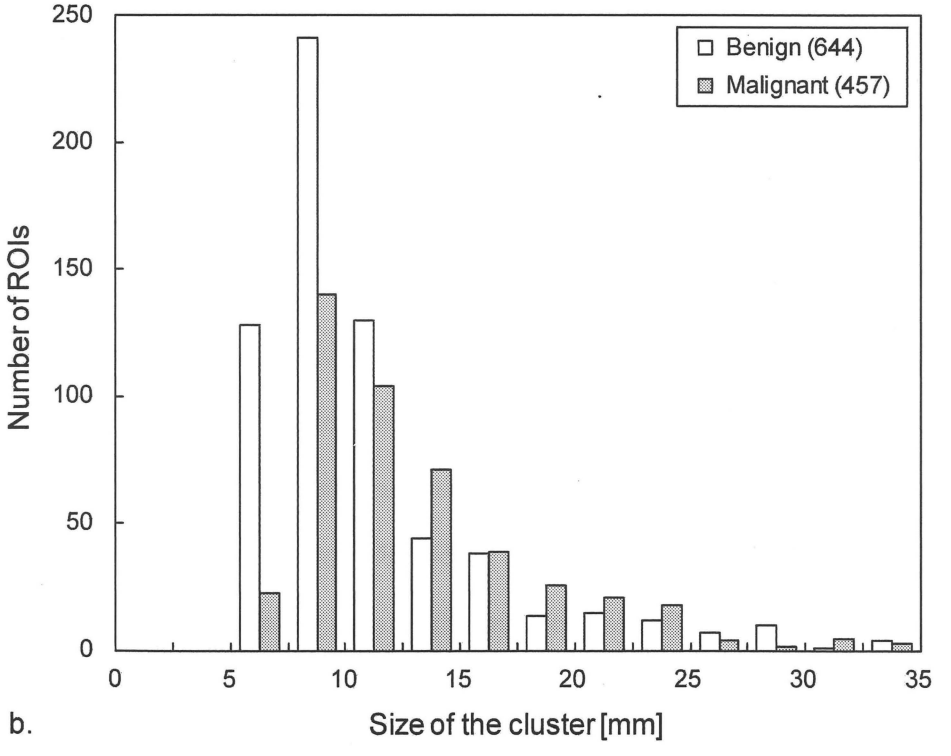


Figure 5.2: (a) Distribution of the number of microcalcifications in the clusters. (b) Distribution of the size of clusters included in our database.

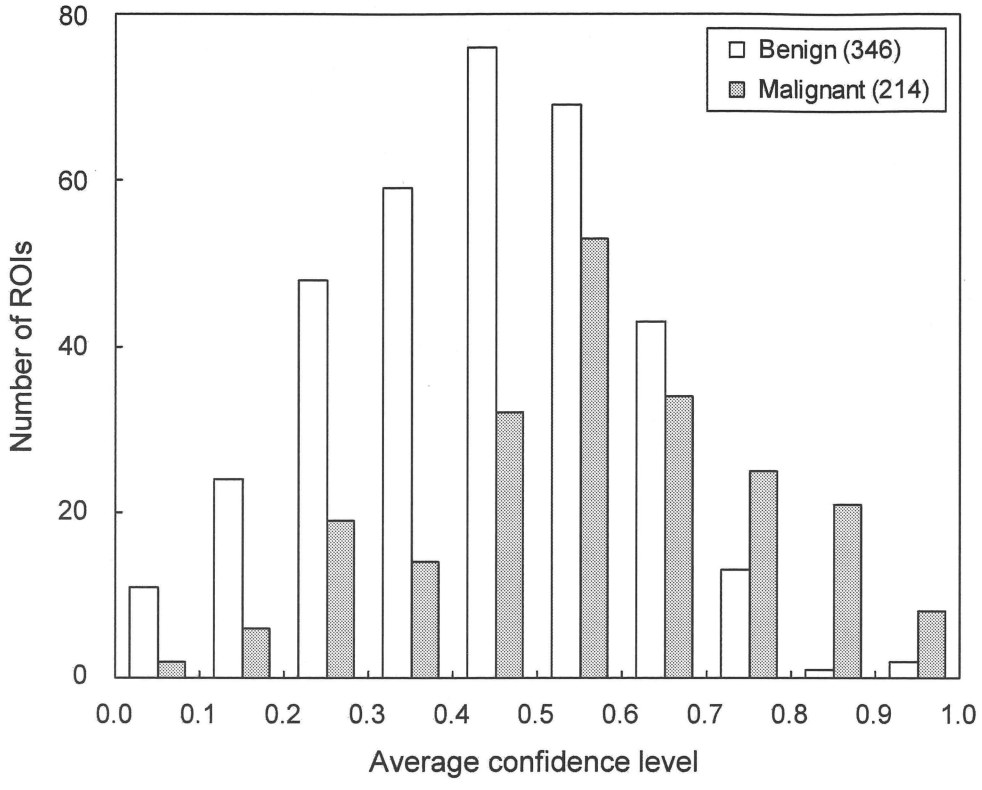


Figure 5.3: Distribution of average confidence levels of malignancy by three attending breast radiologists.

similar images if some of these lesions are used as "unknown" cases because of the limited number of cases available. Of the remaining 955 ROIs, 790 ROIs included both cranio-caudal and medio-lateral-oblique views for 395 clustered microcalcifications, whereas 165 ROIs included only one view. Only one view from each case was selected by an attending breast radiologist based on the quality of the clustered microcalcifications. As a result, the cases in this study consisted of 560 ROIs, including 346 benign and 214 malignant clustered microcalcifications obtained from 560 patients (mean age, 57.4 years; age range, 32-87 years; 560 women).

Three attending breast radiologists (H.A., C.S., R.A.S.; 6-26 years experience) independently provide their confidence level regarding the malignancy (or benignity) on a continuous rating scale from 0 to 1 corresponding to "definitely benign" and "definitely

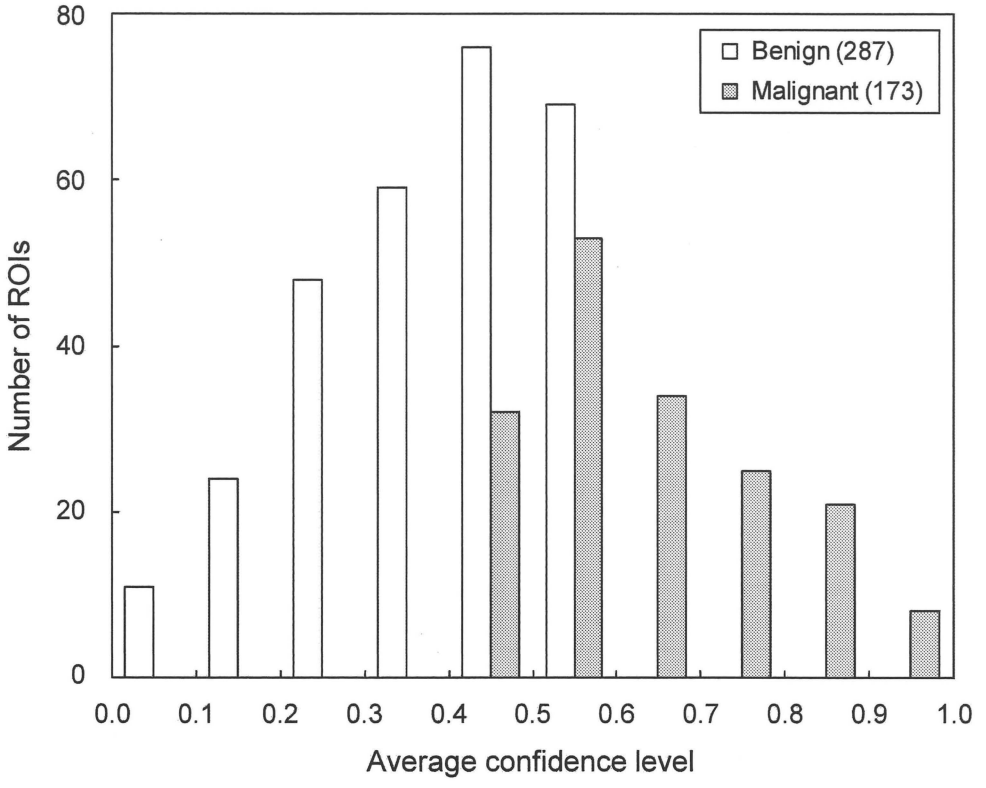


Figure 5.4: Distribution of average confidence levels of malignancy after removing benign-looking malignant and malignant-looking benign lesions.

malignant," respectively, for each of 560 clustered microcalcifications, which is displayed on a high-resolution liquid-crystal display (LCD) monitor (ME511L/P4, 21.3 in., 2048 by 2560 pixels, 410 cd/m2 luminance; Totoku Electric Co., Ltd., Tokyo, Japan). Figure 5.3 shows the distribution of the average confidence levels of malignancy by the three breast radiologists. The average inter-observer correlation coefficient for all possible pairs of 3 observers was 0.554. We assume in this study that benign lesions with the average confidence levels greater than 0.6 would be considered malignant-looking benign lesions, whereas malignant lesions with average confidence levels less than 0.4 would be considered benign-looking malignant lesions. Therefore, one hundred ROIs with those lesions were removed, and the cases to be used in our observer study consisted of 460 ROIs including 287 benign and 173 malignant clustered microcalcifications. Figure 5.4 shows the

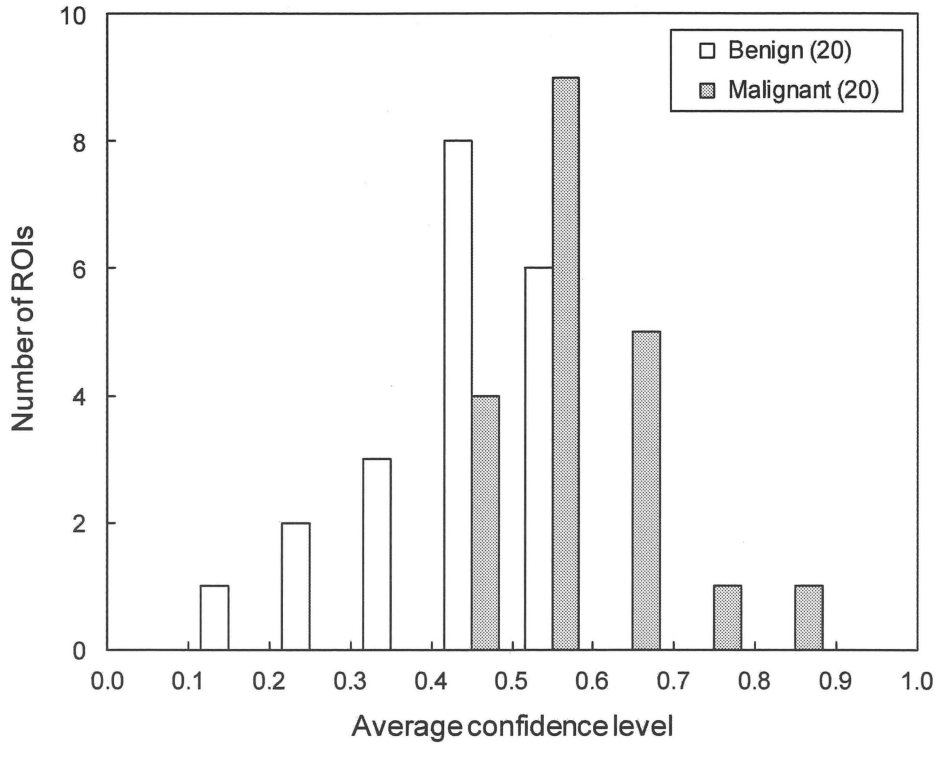


Figure 5.5: Distribution of average confidence levels of malignancy for 40 unknown cases.

distribution of the average confidence levels of malignancy for the remaining 460 cases used in our observer study.

In order to include moderately difficult and indeterminate cases as unknown cases in the observer study, we select forty unknown cases by use of a stratified randomization method based on the average confidence levels of malignancy. Twenty malignant lesions were selected as unknown cases such that the average confidence levels of malignancy for unknown malignant lesions would be distributed approximately normally in the range from 0.40 to 0.90, as shown in Fig. 5.5, whereas twenty benign lesions were also selected as unknown cases such that those for unknown benign lesions would be distributed approximately normally in the range from 0.10 to 0.60.

Similar images for each unknown case should be selected by radiologists because the purpose of this study is to evaluate the usefulness of images which should be similar to the

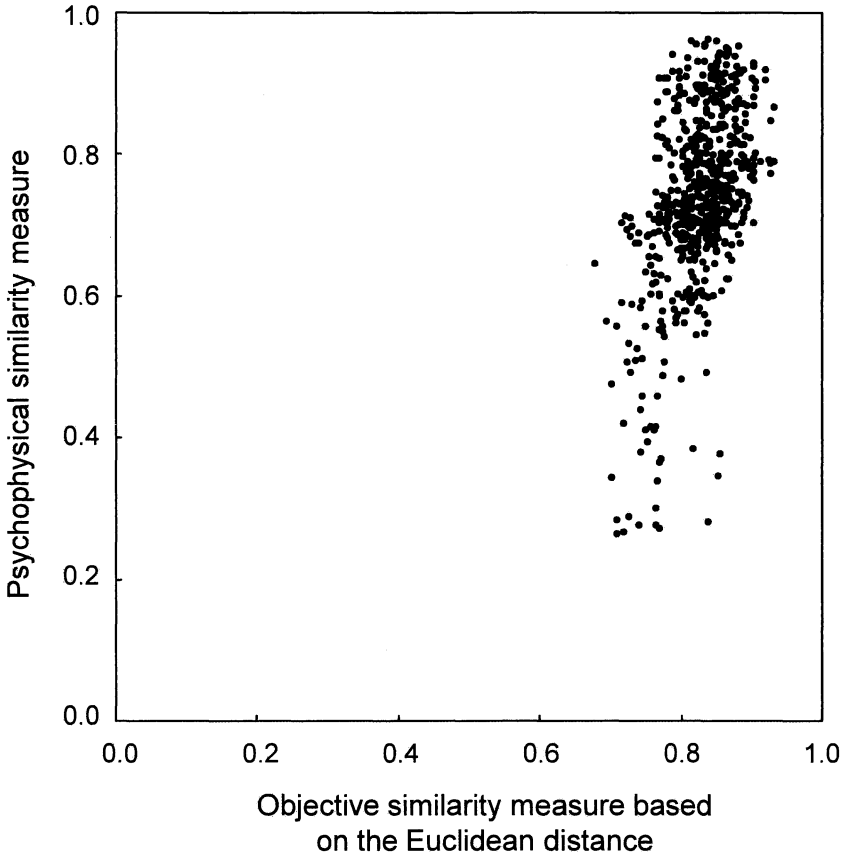


Figure 5.6: Relationship between the objective similarity measure based on the Euclidean distance and the psychophysical similarity measure for all pairs of 40 unknown cases with 8 similar cases selected from each of benign and malignant known images

unknown lesions in terms of radiologists' visual perception. However, it would be impractical for radiologists to select the similar images for each unknown case from the many cases included in our database. Therefore, we decided first to select a set of similar images by an automated computerized method; these are then selected subjectively by a radiologist to constitute the final sets of similar images. For each unknown case, eleven "known" benign lesions and eleven "known" malignant lesions were first pre-selected as the similar images from our database by use of an objective similarity measure based on the short Euclidean distance in image feature space [94], i.e., a set of objective image features for an unknown lesion would be comparable or similar to those of pre-selected lesions.

Subsequently, eight similar images for each of the benign and malignant lesions were selected based on the eight highest psychophysical similarity measures [93]. Figure 5.6 shows the relationship between the objective similarity measure based on the Euclidean distance and the psychophysical similarity measure for all pairs of 40 unknown cases with 8 similar cases pre-selected by the computerized scheme. There was a large variation in psychophysical similarity measures, although all of these cases had relatively high objective similarity measures based on the Euclidean distance. Finally, the four most similar images, which would be very similar to each unknown case, in each of 8 benign and 8 malignant images were selected subjectively based on the overall impression for radiological diagnosis by an attending breast radiologist (H. A.; 6 years experience). This selection of four from eight images was performed without knowing if the group of images was benign or malignant.

5.1.2 Observer Study

In the observer study, an unknown image is displayed in the center of a high-resolution LCD monitor. The observer is asked to mark his/her confidence level regarding the malignancy of the unknown case on the continuous (using units of .01) rating scale from 0 to 1 corresponding to "definitely benign" and "definitely malignant," respectively. After the observer marks the initial confidence level, four benign and four malignant similar images are displayed on the left and right sides of the unknown image, respectively, and the initial mark is erased. The observer is asked again to mark his/her confidence level regarding the malignancy of the unknown case.

Eight observers, including five attending breast radiologists (H.A., C.S., K.K., R.A.S., G.M.N.; 6-30 years experience) and three breast imaging fellows (L.F., A.S., R.S.; 0-1 years experience), participate independently in the observer study. Three of five attending breast radiologists have seen 560 lesions when providing the confidence level regarding the

malignancy for those lesions, and the period between the work for providing their confidence levels and the observer study was more than 3 months. The instructions to the observers included: 1) The purpose of this study is to investigate whether providing the similar known images can assist radiologists in the distinction between benign and malignant lesions on mammograms. 2) Forty unknown cases are included in this study. A training session including two cases is provided at the beginning of the study. 3) You are asked to provide your confidence level regarding the malignancy (or benignity) of a lesion on a bar by use of a mouse first without similar images, and then after observing the similar images. 4) For each unknown case, four most similar images each from benign and malignant lesions in the database are provided. 5) There is no time limit.

5.1.3 Statistical Analysis

For evaluating the radiologists' performances without and with the similar images in the distinction between benign and malignant clustered microcalcifications, we employ a receiver operating characteristic (ROC) analysis based on a sequential-test method [97, 98]. The areas under the ROC curve (AUCs) and the 95% confidence intervals are obtained with a quasi-maximum-likelihood estimation of binormal distribution by use of DBM MRMC software (version 2.2) developed by the University of Iowa and the University of Chicago [99, 100]. The significance of the difference in AUCs between observer readings without and with similar images is tested with use of the Dorfman-Berbaum-Metz method [99, 100], which includes both reader variation and case sample variation by means of an analysis of variance approach. A P value less than .05 is considered to indicate a statistically significant difference. Inter-observer variability is determined by the mean value of the standard deviations of radiologists' confidence levels for each unknown case. An average change in confidence level greater than 0.05 due to the use of similar images by the 8 observers is also assumed to be a beneficial or detrimental change.

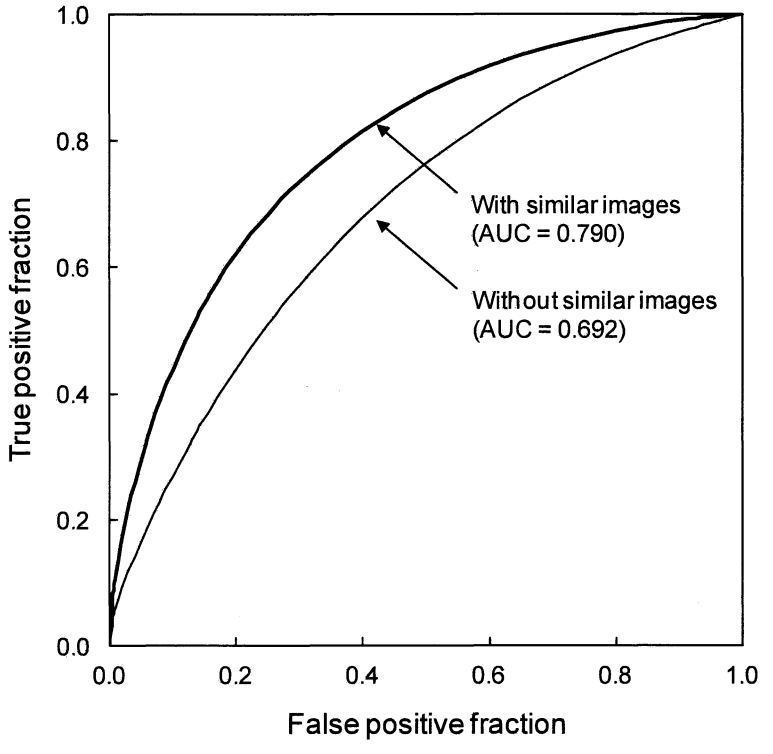


Figure 5.7: Comparison of ROC curves for the average performance of the eight radiologists in the distinction between benign and malignant clustered microcalcifications without and with similar images. With similar images, the average AUC was improved significantly from 0.692 to 0.790 (P = .0009).

5.2 Results

Figure 5.7 shows the average ROC curves for all observers in distinguishing between benign and malignant clustered microcalcifications without and with similar images. The average AUC increased from 0.692 without to 0.790 with the similar images. This difference was statistically significant (P = .0009). Table 5.1 shows the AUCs for each observer without and with similar images. All observers' performances in the differential diagnosis were improved when the similar images were available. The average AUCs for the five attending breast radiologists without and with the similar images were 0.698 and 0.794 (P = .024), respectively, whereas those for the three breast imaging fellows without

Table 5.1: AUCs for radiologists in the distinction between benign and malignant clustered microcalcifications without and with similar images [95% confidence intervals]

Observer	AUC		*P* value
	Without similar images	With similar images	
Attending breast radiologist			
A	0.683 [0.447, 0.919]	0.867 [0.728, 1.001]	.0421
B	0.658 [0.473, 0.842]	0.787 [0.515, 1.058]	.3262
C	0.675 [0.501, 0.849]	0.751 [0.539, 0.964]	.4055
D	0.799 [0.677, 0.922]	0.832 [0.697, 0.967]	.6310
E	0.677 [0.493, 0.861]	0.734 [0.569, 0.901]	.3409
Mean	0.698 [0.589, 0.807]	0.794 [0.708, 0.880]	.0241
Breast imaging fellow			
F	0.734 [0.505, 0.962]	0.806 [0.675, 0.936]	.3959
G	0.677 [0.539, 0.816]	0.765 [0.616, 0.913]	.1902
H	0.632 [0.469, 0.796]	0.778 [0.619, 0.937]	.0686
Mean	0.681 [0.574, 0.788]	0.783 [0.713, 0.852]	.0135
ALL	0.692 [0.596, 0.788]	0.790 [0.723, 0.857]	.0009

and with the similar images were 0.681 and 0.783 ($P = .014$), respectively. The gain in the average AUCs for the attending breast radiologists was comparable to that for the breast imaging observers ($P = .886$). Inter-observer variability without and with similar images was 0.139 and 0.149, respectively.

Figure 5.8 shows the relationship between the average initial confidence levels and the average beneficial or detrimental changes in confidence level for benign and malignant cases due to the use of similar images by the 8 observers. The number of cases with a beneficial effect was much greater than that with a detrimental effect. If an average change in confidence level greater than 0.05 was assumed to be a beneficial or detrimental

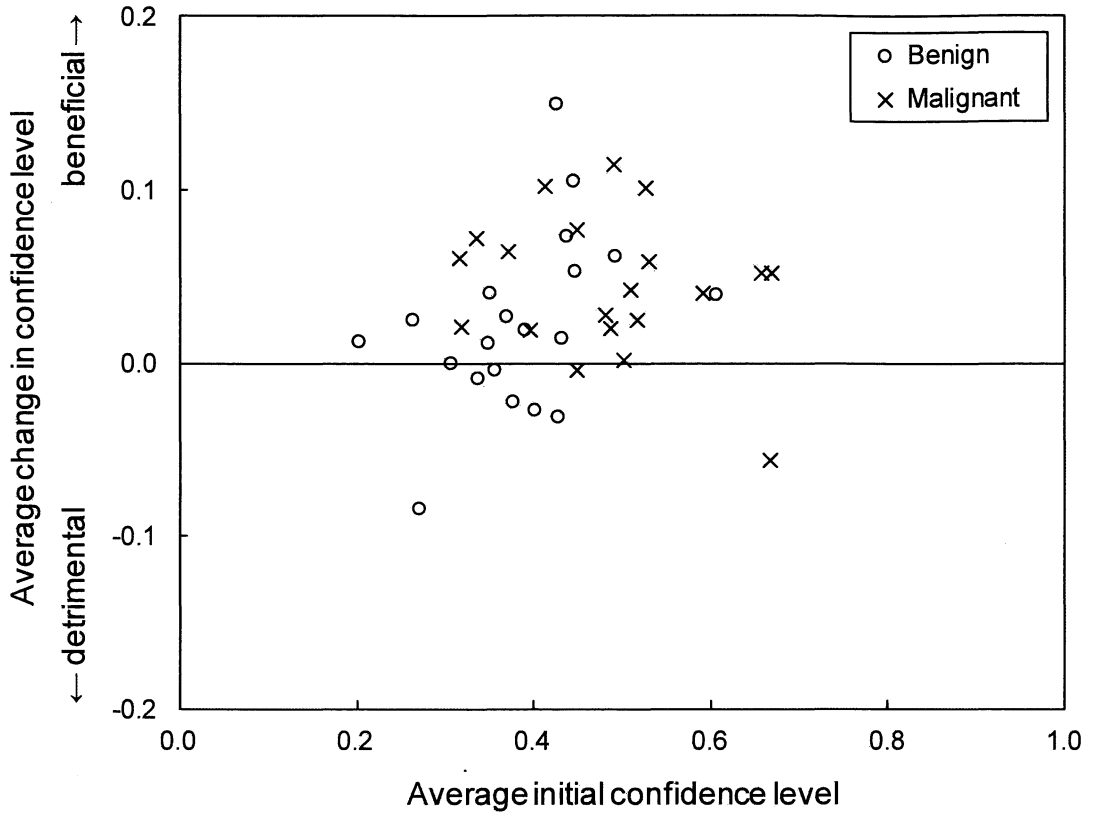


Figure 5.8: Average beneficial or detrimental changes in confidence level due to similar images.

change, the number of beneficial changes was 15, whereas that of detrimental changes was 2. For each observer, the number of beneficial changes was larger than that of detrimental changes.

Figure 5.9 shows the unknown malignant case with the most beneficial change which corresponds to the average confidence level from 0.491 to 0.605 by use of similar images. In this case, six of the eight observers changed their confidence levels of malignancy beneficially after viewing the similar images. The other two hardly changed their confidence levels. Most observers appeared to find that a set of malignant similar images were more similar to this unknown case than were benign similar images. On the other hand, in the unknown benign case with the most beneficial change, most observers appear

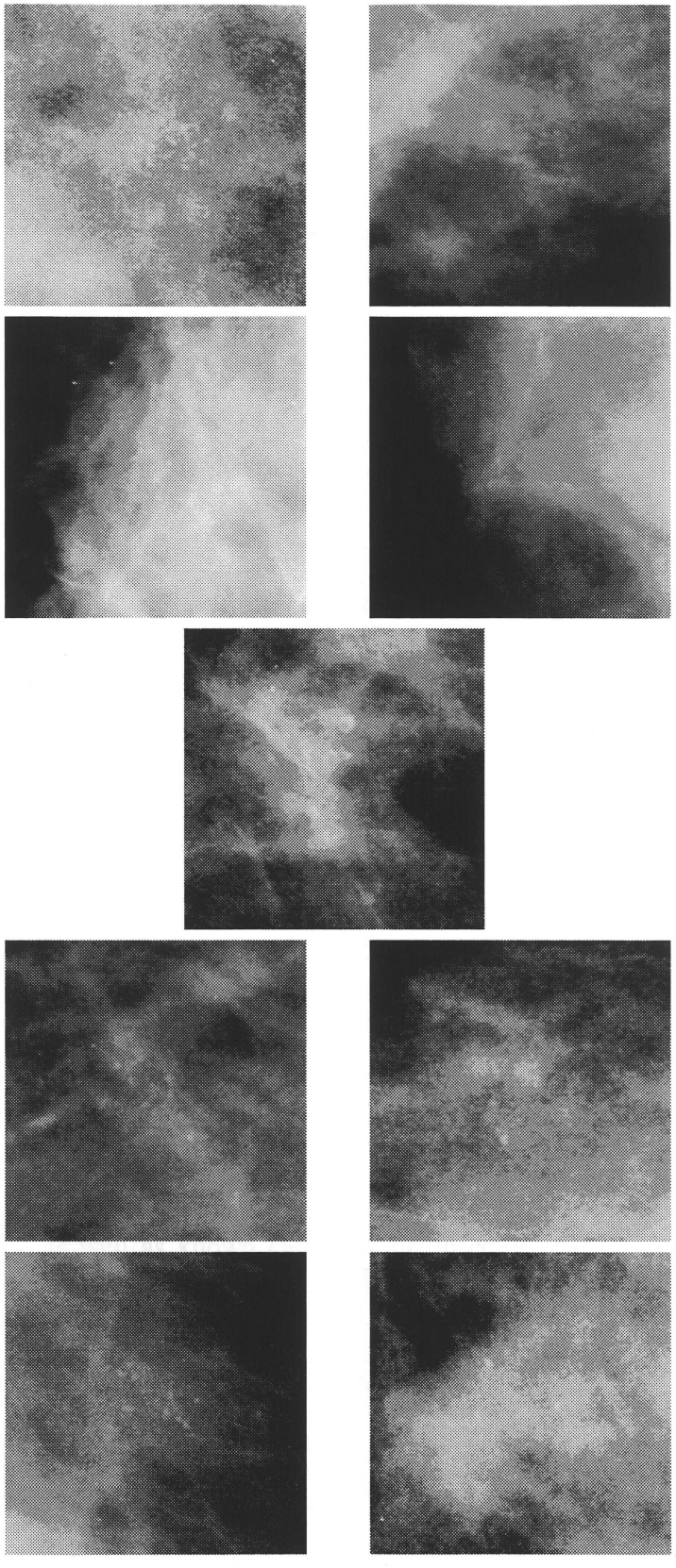


Figure 5.9: Malignant unknown lesion (center) with the selected benign (upper) and malignant (lower) similar lesions. Six of the eight observers increased their confidence levels of malignancy beneficially after viewing the similar images. The other two hardly changed their confidence levels.

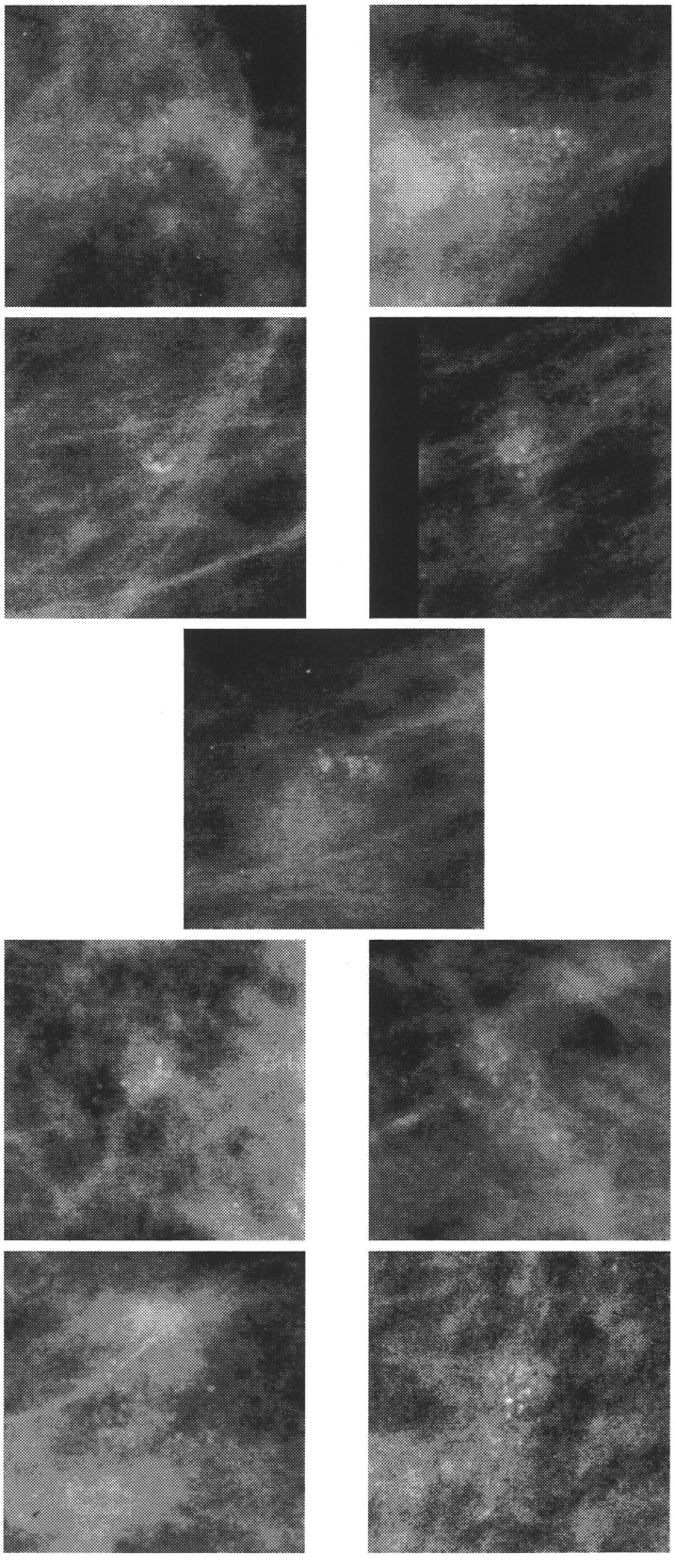


Figure 5.10: Benign unknown lesion (center) with the selected benign (left) and malignant (right) similar lesions. Although the three breast imaging fellows beneficially increased or hardly changed their confidence levels of malignancy after viewing the similar images, four of the five attending breast radiologists decreased their confidence levels of malignancy detrimentally.

to have considered that a particular benign similar image was more similar to the unknown case than any other similar lesions.

Figure 5.10 shows the unknown benign case with the most detrimental change from the average confidence level of 0.270 to 0.354 by use of similar images. In this case, although the three breast imaging observers beneficially changed or hardly changed their confidence levels of malignancy after viewing the similar images, four of the five attending breast radiologists changed their confidence levels of malignancy detrimentally. In this unknown benign case and the unknown malignant case, both having the most detrimental change, some observers might have considered that a few similar images with pathology opposite to the unknown case were more similar. The average confidence levels of these similar images with the opposite pathology by the three attending breast radiologists were between 0.4 and 0.6 in Fig. 5.4, which implies that these similar cases should be considered malignant-looking benign and benign-looking malignant lesions, and thus they should be removed from the database and not be used for similar images in the future.

5.3 Discussion

In this study, the results indicated clearly that radiologists' performance in the differential diagnosis of clustered microcalcifications can be improved by use of similar images. However, malignant-looking benign and benign-looking malignant lesions were not used for unknown cases as well as known cases in our observer study, because these lesions would not contribute to the understanding of the potential usefulness of similar images for the diagnosis. We believe that malignant-looking benign and benign-looking malignant unknown lesions would not be helped at present by use of similar images and/or by any other computerized schemes. In clinical practice, malignant-looking benign lesions would be considered malignant, and would be treated as such. On the other hand, benign-looking malignant lesions would be considered benign, and thus not be subjected to additional

examinations; these lesions may be detected in follow-up examinations. If these lesions were subjected to additional examinations, many unnecessary examinations would result for benign lesions, which would be costly to patients and to society.

If malignant-looking benign and benign-looking malignant lesions were presented as similar images selected from known cases in the database, radiologists would be confused because some "known" benign lesions appear "malignant", and some "known" malignant lesions appear "benign". Therefore, we believe strongly that malignant-looking benign and benign-looking malignant lesions should not be used in observer studies for evaluating the usefulness of similar images. Although we do not know the fraction of malignant-looking benign and benign-looking malignant lesions included in clinical practice at present, about 20% of cases in our study, in which all lesions underwent biopsy, were identified as malignant-looking benign and benign-looking malignant lesions. If the fraction of these lesions would not be very low in clinical practice, radiologists would miss many cancers [11, 16], and unnecessary biopsy would be very common [18, 19].

In our study, we assumed that benign lesions with average confidence levels greater than 0.6 would be considered malignant-looking benign lesions, whereas malignant lesions with average confidence levels less than 0.4 would be considered benign-looking malignant lesions. We applied these threshold values to the selection of both unknown cases and known cases. However, the threshold values used in the selection of known cases should have been different from those used in the selection of unknown cases. The database to be used for known similar lesions should include only benign and malignant lesions which most radiologists can recognize correctly as benign and malignant, respectively. Therefore, the threshold value for selecting known benign cases should be very low, such as 0.4 or lower, whereas the threshold value for known malignant cases should be very high, such as 0.6 or greater. If we had used different threshold values, it is possible that our results would have differed.

By use of MRMC ROC analysis, we analyzed the confidence levels for 35 unknown cases excluding three obviously benign and two obviously malignant cases. Here, we assumed obvious cases to be those for benign lesions with average confidence levels less than 0.3, and malignant lesions with those greater than 0.7. The average AUCs for all observers increased from 0.623 without to 0.741 with the similar images (P = .0016). Although both average AUCs for 35 unknown cases without and with the similar images were less than those for 40 unknown cases, the gain in average AUCs with similar images increased. The performance of all observers in the differential diagnosis was improved when they used similar images. These results indicated that similar images would be more useful for moderately difficult and indeterminate cases with average confidence level, from 0.3 to 0.7 than for obvious cases.

There are some limitations to this study. One limitation is that the similar images used in our observer study were selected subjectively at the final selection process by an attending breast radiologist. Because there are some variations among radiologists in their subjective judgments on similar images, it would be desirable to have subjective judgments by a number of radiologists. Although three of five attending breast radiologists who participated in the observer study had seen 560 lesions when providing the confidence level regarding the malignancy for those lesions, we believed that the recall bias would be minimal because the pathology of 560 lesions was not provided to them, and the period between the work for providing their confidence levels and the observer study was more than 3 months. Another limitation is that we used ROIs instead of whole images as unknown cases in the observer study. However, although the radiologists' performance may be improved by use of the whole image, we believe that the conclusion in this study would not be changed. In addition, magnification views were not included in this study, although radiologists often use magnification images in the differential diagnosis of clustered microcalcifications. The potential usefulness of similar images for magnification

mammograms is another research area which would require further investigation.

CHAPTER 6
CONCLUSION

The motivation in this dissertation was to improve the diagnostic accuracy and consistency of radiologists' image interpretation. The goal was to develop computerized analysis for both detection aid and differentiation aid of clustered microcalcifications on mammograms.

For computerized analysis for detecting clustered microcalcifications, we constructed a novel filter bank by introducing the concept of a Hessian matrix into a commonly used filter bank. This filter bank has three important features: i) it can enhance the nodular component in an image; ii) it can enhance the nodular and linear component in an image; and iii) it can reconstruct an original image from decomposed subimages of the original image. The objective features for the nodular structures and the linear structures with various sizes in each ROI were obtained by using features i) and ii) of the filter bank. We showed that there were differences in these objective features among abnormal ROIs with clustered microcalcifications, normal ROIs with blood vessels, and normal ROIs without blood vessels. We demonstrated that the computerized detection method based on the classification results among three ROIs had a high detection performance.

For computerized analysis for identifying histological classifications of clustered microcalcifications, six objective features on clustered microcalcifications on each of follow-up magnification mammograms were determined from the microcalcifications segmented by the filter bank defined in chapter 2. We showed these objective features were useful statistically for the distinction between five different types of histological classifications. In identifying histological classification of clustered microcalcifications, the histological classification of an unknown new case in question was assumed to be the same as that of the nearest neighbor case which has the shortest Euclidean distance in a feature-space. The feature-spaces for the nearest neighbor case consisted of six objective

features obtained from the previous magnification mammogram (previous features), six objective features obtained from the current magnification mammogram (current features), and the set of the six previous features and the six current features. The classification accuracies were the highest in the feature-space with set of the six previous features and the six current features. This result indicated that the differences in not only image features but also growth speeds were useful in identifying the histological classifications. The computerized identification of histological classifications based on the nearest neighbor case also had high classification accuracies. This result would verify the assumption for identifying histological classification.

For computerized analysis for retrieving lesions similar to unknown lesions on mammograms, we investigated four objective similarity measures. Measures A and B were based on the Euclidean distance in feature space and the psychophysical similarity measure, respectively. Measure C was the sequential combination of B and A, whereas measure D was the sequential combination of A and B. In this study, we selected 100 lesions each for masses and clustered microcalcifications randomly from our database, and we selected five pairs of lesions from 4,950 pairs based on all combinations of the 100 lesions by use of each measure. In two observer studies for 20 mass pairs and 20 calcification pairs, six radiologists compared all combinations of 20 pairs by using a two alternative forced-choice method to determine the subjective similarity ranking score which was obtained from the frequency with which a pair was considered as more similar than the other 19 pairs. In both mass and calcification pairs, pairs selected by use of measure D had the highest mean value of the average subjective similarity ranking scores. The sequential combination of the objective similarity measure based on the Euclidean distance and the psychophysical similarity measure would be useful in the selection of images similar to those of unknown lesions.

The potential usefulness of the presentation of images of lesions with known pathology

similar to an unknown lesion was evaluated in distinguishing between benign and malignant clustered microcalcifications on mammograms. A total of 20 benign and 20 malignant lesions were selected using a stratified randomization method, and it was these lesions that served as unknown cases in this observer study. For each unknown case, eight similar images of benign lesions and eight similar images of malignant lesions were preselected with a computerized scheme. From these preselected images, a breast radiologist subjectively selected the four most similar images of benign lesions and the four most similar images of malignant lesions. Five attending breast radiologists and three breast-imaging fellows participated in the observer study. Observers provided their confidence level regarding malignancy of the unknown case before and after they viewed the similar images. The results were evaluated with multireader multicase ROC analysis. For all observers, the AUCs were improved when similar images were used. The average AUC for all observers increased from 0.692 without use of similar images to 0.790 with use of similar images. The presentation of similar images can improve radiologists' performance in the differential diagnosis of clustered microcalcifications on mammograms.

In this dissertation, we have proposed a new approach to achieve high detection performance and/or high classification accuracies which would certainly help radiologists to improve their diagnosis accuracy of clustered microcalcification in a clinical setting. In this research, we used a digitized mammogram with a pixel size of 0.0435 mm x 0.0435 mm and a 12-bit gray scale by use of a laser scanner. As digital mammograms have gradually become available in more and more clinical settings, and as the number of specialists for diagnosing mammograms is limited in a community, these digital mammograms are going to be networked and analyzed at a key station located in a community. Mie University Hospital is one such key station in Mie prefecture, and our method is going to be routinely used to help radiologists. Then, there arises a new challenge for us to explore a way to treat much larger pixel sizes than those used in this

dissertation since using larger pixel sizes will greatly reduce the sizes of data and cost of communication.

APPENDIX I

Eigenvalues of Hessian Matrix

By using $x=r\cos\theta$ and $y=r\sin\theta$, the function $z=f(x,y)$ is represented by

$$\frac{\partial z}{\partial r}=\frac{\partial z}{\partial x}\cos\theta+\frac{\partial z}{\partial y}\sin\theta,$$

$$\frac{\partial z^2}{\partial^2 r}=\frac{\partial z^2}{\partial^2 x}\cos^2\theta+2\frac{\partial z^2}{\partial x\partial y}\cos\theta\sin\theta+\frac{\partial z^2}{\partial^2 y}\sin^2\theta.$$

The second derivative of the function $f(x,y)$ in an arbitrary direction θ is given by

$$\frac{\partial z^2}{\partial^2 r}=f''=x^T Hx,$$

where $x^T=(\cos\theta \quad \sin\theta)$, and H is a Hessian matrix. Since H is a symmetry matrix, there is an orthogonal matrix P such that $P^T HP=diag[\lambda_1 \quad \lambda_2]$. λ_1 and λ_2 ($\lambda_1 \geq \lambda_2$) indicate the small eigenvalue and the large eigenvalue of the Hessian matrix, respectively. If we change the basis with $x=Py$, the second derivative of the function $f(x,y)$ is

$$f''=y^T P^T HPy=\lambda_1 y_1^2+\lambda_2 y_2^2.$$

By using the relation

$$y_1^2+y_2^2=y^T y=(P^T x)^T P^T x=x^T x=\cos^2\theta+\sin^2\theta=1,$$

it becomes

$$f''=\lambda_1(y_1^2+y_2^2)+(\lambda_2-\lambda_1)y_2^2=\lambda_1+(\lambda_2-\lambda_1)y_2^2\geq\lambda_1.$$

Therefore, the small eigenvalue λ_1 of the Hessian matrix is given by the smallest value of f''. By using the same relation, we have

$$f''=\lambda_2(y_1^2+y_2^2)+(\lambda_1-\lambda_2)y_1^2=\lambda_2+(\lambda_1-\lambda_2)y_1^2\leq\lambda_2.$$

The large eigenvalue λ_2 of the Hessian matrix is given by the largest value of f''.

APPENDIX II

Modified Bayes Discriminant Function

The relationship between the covariance matrix ${}_l\Sigma$ of each class l based on histological classifications, its i-th eigenvalue ${}_l\lambda_i$ $({}_l\lambda_i \geq {}_l\lambda_{i+1})$, and its i-th eigenvector ${}_l\Phi_i$ satisfy the following equation:

$$ {}_l\Sigma = \sum_{i=1}^{n} {}_l\lambda_i \, {}_l\Phi_i \, {}_l\Phi_i^t , $$

where n is the number of dimensions of the feature vector. Therefore, the Bayes discriminant function (BDF) is defined as

$$ g_0^l(x) = \sum_{i=1}^{n} \frac{(x - {}_l\mu, {}_l\Phi_i)^2}{{}_l\lambda_i} + \ln \prod_{i=1}^{n} {}_l\lambda_i . $$

Here, x and ${}_l\mu$ are the input feature vector and the mean vector of class l, respectively. In the BDF, the estimation error of the eigenvectors becomes large when the number of training samples is not large enough compared with the number of dimensions of the feature vector. In particular, the estimation error of higher-order eigenvectors is much larger than that of lower-order eigenvectors.

For solving this problem, the Modified Bayes discriminant function (MBDF) was employed for distinguishing between the five different types of histological classifications. The MBDF is given by

$$ g^l(x) = \sum_{i=1}^{k} \frac{(x - {}_l\mu, {}_l\Phi_i)^2}{{}_l\lambda_i} + \sum_{i=k+1}^{n} \frac{(x - {}_l\mu, {}_l\Phi_i)^2}{{}_l\lambda_{k+1}} + \ln\left(\prod_{i=1}^{k} {}_l\lambda_i \prod_{i=k+1}^{n} {}_l\lambda_{k+1} \right), $$

where $k\,(1 \leq k < n)$ is an integer. Here, the estimation error of higher-order eigenvectors is reduced by using ${}_l\lambda_{k+1}$ as an approximate value of ${}_l\lambda_i$ $(i = k+2, \cdots, n)$. In the case of $k = n-1$, the MBDF is equal to the BDF. When k is about one third of the number of dimensions of the feature vector, it is known to show the highest classification performance. In this study, k was given as one third of the number of dimensions of the feature vector.

REFERENCES

[1] American Cancer Society. Cancer Facts and Figures 2007. American Cancer Society, Atlanta, 2007.

[2] Ministry of Health, Labor and Welfare. Population Survey Report. Ministry of Health, Labor and Welfare, Tokyo, 2006 (in Japanese).

[3] D.A. Berry, K.A. Cronin, S.K. Plevritis, D.G. Fryback, L. Clarke, M. Zelen, J.S. Mandelblatt, A.Y. Yakovlev, J.D.F. Habbema, E.J. Feuer, "Effect of screening and adjuvant therapy on mortality from breast cancer," New Engl J Med 353:1784-1792, 2005.

[4] L. Tabar G. Fagerberg, S. W. Duffy, N. E. Day, A. Gad, O. Grontoft, "Update of the Swedish two-county program of mammographic screening for breast cancer," Radiol. Clin. North Am. 30: 187-210, 1992.

[5] S. Shapiro, W. Venet, P. Strax, L. Venet, R. Roeser, "Selection, follow-up, and analysis in the health insurance plan study: A randomized trial with breast cancer screening," J. Natl. Cancer Inst. Monogr. 67: 65-74, 1985.

[6] L.L. Humphrey, M. Helfand, B.K.S. Chan, S.H. Woolf, "Breast cancer screening: a summary of the evidence for the U.S. preventive services task force," Annals of Internal Medicine 137: E-347-E-367, 2002.

[7] D.B. Kopans. Breast imaging, second edition. Lippincott-Raven publishers, New York, 1997.

[8] T. Morimoto and M. Sasa. Atlas of Screening Mammography. Digital Press, Tokyo, 1996 (in Japanese).

[9] E.A. Sickles, "Mammographic features of early breast cancer," Am J Roentgenol 143: 461-464, 1984.

[10] E.A. Sickles, "Mammographic features of 300 consecutive nonpalpable breast

cancers," Am J Roentgenol 146: 661-663, 1986.

[11] R.E. Bird, T.W. Wallace, B.C. Yankaskas, "Analysis of cancers missed at screening mammography," Radiology 184: 613-617, 1992.

[12] S.K. Goergen, J. Evans, G.P.B. Cohen, J.H. MacMillan, "Characteristics of breast carcinomas missed by screening radiologists," Radiology 204: 131-135, 1997.

[13] B. Vitak, "Invasive interval cancers in the Ostergotland mammographic screening programme: Radiological analysis," Eur Radiol 8: 639-646, 1998.

[14] R. Taylor, R. Supramaniam, M. Rickard, J. Estoesta, C. Moreira, "Interval breast cancers in New South Wales, Australia, and comparisons with trials and other mammographic screening programmes," J Med Screen 9: 20-25, 2002.

[15] M. Muttarak, S. Pojchamarnwiputh, B. Chaiwun, "Breast carcinomas: why are they missed?" Singapore Med J 47: 851-857, 2006.

[16] H. Burhenne, L. Burhenne, F. Goldberg, T. Hislop, A.J. Worth, P.M. Rebbeck, and L. Kan, "Interval breast cancers in the screening mammography program of British Columbia: Analysis and classification," Am J Roentgenol 162: 1067-1071, 1994.

[17] D.B. Kopans, "The positive predictive value of mammography," Am J Roentgenol 158: 521-526, 1992.

[18] F.M. Hall, J.M. Storella, D.Z. Silverstone, and G. Wyshak, "Nonpalpable breast lesion: recommendations for biopsy based on suspicion of carcinoma at mammography," Radiology 167: 353-358, 1988.

[19] D.D. Adler and M.A. Helvie, "Mammographic biopsy recommendations," Current Opinion in Radiology 4: 123-129, 1992.

[20] K. Doi, M.L. Giger, H. MacMahon, "Computer-aided diagnosis: Development of automated schemes for quantitative analysis of radiographic images," Semin Ultrasound CT MRI 13: 140-152, 1992.

[21] K. Doi, H. MacMahon, S. Katsuragawa, R.M. Nishikawa, and Y. Jiang,

"Computer-aided diagnosis in radiology: potential and pitfalls," European Journal of Radiology, 31: 97-109, 1997.

[22] C.J. Vyborny, "Can computers help radiologists read mammograms?" Radiology 191: 315-317, 1994.

[23] M.L. Giger, Z. Huo, M.A. Kupinski, C.J. Vyborny, "Computer-aided diagnosis in mammography," In: The Handbook of Medical Imaging Vol.2 Medical Imaging Processing and Analysis. J.M. Fitzpatrick, M. Sonka (eds) SPIE 915-1004, 2000.

[24] H.P. Chan, K. Doi, C.J. Vyborny, K.L. Lam, and R.A. Schmidt, "Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study," Invest Radiol 23: 664-671, 1988.

[25] H.P. Chan, K. Doi, C.J. Vyborny, R.A. Schmidt, C.E. Metz, K.L. Lam, T. Ogura, Y.Z. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," Invest Radiol 25: 1102-1110, 1990.

[26] H. Li, K.J. Liu, and S. Lo, "Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms," IEEE Trans Med Imag 16: 785-798, 1997.

[27] N. Karssemeijer, "Recognition of clustered microcalcifications using a random field mode, biomedical image processing and biomedical visualization," in Proc SPIE 1905: 776-786, 1993.

[28] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms" Int J Pattern Recognit Artificial Intell 7: 1357-1376, 1993.

[29] S. Mallat, "Multifrequency channel decompositions of images and wavelet models," in Proc IEEE Trans Acoust Speech Sig 37: 2091-2110, 1989.

[30] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans Pattern Anal Machine Intell 11: 674-693, 1989.

[31] R.N. Strickland and H.I. Hahn, "Wavelet transform for detecting microcalcifications in mammograms," IEEE Trans Med Imag 15: 218-229, 1996.

[32] R.N. Strickland and H. I. Hahn, "Wavelet transform methods for objects detection and recovery," IEEE Trans Image Processing 6: 724-735, 1997.

[33] H. Yoshida, K. Doi, and R.M. Nishikawa, "Automated detection of clustered microcalcifications, in Digital Mammograms Using Wavelet Transform Techniques," in Proc SPIE 2167: 868-886, 1994.

[34] H. Yoshida, K. Doi, R.M. Nishikawa, M.L. Giger, and R.A. Schmidt, "An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms," Acad Radiol 3: 621-627, 1996.

[35] L.P. Clarke, M. Kallergi, W. Qian, H.D. Li, R.A. Clark, and M.L. Silbiger, "Three-structured nonlinear filter and wavelet transform for microcalcification segmentation in digital mammography," Cancer Lett 77: 173-181, 1994.

[36] W. Qian, L.P. Clarke, M. Kallergi, and R.A. Clark, "Three-structured nonlinear filters in digital mammography," IEEE Trans Med Imag 13: 25-36, 1994.

[37] W. Qian, L.P. Clarke, B. Zheng, M. Kallergi, and R.A. Clark, "Computer assisted diagnosis for digital mammography," IEEE Eng Med Biol 14: 561-569, 1995.

[38] A.F. Laine, S. Schuler, J. Fan, and W. Huda, "Mammographic feature enhancement by multiscale analysis," IEEE Trans Med Imag 13: 725-740, 1994.

[39] A.F. Laine, J. Fan, and W. Yang, "Wavelets for contrast enhancement of digital mammography," IEEE End Med Biol 14: 536-550, 1995.

[40] Y. Jiang, R.M. Nishikawa, D.E. Wolverton, C.E. Metz, M.L. Giger, R.A. Schmidt, C.J. Vyborny, K. Doi, "Malignant and benign clustered microcalcifications: Automated feature analysis and classification," Radiology 198: 671-678, 1996.

[41] H.P. Chan, D. Wei, L.T. Niklason, M.A. Helvie, K.L. Lam, M.M. Goodsitt, and D.D. Adler, "Computer-aided classification of malignant/benign microcalcifications in

mammography," Med Phys 21: 875, 1994.

[42] H.P. Chan, B. Sahiner, N. Petric, M.A. Heavie, K.L. Lam, D.D. Adler and M.M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network," Phys Med Biol 42: 549-567, 1997.

[43] H.P. Chan, B. Sahiner, K.L. Lam, N. Petric, M.A. Helvie, M.M. Goodsitt, and D.D Adler, "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," Med Phys 25: 2007-2019, 1998.

[44] R. Nakayama, Yoshikazu Uchiyama, Koji Yamamoto, Ryoji Watanabe, Kiyoshi Namba, Kakuya Kitagawa, Kan Takeda, "Discrimination of Malignant and Benign Micro-calcification Clusters on Mammograms," Japanese Journal of Computer Aided Diagnosis of Medical Images 3: 1-7, 1999.

[45] R. Nakayama, Yoshikazu Uchiyama, Isamu Hatsukade, Koji Yamamoto, Ryoji Watanabe, Kiyoshi Namba, Kakuya Kitagawa, Kan Takeda, "Computerized Discrimination of Malignant and Benign Microcalcification Clusters on Mammogram," Japanese Journal of Radiological Technology 56: 391-397, 2000 (in Japanese).

[46] L. Shen, R.M. Rangayyan, and J.E.L. Desautels, "Application of shape analysis to mammographic calcifications," IEEE Trans Med Imaging 13: 263-274, 1994.

[47] M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters," Med Phys 31: 314-326, 2004.

[48] I. Leichter, R. Lederman, S.S. Buchbinder, P. Bamberger, B. Novak, S. Fields, "Computerized evaluation of mammographic lesions: What diagnostic role does the shape of the individual microcalcifications play compared with the geometry of the cluster?" Am J Roentgenol 182: 705-712, 2004.

[49] Y. Jiang, R.M. Nishikawa, R.A. Schmidt, C.E. Metz, M.L. Giger, and K. Doi,

"Improving breast cancer diagnosis with computer-aided diagnosis," Acad Radiol 6: 22-33, 1999.

[50] G. Sakamoto and S. Haga. Fundamental and clinic of ductal carcinoma in situ. Shinoharashinsha, Tokyo, 2001 (in Japanese).

[51] M. Heath, K. Bowyer, D. Kopans, R. Moore, P. Kegelmeyer Jr, "Current states of the digital database for screening mammography," Digital mammography Kluwer Academic Publishers, Dordrecht, Holland, 1998.

[52] A. Shimizu, J. Hasegawa, and J. Toriwaki, "Characteristics of minimum directional difference filter which extracts circumscribed shadows in chest x-ray images," Systems and Computers in Japan 25: 56-66, 1994.

[53] A. Shimizu, J. Hasegawa, and J. Toriwaki, "Characteristics of rotatory second order difference filter for computer aided diagnosis of medical images," Systems and Computers in Japan 26: 38-51, 1995.

[54] Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans." Med Phys 30: 2040-2051, 2003.

[55] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, "Three dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," Med Image Anal 2: 143-168, 1998.

[56] Y. Sato, C. F. Westin, A. Bhalerao, S. Nakajima, N. Shiraga, S. Tamura, and R. Kikinis, "Tissue classification based on 3D local intensity structures for volume rendering," IEEE Trans Vis Comput Graph 6: 160-180, 2000.

[57] G. Strang and T. Nguyen. Wavelets and Filter Banks. Wellesley-Cambridge, MA, 1996.

[58] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. Wiley & Sons, Inc., 2001.

[59] C.E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies" Invest Radiol 24: 234-245, 1989.

[60] R.A. Johnson and D.W. Wichern. Applied Multivariate Statistical Analysis. Englewood Cliffs , NJ, 1992.

[61] R.C. Gonzales and R. E. Woods. Digital Image Processing. Addison-Wesley, MA, 1992.

[62] J. Sera. Image Analysis and Mathematical Morphology. Academic Press, London, 1982.

[63] R. Nakayama, R. Watanabe, K. Namba, K. Yamamoto, K. Takeda, S. Katsuragawa, and K. Doi, "Potential usefulness of multiple-mammographic views in computer-aided diagnosis scheme for identifying histological classification of clustered microcalcification." in Proc. of International Workshop on Digital Mammography 4046: 229-236, 2006.

[64] G. Hermann, R.J. Keller, S. Drossman, B.A. Caravella, P. Tartter, R.A. Panetta, and I.J. Bleiweiss, "Mammographic pattern of microcalcifications in the preoperative diagnosis of comedo ductal carcinoma in situ: histopathologic correlation," Can Assoc Radiol J 50: 235-240, 1999.

[65] E. Fondrinier, G. Lorimier, V. Guerin-Boblet, A.F. Bertrand, C. Mayras, and N. Dauver, "Breast microcalcifications: multivariate analysis of radiologic and clinical factors for carcinoma," World J Surg 26: 290-296, 2002.

[66] M. Aoyama, Q. Li, S. Katsuragawa, H. MacMahon, and K. Doi, "Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images," Med Phys 29: 701-708, 2002.

[67] T. Sergios, K. and Konstantinos. Pattern Recognition. Academic Press, London, 2003.

[68] F. Kimura, K. Takahashi, S. Tsuruoka, Y. Miyake, "On avoiding peaking phenomenon of the quadratic discriminant function," IEICE J-69-D: 1328-1334, 1986.

[69] S. Tsuruoka, M. Kurita, T. Harada, F. Kimura, Y. Miyake, "Handwriting 'KANJI' and 'HIRAGANA' character recognition using weighted direction index histogram method," IEICE J-70-D: 1390-1397, 1987.

[70] S. Tsuruoka, H. Morita, F. Kimura, Y. Miyake, "Handwritten character recognition adaptable to the writer," IEICE J-70-D: 1953-1960, 1987.

[71] C.P. Langlotz, "Fundamental measures of diagnostic examination performance: Usefulness for clinical decision making and research," Radiology 228: 3-9, 2003.

[72] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," Br J Radiol 78:s3-s19, 2005.

[73] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," Comput Med Imaging Graph 31:198-211, 2007.

[74] K. Doi, "Computer-aided diagnosis moves from breast to other systems," Diagnostic Imaging 37-40, 2007.

[75] S. Kumazawa, C. Muramatsu, Q. Li, F. Li, J. Shiraishi, P. Caligiuri, R.A. Schmidt, H. MacMahon, K. Doi, "An investigation of radiologists' perception of lesion similarity: Observations with paired breast masses on mammograms and paired lung nodules on CT images," Acad Radiol 15:887-894, 2008.

[76] G. Bucci, S. Cagnoni, R. De Dominicis, "Integrating content-based retrieval in a medical image reference database," Comput Med Imaging Graph 20:231–241, 1996.

[77] S.T. Wong, H.K. Huang, "Design methods and architectural issues of integrated medical image data base systems," Comput Med Imaging Graph 20:285–299, 1996.

[78] H.A. Swett, P.R. Fisher, A.I. Cohn, P.L. Miller, P.G. Mutalik, "Expert system controlled image display," Radiology 172:487-493, 1989.

[79] H.A. Swett, P.G. Mutalik, V.P. Neklesa, L. Horvath, C. Lee, J. Richter, I. Tocino, P.R. Fisher, "Voice-activated retrieval of mammography reference images," J Digit Imaging 11:65-73, 1998.

[80] H. Qi, W.E. Snyder, "Content-based image retrieval in picture archiving and communications systems," J Digit Imaging 12(2 Suppl 1):81-83, 1999.

[81] J. Sklansky, E.Y. Tao, M. Bazargan, C.J. Ornes, R.C. Murchison, S. Teklehaimanot, "Computer-aided, case-based diagnosis of mammographic regions of interest containing microcalcifications," Acad Radiol 7:395-405, 2000.

[82] M.L. Giger, Z. Huo, C.J. Vyborny, "Intelligent CAD workstation for breast imaging using similarity to known lesions and multiple visual prompt aids," Proc. of SPIE 4684:768–773, 2002.

[83] U. Sinha U, H. Kangarloo, "Principal component analysis for content-based image retrieval," Radiographics 22:1271–1289, 2002.

[84] A.M. Aisen, L.S. Broderick, H. Winer-Muram, C.E. Brodley, A.C. Kak, C. Pavlopoulou, J. Dy, C.R. Shyu, A. Marchiori, "Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment," Radiology 228:265-270, 2003

[85] Y. Kawata, N. Niki, H. Ohmatsu, N. Moriyama, "Example-based assisting approach for pulmonary nodule classification in three-dimensional thoracic computed tomography images," Acad Radiol 10:1402-1415, 2003.

[86] I. El-Naqa, Y. Yang, N.P. Galatsanos, R.M. Nishikawa, M.N. Wernick, "A similarity learning approach to content-based image retrieval: application to digital mammography," IEEE Trans Med Imaging 23:1233-1244, 2004.

[87] B. Zheng, A. Lu, L.A. Hardesty, J.H. Sumkin, C.M. Hakim, M.A. Ganott, D. Gur, "A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment," Med Phys 33:111-117, 2006.

[88] R. Nakayama, R. Watanabe, K. Namba, K. Takeda, K. Yamamoto, S. Katsuragawa, K. Doi, "An improved computer-aided diagnosis scheme using the nearest neighbor criterion for determining histological classification of clustered microcalcifications,"

Methods Inf Med 46:716-722, 2007.

[89] Q. Li, F. Li, J. Shiraishi, S. Katsuragawa, S. Sone, K. Doi, "Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules," Med Phys 30:2584-2593, 2003.

[90] C. Muramatsu, Q. Li, K. Suzuki, R.A. Schmidt, J. Shiraishi, G.M. Newstead, K. Doi, "Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary results," Med Phys 32:2295-2304, 2005.

[91] C. Muramatsu, Q. Li, R.A. Schmidt, K. Suzuki, J. Shiraishi, G.M. Newstead, K. Doi, "Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: observer study results," Med Phys 33:3460-3468, 2006.

[92] C. Muramatsu, Q. Li, R.A. Schmidt, J. Shiraishi, K. Suzuki, G.M. Newstead, K. Doi, "Determination of subjective similarity for pairs of masses and pairs of clustered microcalcifications on mammograms: comparison of similarity ranking scores and absolute similarity ratings," Med Phys 34:2890-2895, 2007.

[93] C. Muramatsu, Q. Li, R.A. Schmidt, J. Shiraishi, K. Doi, "Investigation of psychophysical similarity measures for selection of similar images in the diagnosis of clustered microcalcifications on mammograms," Med Phys 35:5695-5702, 2008.

[94] C. Muramatsu "Investigation of similarity measures for selection of similar images in computer-aided diagnosis of breast lesions on mammograms," Ph.D. Dissertation, The University of Chicago, Chicago, IL. (ProQuest/UMI, Ann Arbor, MI, 2008)

[95] M. Kendall, J.D. Gibbons. Rank correlation methods, 5th ed. New York, Oxford University Press, 1990

[96] K. Horsch, M.L. Giger, C.J. Vyborny, L. Lan, E.B. Mendelson, R.E. Hendrick, "Classification of breast lesions with multimodality computer-aided diagnosis:

observer study results on an independent clinical data set," Radiology 240: 357-368, 2006.

[97] T. Kobayashi, X.W. Xu, H. MacMahon, C.E. Metz, K. Doi, "Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs," Radiology 199: 843-848, 1996.

[98] T. Uozumi, K. Nakamura, H. Watanabe, H. Nakata, S. Katsuragawa, K. Doi K, "ROC analysis of detection of metastatic pulmonary nodules on digital chest radiographs with temporal subtraction," Acad Radiol 8: 871-878, 2001.

[99] D.D. Dorfman, K.S. Berbaum, C.E. Metz, "Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method," Invest Radiol 27: 723-231, 1992.

[100] S.L. Hillis, K.S. Berbaum, C.E. Metz, "Recent developments in the Dorfman Berbaum Metz procedure for multireader ROC study analysis," Acad Radiol 15: 647-661, 2008.

LIST OF RELATED PUBLICATIONS

Original Papers

[1] Ryohei Nakayama, Hiroyuki Abe, Junji Shiraishi, Kunio Doi, "Evaluation of Objective Similarity Measures for Selecting Similar Images of Mammographic Lesions," Journal of Digital Imaging (in press).

[2] Ryohei Nakayama, Hiroyuki Abe, Junji Shiraishi, Kunio Doi, "Potential Usefulness of Similar Images in the Differential Diagnosis of Clustered Microcalcifications on Mammograms," Radiology, vol.253, no.3, pp625-631, 2009.

[3] Ryohei Nakayama, Ryoji Watanabe, Kiyoshi Namba, Kan Takeda, Koji Yamamoto, Shigehiko Katsuragawa, and Kunio Doi, "An Improved Computer-aided Diagnosis Scheme Using the Nearest Neighbor Criterion for Determining Histological Classification of Clustered Microcalcifications," Method Inform Med., vol.46, no.6, pp.716-722, Dec. 2007.

[4] Ryohei Nakayama, Ryoji Watanabe, Kiyoshi Namba, Kan Takeda, Koji Yamamoto, Shigehiko Katsuragawa, and Kunio Doi, "Computer-Aided Diagnosis Scheme for Identifying Histological Classification of Clustered Microcalcifications by Use of Follow-up Magnification Mammograms," Acad Radiol., vol.13, no.10, pp.1219-1228, Oct. 2006.

[5] Ryohei Nakayama, Yoshikazu Uchiyama, Koji Yamamoto, Ryoji Watanabe, and Kiyoshi Namba, "Computer-Aided Diagnosis Scheme Using a Filter Bank for Detection of Microcalcification Clusters in Mammograms," IEEE Trans Biomed Eng., vol.53, no.2, pp.273-83, Feb. 2006.

International Conference (Proceedings)

[1] Akiyoshi Hizukuri, Ryohei Nakayama, Nobuo Nakako, Naoki Nagasawa, Shigeki

Kobayashi, Kan Takeda, Shinji Tsuruoka, "Introduction of Computerized Detection Method to Breast Cancer Screening in Mie Prefecture," Proc. of the Second International Workshop on Regional Innovation Studies - Biomedical Engineering - (IWRIS2010), Tsu, pp.17-20, Oct. 2010.

[2] Akiyoshi Hizukuri, Ryohei Nakayama, Nobuo Nakako, Takanori Ogino, Hiroharu Kawanaka, Haruhiko Takase, Shinji Tsuruoka, "Computerized Identification Method for Current Dental Radiographs Based on the Arranged Previous Dental Radiographs," Proc. of IWRIS2010, Tsu, pp.21-24, Oct. 2010.

[3] Akiyoshi Hizukuri, Ryohei Nakayama, Toshihito Kawamura, Takanori Ogino, Shinji Tsuruoka, "Computerized Segmentation Method of Calcifications within Clustered Microcalcifications on Mammogram," Proc. of the First International Workshop on Regional Innovation Studies - Biomedical Engineering - (IWRIS2009), Tsu, pp.47-50, Oct. 2009.

[4] Toshihito Kawamura, Ryohei Nakayama, Takanori Ogino, Akiyoshi Hizukuri, Hiroharu Kawanaka, Shinji Tsuruoka, "Computerized Extraction Method of Hepatic Vessels in Contrasted Abdominal X-ray CT Images," Proc. of IWRIS2009, Tsu, pp.51-54, Oct. 2009.

[5] Takanori Ogino, Ryohei Nakayama, Toshihito Kawamura, Akiyoshi Hizukuri, Kan Takeda, "Computer-aided Detection Scheme for Sentinel Lymph Nodes on Lymphoscintigrams," Proc. of IWRIS2009, Tsu, pp.43-46, Oct. 2009.

[6] Toshihito Kawamura, Ryohei Nakayama, Yosuke Mizutani, Kan Takeda, Hiroharu Kawanaka, Takahiro Takada, Koji Yamamoto, and Tsuruoka Shinji, "Computer-aided Diagnosis Scheme for Detection of Mass with Bilateral Subtraction Technique in Mammograms," Proc. of Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems (SCIS & ISIS 2008), Nagoya, pp.1853-1858, Sep. 2008.

[7] Toshihito Kawamura, Ryohei Nakayama, Kan Takeda, Hiroharu Kawanaka, Takahiro Takada, Koji Yamamoto, and Shinji Tsuruoka, "Development of Computerized Identification Method for Nipple in Mammograms," Proc. of SCIS & ISIS 2008, Nagoya, pp.1847-1852, Sep. 2008.

[8] Takanori Ogino, Ryohei Nakayama, Toshihito Kawamura, Takahiro Takada, Koji Yamamoto, Naoki Nagasawa, and Kan Takeda, "Computerized Scheme for Detecting Sentinel Lymph Nodes on Lymphoscintigram using Subtraction Technique based on Symmetry of Mapped Injection Site," Proc. of SCIS & ISIS 2008, Nagoya, pp.1995-1998, Sep. 2008.

[9] Ryohei Nakayama, Ryoji Watanabe, Kiyoshi Namba, Koji Yamamoto, Kan Takeda, Shigehiko Katsuragawa, and Kunio Doi, "Potential Usefulness of Multiple-Mammographic Views in Computer-aided Diagnosis Scheme for Identifying Histological Classification of Clustered Microcalcification," Proc. of International Workshop on Digital Mammography 2006, pp.229-236, June 2006.

International Conference (Abstract)

[1] Ryohei Nakayama, Yumi Kashikura, Kiyoshi Namba, Shigeki Kobayashi, Kan Takeda, Tomoko Ogawa, "Computer-aided Diagnosis Scheme for Determining Histological Classifications of Breast Masses on Ultrasonographic Image," Radiological Society of North America 2010 95st Scientific Assembly and Annual Meeting (RSNA2010), Chicago, Dec. 2010, (in press).

[2] Nobuo Nakako, Takahiro Takada, Ryohei Nakayama, Naoki Nagasawa, Shigeki Kobayashi, Kan Takeda, "Construction of PACS under the IPV6 Environment," RSNA2010, Chicago, Dec. 2010, (in press).

[3] Ryohei Nakayama, Akiyoshi Hizukuri, Toshihito Kawamura, Takanori Ogino, Naoki Nagasawa, Kan Takeda, "Computer-Aided diagnosis (CAD) scheme for detection of

sentinel lymph nodes on lymphoscintigrams based on symmetry of mapped injection site." Computer Assisted Radiology and Surgery (CARS) 2010, Geneva, June 2010, pp.S379

[4] Akiyoshi Hizukuri, Ryohei Nakayama, Toshihito Kawamura, Kan Takeda, Shinji Tsuruoka, "Computerized segmentation method of calcifications within clustered microcalcifications on mammograms using multiresolution analysis," CARS2010, Geneva, June 2010, pp.S375

[5] Ryohei Nakayama, Hiroyuki Abe, Junji Shiraishi, Kunio Doi, "Improved Differential Diagnosis of Clustered Microcalcifications on Mammograms with Use of Similar Images: Observer Study Result," Radiological Society of North America 2009 94st Scientific Assembly and Annual Meeting (RSNA2009), Chicago, Dec. 2009, pp.606

[6] Naoki Nagasawa, Takanori Ogino, Ryohei Nakayama, Toshihito Kawamura, Takahiro Takada, Kan Takeda, "Performance Evaluation of Computerized Scheme for Identifying Sentinel Lymph Node Using Subtraction Technique on Lymphoscintigram," The 2009 SNM Annual Meeting, Toronto, June 2009.

[7] Koji Yamamoto, Ryohei Nakayama, Ryoji Watanabe, Takahiro Takada, and Kan Takeda, "Computer-aided diagnosis scheme for detection of architectural distortion on mammograms using multiresolution analysis," CARS2008, Barcelona, June 2008.

[8] Ryohei Nakayama, Ryoji Watanabe, Kiyoshi Namba, Koji Yamamoto, Kan Takeda, Shigehiko Katsuragawa, and Kunio Doi, "Computer-aided Diagnosis (CAD) Scheme for Determining Histological Classification of Clustered Microcalcifications on Magnification Mammograms Based on Nearest Neighbour Case," RSNA2007, Chicago, Nov. 2007, pp.979 .

[9] Ryohei Nakayama, Ryoji Watanabe, Kiyoshi Namba, Koji Yamamoto, Kan Takeda, Shigehiko Katsuragawa, and Kunio Doi, "Computer-aided Diagnosis (CAD) Scheme for Determining Histological Classification of Clustered Microcalcifications Based on

the Nearest Neighbour Criterion in the Follow-up Feature Space," CARS2007, Berlin, June 2007, pp.513.

[10] Ryohei Nakayama, Ryoji Watanabe, Kiyoshi Namba, Koji Yamamoto, Kan Takeda, Shigehiko Katsuragawa, and Kunio Doi, "Computer-aided Diagnosis (CAD) Scheme for Identifying Histological Classification of Clustered Microcalcifications Based on the Similarity Measure in Feature-Space," RSNA2006, Chicago, Nov. 2006, pp.283.

[11] Ryohei Nakayama, Ryoji Watanabe, Kiyoshi Namba, Koji Yamamoto, Kan Takeda, Shigehiko Katsuragawa, and Kunio Doi, "Computerized Analysis for Identifying Histological Classification of Clustered Microcalcifications in Follow-up Magnification Mammograms," RSNA2006, Chicago, Nov. 2006, pp.779.