[Original Paper]

Performance Evaluation of a CAD System for Detecting Masses on Mammograms by Using the MIAS Database

Shenglan LI[†], [*], Takeshi HARA[†], Yuji HATANAKA[†], Hiroshi FUJITA[†],
Tokiko ENDO[††] and Takuji IWASE[†††]

[†] Department of Information Science, Faculty of Engineering, Gifu University
Yanagido 1-1, Gifu-shi 501-1193, Japan
[††] Department of Radiology, National Hospital of Nagoya
Sannomaru 4-1-1, Naka-ku, Nagoya-shi 460-0001, Japan
[†††] Department of Breast Surgery, Aichi Cancer Center Hospital
Kanokoden 1-1, Chikusa-ku, Nagoya-shi 464-8681, Japan
[*] S. Li is presently with N-TECH Company, Nakazone 520, Ogaki-shi 503-0986, Japan
(Received May 21, 2001, in final form, June 19, 2001)

Abstract : Since the evaluation is influenced by the case selection, it is necessary to apply some common databases in the performance evaluation. To evaluate our computer-aided diagnosis (CAD) system for detecting masses, 320 images from the Mammogaphy Image Analysis Society (MIAS) database in UK were applied in this study. Our algorithm for detecting masses was based on a standard adaptive thresholding technique which had been developed in our group. However, the preliminary result was not as well as those for a Japanese database. After we adjusted some thresholding values of our system, a 90% sensitivity with 0.8 false positive (FP) was achieved which indicated that our scheme was effective for the different databases in both Japan and UK. Moreover, the differences between the MIAS and a Japanese database were also discussed.

Key words : Mass, Mammogram, Computer-aided diagnosis, Performance study

1. Introduction

Mammography is considered to be a major significant way for detecting abnormalities in breast as early as possible. Breast cancer commonly presents as a mass. Currently there are a number of research groups who have been developing CAD systems to detect masses on mammograms. There are two general approaches achieved in mammographic mass detection and analysis : one is a single-image segmentation and the other is a bilateral image subtraction.

In the first approach, several techniques based on computerized feature extraction have been employed. Lai et al. presented a method by detecting circumscribed masses. They reported that their sensitivity for masses was 100% with an average of 1.7 false positive (FP) per mammogram [1]. Li et al. used a technique based on adaptive threshold. Their results with a database of 95 images indicated that a sensitivity of 90% was achieved at the expense of two FPs per image [2]. A

new rubber band straightening transform (RBST) was introduced. The classification accuracy for masses described by an area (Az) of 0.94 under the ROC curve [3].

The second approach is based on the analysis of the symmetry between both sides of mammograms. Yin et al. developed a nonlinear bilateral subtraction technique to identify asymmetries between the right and left breast images. They correctly distinguished 95% masses from 46 pairs of mammograms with three FPs per image [4]. Different to Yin's approach, Mendez et al. have characterized the asymmetries by using only one thresholding instead of more complex methods of linking multiple subtracted images. A true positive (TP) rate of 71% was achieved at an average number of 0.67 FP per image [5].

An automated CAD system for detecting masses and clustered microcalcifications in digital mammograms has been developed in our group. A series of results for detecting masses has been reported by using Japanese databases [6-8]. So far, the detection result reached a sensitivity of 91 % at an average of 0.9 FP per image for a testing database. In addition, a mass-classification system for mammograms has been developing in our group. It was reported that the classification sensitivity was 85.7% and the specificity was 81.8% [9].

In 1994, the result in the paper of Nishikawa [10] showed that the accuracy of a CAD system was depended on the case selection. As a possible long-term solution for dealing with this problem, a common database was suggested to evaluate CAD schemes.

Responding to Nishikawa's suggestion, our group has tried to find some common databases to examine the performance of our CAD system for detecting masses. Several years ago, as the first step we used the CADM database for our

evaluation that was produced by Japan Society of Computer Aided Diagnosis of Medical Images. It was concluded that our scheme was effective for Fuji computed radiography (FCR) images because of the high detection performance. Furthermore, we had been looking forward to evaluating our CAD system by some databases from overseas such as the database produced by the Mammographic Image Analysis Society (MIAS) in UK 1994.

In this paper, we employed the MIAS database to evaluate the performance of our CAD scheme for detecting masses on mammograms. Since the performance of our system to detect the clustered microcalcifications on the MIAS database has been demonstrated in our previous paper with a sensitivity of 95.8% at the number of FPs as 1.84 clusters per image [11], we will only discuss the detection ability for masses in below.

2. MIAS database

The MIAS database contained 161 pairs of mammograms selected carefully from the United Kingdom National Breast Screening Programme. Each of the mammograms was obtained from the medio-lateral oblique view and was digitized by a Joyce-Loeble SCANDIG-3 microdensitometre at a spatial resolution of 50-μm sampling distance with an 8-bit density resolution. The linear response of scanning microdensitometre was in the range from 0.0 to 3.2 optical densities. There were four sizes of image used, depending on the breast sizes: small, medium, large and extra large in the MIAS database.

There were 118 abnormal and 204 normal mammograms in the database. Classes of abnormality are presented as calcification; well-defined and circumscribed masses; spiculated masses; other, ill-defined masses; architectural distortion;
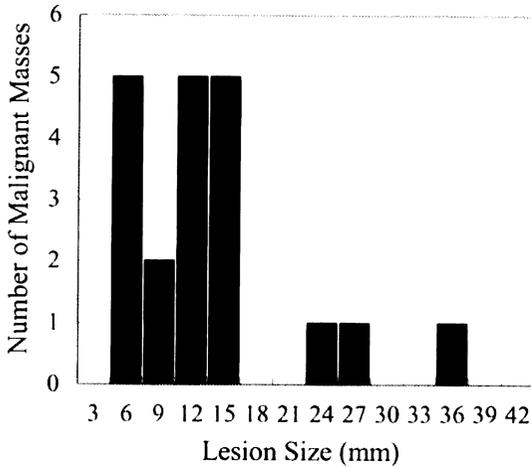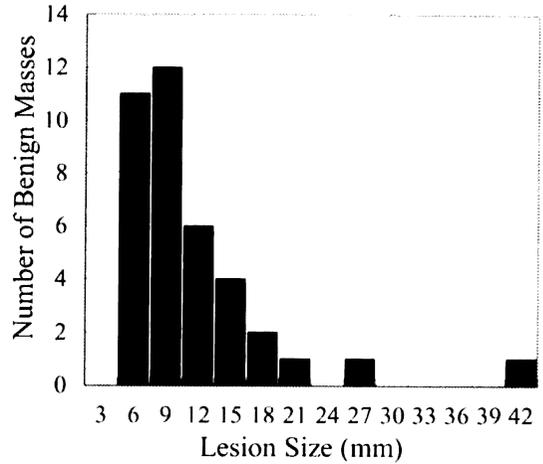
Fig. 1 Histograms of size in radius for malignant (a) and benign (b) masses in the MIAS database. The mean size for malignant was 12.2mm, and for benign was 9.8mm. It should be noted that the values of vertical coordinate in (a) and (b) were different.

and asymmetry. All abnormal mammograms had been biopsy-proven. There are 58 masses (20 malignants) in 55 images. The mammograms in database had been classified depending on the characteristic of background tissues into three categories, fatty, fatty-glandular and dense [12].

We calculated the histograms of 58 mass sizes that were shown for malignant (a) and benign (b) in Fig. 1, respectively. The average approximate radius of the masses were 12.2mm for malignant as well 9.8mm for benign. From above histogram, it was not difficult to find that some of malignant masses were very obvious, but some were rather subtle.

In this study, considering that applying part of common database would still cause the evaluating problem, we employed 320 out of all 322 mammograms (except two images which were not included in our type) from the MIAS database that was rather different with Japanese databases. Until now, such kind of research has never done by

any other groups. For instance, there were only 37, 60 and 14 mammograms from the MIAS database used in the studies of Rangayyan et al. [13], Brake et al. [14] and Kok-Wiles et al. [15], respectively. It was no doubt that applying nearly all cases from a rather different common database would make the detection more difficult than those evaluations already done by other groups before. However, we believe that the advantage for such evaluation will make our result more comparable as well as more objectively.

In addition, it should be noted that since the number of malignant masses (those are our main interest in this study) in the MIAS database was as less as 20 in all of 322 images, it was impossible for us to separate the MIAS database into two parts, in which one for training and the other for testing. In below, we provided two results by employing the MIAS database both as a training set and as a testing set.

3. Method of CAD scheme

Our scheme, which was based on single-image segmentation, has developed by employing a standard adaptive thresholding technique in the field of image processing and analysis [6-8, 16]. In the process of eliminating FPs, a method by comparing both right and left images was employed as well [17].

There were 7 main stages developed in our algorithm, which were (1) image input and digitization ; (2) extraction of breast area ; (3) classification and segmentation of breast region ; (4) extraction of suspicious area ; (5) re-analysis of detected candidate region ; (6) elimination of FPs ; (7) indication of detected masses.

3.1 Image input and digitization

Each of the mammograms in the MIAS database was first compressed to an image with a pixel size of 0.4mm×0.4mm and was linearly converted to 12-bit pixel values.

3.2 Extraction of breast area

The border of skin line in each mammogram was extracted by investigating the change of density profile. In this technique, not only the differences of the image density but also the weight of location information of breast border were considered [18].

3.3 Classification and segmentation of breast region

First, by using the characteristics of Sobel density gradient, the area of pectoralis muscle was decided. Then the gray-level histogram for each mammogram without the pectoralis muscle's area was achieved. According to the histogram, the digital mammogram was divided into glandular-fatty, fatty and dense in regard with the three

categories reported on the MIAS database. Afterwards, the segmentation was undergone especially for glandular-fatty image. Generally, the glandular-fatty images were segmented into two parts, one was the thick mammary-gland area and the other was the area on which the fatty was able to part away from mammary-gland [7, 8, 16].

3.4 Extraction of suspicious area

The low-density areas were generally considered as the first candidates for mass. A mask (4 pixels ×4 pixels) was used in each of three categorized images to find out the suspicious candidates by several different threshold values depending on the categories and the segmented parts of mammogram [7, 8].

3.5 Re-analysis of detected candidate region

Because there were a number of normal candidates in low-density areas, the general feature analysis was employed for decreasing the number of candidates as a regular step for classifying the suspicious mass.

The malignant candidates were determined repeatedly in the low-density areas by using some texture features analysis. For instance, when a candidate's circularity value was greater than the threshold, it would be classified as a true mass. Otherwise this candidate would be considered as a re-analysis candidate. For the re-analysis candidate, a new threshold was determined. According to that threshold, we transformed the re-analysis candidate to a new candidate that was defined as its pixel value changed from the lowest density to this new threshold. Then the thresholds of size, contrast, circularity and standard deviation would examine once more the new candidate. Such re-analysis procedure was iteratively performed until either the new candidates' size was smaller than the threshold for size or it was classified as a

true malignant candidate [7, 8].

3.6 Elimination of FPs

Since there was still a lot of FPs left after the general feature analysis was applied, it seemed necessarily to eliminate the FPs by using some new methods [6, 19].

There were three methods developed by our group : (1) eliminating FP masses by using second -order statistics [21], (2) decreasing the funicular -shape FPs [20], (3) decreasing the false candidates by comparing right and left mammograms [17].

3.6.1 Eliminating FP masses by using second-order statistics

We also noted that there were some differences between TP and false one in the co-occurrence matrix of the image. There were four parameters, which significantly impacted to the classification by a series of experiments. From the gray-level co -occurrence matrix, three second-order statistics values (angular second moment : ASM, inverse difference moment : IDM, and entropy : ENT) were determined. In addition, from the matrix based on the gray-level difference method, a contrast (CNT) was set.

By investigating the data from our experiment, we recognized that the quantities of TPs in ASM was smaller than those in the FPs but the magnitude of IDM, ENT and CNT in the TPs were greater than those in the FPs. The FP candidates were effectively discriminated by using such characteristics of the four parameters. As an example in a Japanese mammogram database, the number of false candidates was decreased from 4.0 to 1.9 per image at the sensitivities around 85% [20].

3.6.2 Decreasing funicular-shape FPs

There were lots of funicular-shape candidates

(nearly 1/3 in total FPs) detected by our scheme in regard to their higher circularity. In fact, they were not TPs but were either the blood vessels or parts of the mammary-gland.

The higher circularity for funicular-shape candidates was mainly caused by the general definition for circularity in re-analysis procedure of our program, as introduced in section 3.5. In order to classify the funicular-shape false candidates from the masses, we presented eleven parameters to examine the funicular-shape candidates such as : (1) length-to-width ratio, (2) minimum width, (3) circularity, (4) average contrast in the candidate, (5) average contrast in the central part of the candidate, (6) average of the standard deviations of pixel-value distributions in the equal distance, (9) percentage of the gradient-omponent ratio in constant directions determined by nipple position, (10) standard deviation for the central part of the candidate in unsharp-mask processed image, and (11) gradient ratio for each direction obtained using the gravity. After the region of interest (ROI) automatically extracted by a rectangular window that contained a minimum area including the border of suspicious mass candidate, all eleven parameters would be calculated one by one. Then the funicular-shape case was distinguished by the Mahalanobis' generalized distance. To calculate the Mahalanobis' generalized distance, a dictionary database was necessary [21].

3.6.3 Decreasing the false candidates by comparing right and left mammograms

There were several systems developed for detecting masses, which were based on the deviation from the usual architectural symmetry of normal right and left [4, 5]. Differing with others, we introduced a method based on the same point but rather to utilize it in the stage of eliminating the false candidates than to use it for detecting

masses.

At first, the images with rather enhancing border and edge were derived from a dynamic-range compression processing and a filter processing for density. Secondly, the alignment procedure for mammograms was decided by the reference information from border. Then, the correlative regions were determined. Finally, the correlation of the right and left mammograms was examined in both correlative areas. There were four feature values introduced such as the maximum of crosscorrelation coefficient, maximum of difference for pixel value, mean absolute difference on gradient, and correlation coefficient on gradient. By setting four thresholds for those parameters, the FP candidates were decreased by comparison of right and left images [17].

3.7 Indication of detected masses

An arrow described the detected masses. At same time the center and the size of detected masses were introduced as well.

4. Results and discussion

4.1 Result when using the MIAS database as a testing set

An initial result for malignant masses (TP rate : 55%, the number of FPs : 1.4 per image) was poor while the MIAS database was applied to our CAD without any modifications. Such result suggested that some modifications are necessary for our scheme when employing the MIAS in regarding our result to a Japanese database (TP rate : 91%, the number of FPs : 0.7 per image). There is a similar result reported by Brown et al. By applying the 50 cases from the MIAS database for training their CAD system (TP rate : 95 %, the number of FPs : 1.5 per image), the detection sensitivity for an Australia database was only 60% at the 1.5 FPs per image [22]. To present some significant differences in mammograms from the MIAS database and a Japanese database, some examples were shown in Fig. 2. It is not hard to find out from these examples
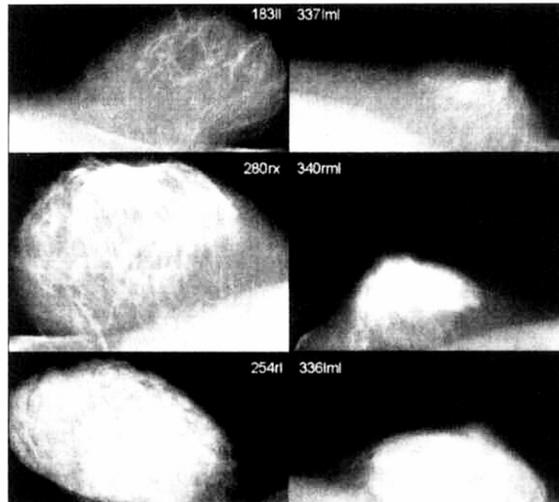
Fig. 2 Examples to show the differences between the mammograms of the MIAS database and a Japanese database. The Japanese mammograms came from Aichi Cancer Center Hospital, Japan (1998, March). In order to show these images clearly in print, some image processing methods were applied.
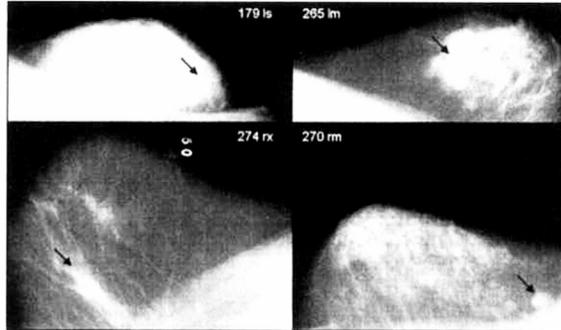
Fig. 3 Examples in which we failed to detect the masses correctly. The arrows point to the malignant masses not be detected by our CAD system. Same as the images in Fig. 2, all examples were modified to view clearly in here.

Table 1 Results for the detection of malignant masses in the MIAS database.

	TP rate	No. of FPs per image
Result 1	90% (18/20)	0.8 (265/320)
Result 2	52% (30/58)	0.8 (253/320)
Result 3	59% (23/39)	0.8 (260/320)

Table 2 A comparison of detection ability for malignant masses from the MIAS database and one of Japanese database.

	MIAS	Japanese
TP rate	90% (18/20)	91% (21/23)
FPs (per image)	0.8 (265/320)	0.7 (628/888)
TN rate	40% (128/320)	45% (400/888)

that there are more linear materials and more mammary-gland materials in the images of MIAS database. Furthermore, the size of breast area in the MIAS database is usually larger than those in Japanese databases. Such characteristics in the MIAS database not only make the detection of masses rather difficult, but also tend to increase the number of FPs. We think that the different characteristics in the MIAS database are the main reason which caused the low initial detection result introduced above.

4.2 Results when using the MIAS database as a training set

After some of our parameters were adapted by using the information of 20 malignant masses, a feasible result (TP rate : 90%, FP : 0.8), as Result 1 Table 1, was achieved as significant as the result while utilizing a database from Japan. When not only the malignant but also benign masses were considered as abnormal, the TP rate was 52 % with 0.8 FP shown as Result 2. In addition, if we regarded all of malignant masses, architectural distortions and asymmetries of mammary gland as abnormal, the sensitivity was 59% with the expense of 0.8 FP as Result 3. We trained our CAD system by using only the information from 20 malignant masses in the MIAS database. Above result is similar to those in Brown et al. paper [22]. It seems that some modifications might be necessary when the CAD system is applied to different databases.

It is important to point out that as suggested by Kallergi et al. recently [23], when evaluating the performance of detecting masses in digital mammography, two criteria should be paid attention to. The detected area should be at least 50% of the true mass area and no more than four times of

the true area in order to be considered as TP [23]. We think that the evaluation according to their suggestion would allow a better understanding of the performance. Therefore, all TPs presented in Table 1 were checked by these two criteria one by one.

In **Fig. 3**, there are four images in which our scheme failed to detect the malignant mass correctly in Result 1. The mass on image of 179ls was not detected because of its greater value of the maximum crosscorrelation when the FPs were eliminated by comparing left and right mammograms. Both images of 265lm and 274rx were unsuccessfully detected because they were not met with some thresholds in our system at the stage of eliminating the FPs. The reason we failed to detect the malignant mass in 270rm mammogram was that the suspicious candidate including the malignant mass was extracted improperly with the pectoralis muscle's area. In the stage of re-analysis of detected candidate region, when the true suspicious candidate was finally separated from the pectoralis muscle's area, unfortunately, its area with true mass had become too small to delete as a FP. Understanding from this case, it is necessary to improve our scheme in order to extract the suspicious candidates more efficiently, especially for those malignant masses in or partly in the pectoralis muscle area.

4.3 A detection comparison between the MIAS database and one of Japanese databases

Table 2 shows a comparison between the MIAS database and one of Japanese databases for detecting the malignant masses in mammograms. In the Japanese database, there are 888 mammograms from the Aichi Cancer Center Hospital in 1996. Among those images, 23 malignant masses were presented. It was not difficult to recognize from the Table 2 that we achieved the similar results for either the MIAS or the Japanese database while using them as training sets.

4.4 A discussion for three methods of eliminating the FPs

There were three methods used in our CAD algorithm for eliminating FP candidates. After these three methods were employed, the total number of FPs for 320 mammograms was reduced from 1251 to 265. It was obvious that those three methods have played an important role to improve the performance of our system for detection of masses. A discussion was provided in below for three methods, named as second-order statistics method, decreasing the funicular-shape FPs method and comparing right and left method. In short, we called them as Method 1, Method 2 and Method 3.

From our investigation, the Method 3 performed better than the other two methods for the MIAS database. On the contrary, for a Japanese database, we found that the Method 1 performed much better than Methods 2 and 3. It should be noted that the methods used for decreasing the number of FPs was strongly sensitive to some specific types of FPs. By our observation, the Method 1 enabled to decrease the smaller and low contrast FPs. The candidates eliminated by the Method 3 varied and there was no limitation of one or two types of FPs. Hence, the Method 3 was able to delimit the numerous FPs. However, it should be mentioned that there was one type of FPs eliminated efficiently only by Method 3, some FPs located in the highest density part of mammary-gland. Such kind of FPs was usually regarded to be the most difficult FPs for eliminating.

5. Conclusion

We studied the performance of our CAD scheme for detecting malignant masses by using all mammograms from the MIAS database. The initial result while using the MIAS database as a testing set was not as well as that of Japanese database. The reason for this is that our system had been training by some of Japanese databases, which are not the similar databases with the MIAS database. It seems that our detection result was deeply affected by the different characteristics in the MIAS database and the Japanese databases. After modified some of our parameters, we successfully achieved the sensitivity of 90% at the expense of FPs at 0.8. By comparing the both detection results from the MIAS database and a Japanese database, we found that we have reached at the similar sensitivities while applying the two databases as training sets. Such feasible results demonstrated that our CAD system for mass detection was effective not only on the Japanese databases, but also on the MIAS database.

In the discussion of three methods for elimination of false candidates, our result illustrated that the method of decreasing the FPs by comparing right and left mammograms was more efficient than the second-order statistics method and decreasing the funicular-shape FPs method.

We are looking forward to utilizing as many databases from overseas as possible to evaluate and improve the performance of our CAD scheme in order to provide an effective second opinion to radiologists.

References

[1] Lai SM, Li X and WF Bischof : On techniques for detecting circumscribed masses in mammograms, IEEE Trans.,8(4),377-386, 1989.

[2] Li HD, Kallergi M, Clarke LP, et al.: Markov random field for tumor detection in digital mammography, IEEE Trans. Med. Imag., 14(3),565-576, 1995.

[3] Sahiner B, Chan HP, Petrick N, et al.: Computerized characterization of masses on mammograms : the rubber band straightening transform and texture analysis, Med. Phys., 25(4),516-526,1998.

[4] Yin FF, Giger ML, Doi K, et al.: Computerized detection of masses in digital mammograms : automated alignment of breast images and its effect on bilateral-subtraction technique, Med. Phys., 21(3),445-452, 1993.

[5] Mendez AJ, Tahoces PG, Lado MJ, et al.: Computer-aided diagnosis : automatic detection of malignant masses in digitized mammograms, Med. Phys., 25(6),957-964,1998.

[6] Fujita H, Endo T, Matsubara T, et al.: Automated detection of masses and clustered microcalcification on mammograms, Proc. SPIE, 2434,265-269,1995.

[7] Matsubara T, Fujita H, Endo T, et al.: Development of a high-speed processing algorithm for mass detection based on thresholding technique in mammograms, Med. Imag. Tech., 15(1),1-13,1997.

[8] Matsubara T, Kasai S, Seki K, et al.: Development of a computer-aided diagnostic system for mammograms : improvement of the method of extracting low-density regions during automated mass detection, J. Jpn. Assoc. Breast Cancer Screen, 7(1), 87-101, 1998.

[9] Hara T, Tani Y, Fujita H, et al.: Development of a mass-classification system for mammograms, Med. Imag. Tech., 17(5), 577-584, 1999.

[10] Nishikawa RM, Giger ML, Doi K, et al. : Effect of case selection on the performance of computer-aided detection scheme, Med. Phys., 21(2), 265-269, 1994.

[11] Ibrahim N, Fujita H, Hara T, et al. : Automated detection of clustered microcalcifications on mammograms : CAD system application to MIAS database, Phys. Med. Biol., 42, 2577-2589, 1997.

[12] Suckling J, Parker J, Dance DR, et al. : The mammographic image analysis society digital mammogram database, Digital Mammography (Computational Imaging and Vision, vol. 13), Kluwer Academic Publishers, 375-378, 1994.

[13] Rangayyan RM, Elfaramawy NM, Desautels JEL, et al. : Measures of acutance and shape for classification of breast tumors, IEEE Trans. Med. Imag., 16(6), 799-810, 1997.

[14] Brake GM and Karssemeijer N : Comparison of three mass detection methods, Digital Mammography (Computational Imaging and Vision, vol. 13), Kluwer Academic Publishers, 119-126, 1998.

[15] Kok-Wiles SL, Brady JM and Highnam RP : Comparing mammogram pairs for the detection of lesions, Digital Mammography (Computational Imaging and Vision, vol. 13), Kluwer Academic Publishers, 103-110, 1998.

[16] Matsubara T, Fujita H, Hara T, et al. : Development of a new algorithm for detection of mammographic masses, Digital Mammography (Computational Imaging and Vision, vol. 13), Kluwer Academic Publishers, 139-142, 1998.

[17] Kasai S, Fujita H, Hara T, et al. : Elimination of false-positive candidates by comparing right and left mammograms in an automated mass detection algorithm, Med. Imag. Tech., 16(6), 655-665, 1998.

[18] Kato M, Fujita H, Hara T, et al. : Improvement of automated breast-region-extraction algorithm in a mammogram CAD system, Med. Imaging Inform. Sci., 14(2), 104-113, 1997.

[19] Wei D, Chan HP, Petrick N, et al. : False-positive reduction technique for detection of masses on digital mammograms : global and local multiresolution texture analysis, Med. Phys., 24(6), 903-914, 1997.

[20] Ohtsuka O, Kasai S, Hatanaka Y, et al. : Elimination of false-positive mass candidates using second-order statistics in a mammograms CAD system, Med. Imaging Inform. Sci., 16(1), 13-19, 1999.

[21] Kasai S, Fujita H, Hara T, et al. : Elimination of funicular-shaped false-positive candidates in an automated detection algorithm for mammographic masses, J. of Computer Aided Diagnosis of Medical Images, 3(2), 1-7, 1999.

[22] Brown S, Li R, Brandt L, et al. : Development of a multi-feature CAD system for mammography, Digital Mammography (Computational Imaging and Vision, vol. 13), Kluwer Academic Publishers, 189-196, 1998.

[23] Kallergi M, Carney GM and Gaviria J : Evaluating the performance of detection algorithms in digital mammography, Med. Phys. 26(2), 267-275, 1999.