Automated Detection Method for Clustered Microcalcification in Mammogram Image Based on Statistical Textural Features

Kohei Arai, Indra Nugraha Abdullah, Hiroshi Okumura

Graduate School of Science and Engineering

Saga University

Saga City, Japan

Abstract—**Breast cancer is the most frightening cancer for women in the world. The current problem that closely related with this issue is how to deal with small calcification part inside the breast called micro calcification (MC). As a preventive way, a breast screening examination called mammogram is provided. Mammogram image with a considerable amount of MC has been a problem for the doctor and radiologist when they should determine correctly the region of interest, in this study is clustered MC. Therefore, we propose to develop an automated method to detect clustered MC utilizing two main methods, multi-branch standard deviation analysis for clustered MC detection and surrounding region dependence method for individual MC detection. Our proposed method was resulting in 70.8% of classification rate, then for the sensitivity and specificity obtained 79% and 87%, respectively. The gained results are adequately promising to be more developed in some areas.**

Keywords- Automated Detection Method; Mammogram; Micro calcification; Statistical Textural Features; Standard Deviation.

I.   INTRODUCTION

Uncontrolled growth of breast cells caused by a genetic abnormality is a short meaning of breast cancer. Mostly breast cancer starts from lobules cells, glands or milk producer and duct cells, part that transporting milk from the lobules to the nipple. This cancer is exceptionally rare starts from the stromal tissues and the fatty connective tissues, but if it happens the cell changes and have the ability to divide without control and forming a tumor.

A tumor can be categorized into two types, first is benign type, which is a tumor that nearly same with the normal one in appearance, slow growth, do not spread to the other body parts and the second is malignant type, which has characteristics that vice versa from benign type.

Based on the Globocan, an international World Health Organization agency for cancer located in France, breast cancer is the most frightening cancer for women in the world, and become the most common cancer both in developing and developed regions. In 2008 estimated 1.38 million new cancer cases diagnosed, the proportion of breast cancer was 23% of all cancers.

TABLE I.   SUMMARY OF BREAST CANCER INCIDENCE AND MORTALITY WORLDWIDE IN 2008

Region	Cases	Deaths
World	1384	458
Africa Region (Afro)	68	37
American Region (Paho)	320	82
East Mediterranean Region (Emro)	61	31
Europe Region (Euro)	450	139
South-East Asia Region (Searo)	203	93
Western Pacific Region (Wpro)	279	73

From the above table, we can notice to all regions, the rates of mortality are very high and obviously there is no region in the world that has not affected with this cancer. The most worrisome region is Europe region with the number of incidence cases is 450 and mortality cases is 139. That means the rate of mortality in this region is 0.308 and made this rate is nearly equal to the rate of the world region which is 0.331.

As seen below, first rank occupied by breast cancer and the portion compared to the other cancers is extremely high which represented by age-standardized mortality rates (ASR) with 38.9% for incidence and 12.4% for mortality.
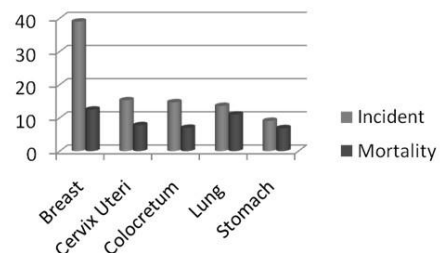

Figure 1. Age-standardized mortality rates (ASR) for women per 100000.

In order to overcome this problem, every woman needs to concern about their health through several continuous tests; Breast cancer tests covering screening tests, diagnostic tests, and monitoring tests.

In this study, we will focus on the test in screening tests called Mammograms, this test has been using for almost 40 years and the most valuable tool not only to screen the cancer, and also to diagnose and evaluate. The cooperation between mammography technician and radiologist can be involved to help the doctor increasing the accuracy of the final decision. Mammogram can read any signs of abnormality such as asymmetry of shape, irregular areas, clusters of small micro calcification (MC) and area of skin thickening. Commonly, the radiologist also operates a Computer Aided Diagnosis (CAD) system. This system will analyze the digital format of mammogram, and the result is a mammogram with any markers in the suspicious areas. The difficulty for the system is to detect clustered extra small calcifications in the form of clusters called with clustered MC.

Many researchers conducted to find the best method detecting the clustered MC. Yu and Guan [5] made a CAD that consist of two steps, first was the detection of MC pixel through classification of wavelet features and gray-level statistical features, and the second was the detection of individual MC objects, surely that the system needs a large amount of time and memory. Then Abdallah et.al [3] reported the efficient technique to detect the ROI using multi-branch standard deviation analysis and resulting the promising result which more than 98% of true positive (TP) cases. The most current one is Tieudeu et.al [1] detect the clustered MC based on the analysis of the their texture. Selection process has done via labeling method of the image that obtained from subtraction the smoothing image from the contrast enhance image, and classification of features successfully completed by neural network.

This method was resulting superfine sensitivity equal with 100% and 87.7% of specificity with proper classification rate 89%.

Therefore in this study we propose to make a system that can automatically detect the clustered MC based on the strengths from the Tieudeu et.al [1] with different enhancement image algorithm combine with detection of individual MC as done by Kim and Park [4] which employed the statistical features to detect the MC.

II. PROPOSED METHOD

A. Segmentation

The data set comes from the Japanese Society of Medical Imaging Technology, and each image has size 2510x2000 pixels and each pixel consists of 10 bits. Three categories can be found in this data set, namely calcification, normal and tumor categories. Before enter to the main process, the data should be preprocessed. The objective is to gain efficiency of time and/or memory processing, in consideration of the large size of image and size of each pixel.

Many studies have been implementing the Otsu threshold method when they want to form a binary image from the gray scale image. The main reasons are both the time processing is remarkably short and provides a satisfaction result. In this study, the segmentation operation is not only the Otsu method itself but also morphological operation being involved.

Otsu threshold method is a binarization method that calculates a measure of spread of the pixel value and iterates all possible values as a threshold. The objective is to find the threshold value based on a minimum value of within class variance and the equation described as below:

$$\sigma_W^2 = W_b \sigma_b^2 + W_f \sigma_f^2. \tag{1}$$

Where σ_W^2 is within class variance, W indicating weights, σ is a variance, b and f are background and foreground, respectively.

As a deficient result from Otsu threshold method from this data set, we need to improve the segmentation method to gain the better result of segmented image. In this study, we are applying one of morphology operations that called erosion operation. This is not ordinary erosion operation but erosion operation with small modification. There still remaining noise in the previous segmented image that must be removed which is the patient tag number, through this method that noise easily be removed. In spite of need much time to process, yet, will produce a satisfied result. The algorithm of our special erosion operation can be seen as below:

```
Input ROW, COL, MAX_ITER
Input N[ROW][COL]
For x=1 to MAX_ITER do
        For i = ROW/2 to ROW do
                For j=1 to COL do
                If N[i][j] = 0 then
                        N[i − 1][ j] = 0
                end if
                if j < COL/2 and N[i][j] = 0 then
                        N[i][ j − 1] = 0
                end if
                if j > COL/2 and N[i][j] = 0 then
                        N[i][ j + 1] = 0
                end if
                end for
        end for
end for
```
Algorithm 1. Our erosion algorithm

B. Detection of Clustered MC

1) Breast Tissue Detection Based on Texture-based Analysis

In this study, we are applying the method that has developed by Tieudeu et.al [1] with modification in one specified area. They are developed the main method by utilizing three methods. First is enhancing the contrast of the original image then produce an image called with contrast enhance image (CI) and the way to get this image become a point of modification. The second is smoothing the original image then produce an image called with smoothed image (SI). The last is subtraction the smoothed image from enhanced image then called with difference image (DI).

This adoption motivated by clustered MC that allied with breast mass can be concluded as a benign or even premalignant cancer. Frequently, MC only associated with extra cell growth inside the breast. Different with the previous study when forming the CI, we are using the histogram equalization method with an aim to spread the most frequent intensity values that make the lower contrast reach a higher contrast. The details represented by the equation below:

$$prob_n = \frac{\#pixels\ intensity\ n}{\#pixels}; \quad n = 0, 1, 2 \dots L. \qquad (2)$$

$$M_i = floor\left((L)\sum_{n=0}^{i} prob_n\right). \qquad (3)$$

Where $prob_n$ denotes the normalized histogram for each gray level value, n is gray level values, L is maximum gray level value and M is image matrix.

2) Multi-branch Standard Deviation Analysis

MC related with local maxima values in the image. This idea became a point to find up a correlation between the local maxima and its neighboring pixels. In this study, we conduct an analysis with make use of standard deviation method to find that correlation as reported by Abdallah et.al [3]. Based on visual observation for calcification category, there is not only one or two clustered MC in one image but even more than five clusters of MC can be found. In relation of that problem, developing a multi-branch point of view become something primary needs. It because highly possible if we find a local maxima in one direction and after take a look in a different direction that point is not a local maxima. That critical point provides promising solution to find the clustered MC in one small area. The illustration provided as below:
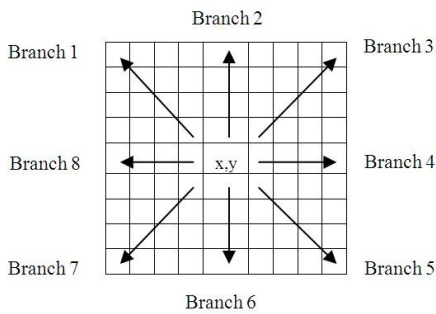

Figure 2. Multi-branch standard deviation analysis to find MC.

Where x, y point is an ideal local maxima if from all branches seen as a local maxima, branch direction move clockwise start from branch 1, branch 2, branch 3, branch 4, branch 5, branch 6, branch 7 and branch 8. At the time that we want to know one point is local maxima from one branch, the threshold value and the counter needed. While calculating the threshold between the central pixel and its neighbor pixels if the standard deviation greater than the threshold value the counter will be increasing by one, whereupon an ideal local maxima is the point that has a counter value equal with eight. Described with the following equation:

$$STD_i = \sqrt{\frac{\sum_{i=1}^{n}(Center - x_i)^2}{n}}; \quad i = 1, 2, \dots, 8. \qquad (4)$$

Where:

STD_i	=	Standard Deviation at branch i
Center	=	Cluster center
x_i	=	Gray level value at the specified position i
n	=	Number of pixels

As said before the counter will have a maximum value 8, that value is equal with a total of branches in this method. Size

of the detection window in this method is 9x9, and that size obtained from the reference that MC in mammogram image can be captured through that size of the mask. ROI as a final result of this section has size 128 x 128 which matched with the most clustered MC size. In this study, one mammogram image represented by one ROI although there is more than one clustered of MC can be found. It because this system's purpose is giving assistance to the doctor and the radiologist when they are facing the final decision, at the moment only one representation of clustered MC is found still means the patient categorized as calcification and need further treatment. Moreover, selection criterion of ROI is the area with the highest number of suspicious local maxima pixels.

C. Detection of Individual MC

1) Surrounding Region Dependance Method

In this part, we will talk about detection of individual MC through the method that previously used by Kim and Park [4]. The method is Surrounding Region Dependence Method (SRDM) which utilizing rectangular and threshold in order to obtain the distribution matrix. This matrix represents a characteristic of the ROI image that related to calcification case or not. Consider two rectangular windows are centered in *x, y* pixel, with largest window has size 5, and intermediate is 3. As shown with the image below:
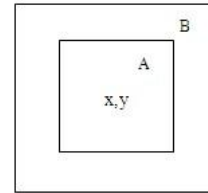

Figure 3. Surrounding Region Dependence Method.

A is inner surrounding region, and B is outer surrounding region.

SRDM involves a *M(q)* or matrix of a surrounding region dependence obtained from transformation of an ROI image and *q* is a given threshold value. The details presented as below:

$$M(q) = [\alpha(i,j)], 0 \le i \le n. \qquad (5)$$

$$\alpha(i,j) = \#\{(x,y)|c_A(x,y) = i\ and\ c_B(x,y) = y, (x,y) \in L_x \times L_y\}. \qquad (6)$$

Where $L_x \times L_y$ is two dimensional image space and c_A, c_B are inner count, outer count, respectively.

Feature extraction is an essential part when dealing with the classification term. Hereafter horizontal, vertical, diagonal, and grid-weighted sums are extracted from the characteristics of the element distribution in the SRDM matrix as textural features.

The distribution for a positive ROI will tend to the right and/or lower right of the matrix and indicate us if subtraction neighbor values in inner and outer rectangles from the center value more than the threshold, those values will be located at

the right part of the matrix. For negative ROI has a contrary description, the distribution will tend to other location of the positive ROI.

III. EXPERIMENTS

A. Segmentation

Segmentation process in this study has an aim to remove the noise which called mammogram's tag number and the backlight. Otsu threshold method successfully removed the backlight from the image and the remaining noise is tag number, this noise removing process is handling by erosion method. We have 65 images in the data set and only three images that categorized as dissatisfied results. The reason of the negative appearance is because breast size of those patients classified as extraordinarily large size and has a round shape that made on both corners of the mammogram image have a less visible area. The satisfy segmented image and dissatisfy segmented image presented respectively as below:
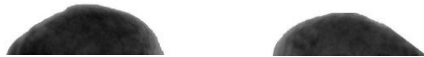
Figure 4. Satisfy segmented images.

Figure 5. Dissatisfy segmented images.

B. Detection of Clustered MC and Individual MC

Through the described method, we obtained all images called the CI, SI and DI. From below DI image we can obviously see the breast tissue area and hereafter this area will be the main concern when finding the clustered MC. As an example, shown with the images below:
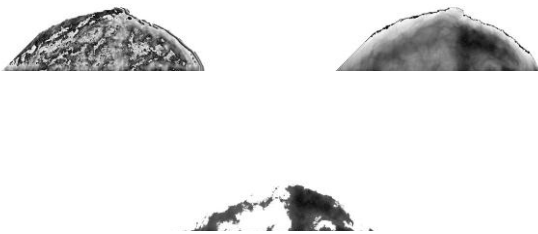
Figure 6. Sample of the CI, SI and DI images.

The naming format of below images is category plus image number in data set, for example, C5 means ROI image that categorized as calcification with number image is 5, in the sequel example are C6U and C6L shown us that U has originality from the upper part of C6, hence, the C6L from the lower part. The others categories denote with T for tumor category, and N for normal. Mostly the MC detected on this

category and obviously showed that this method was suitable to detect the clustered MC. From the experiment, threshold value for clustered MC detection equal with 8 was the maximum threshold value. Hereupon, best threshold value for individual MC detection was 3. According to the proposed method, resulting ROI images as presented below:
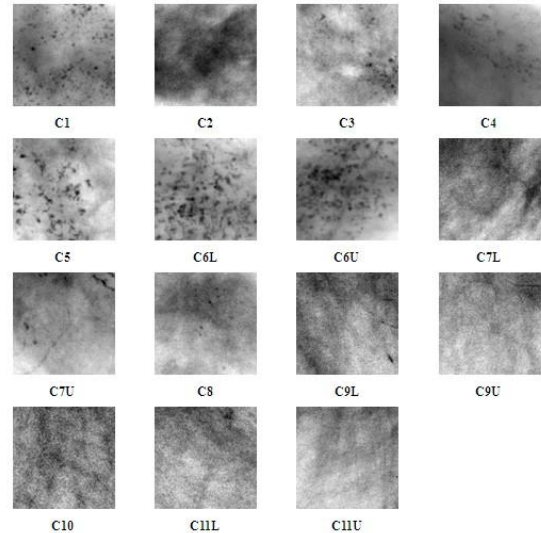
Figure 7. Region of Interests (ROI) from calcification category.

In this part, the data set separated into two parts that are training and testing parts with the data proportion were 50% and 50%, respectively. For training data, we were adding ideal output in the form of ROI from all categories manually to train the classifier and then extracted their features. Manual observation of all data passed, and we acquired the information that in category tumor also found clustered MC. At least, four tumor images possessing clustered MC and those were T1, T2, T4 and T8 images. That finding guidance based on sketch images that provided inside the data set, as seen in the following image:
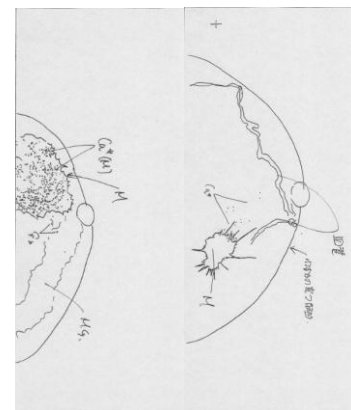
Figure 8. Sketch images, calcification category (left), tumor category contains MC (right).

Classification result for this system was good enough pointed by the classification rate that was 70.8%.

Mostly, true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are the options for

diagnosis decision. TP means similarity cancerous of judgment from an expert and system, TN means similarity a non-cancerous judgment from an expert and system, FP means a non-cancerous classified as cancerous, and last is FN which means a cancerous classified as a non-cancerous. After the experiment, the results shown with the following table:

TABLE II. CONFUSION MATRIX

TP	FP
15	6
FN	**TN**
4	40

Hereafter let we talk about other parameters that could indicate the system whether is acceptable or not which are sensitivity and specificity. Both parameters shown as below:

Sensitivity = TP/(TP+FN)
Specificity = TN/(TN+FP)

Then obtained those values equal with 79% and 87%, respectively. Regarding the sensitivity value was deficient, there is a primary reason, because we were trying to find the MC which had a round shape. In fact on few images the shape of MC included round as well as long shape. The system could not find that shape of calcification precisely. The reason is the window for detecting local maxima pixel that has identified as MC was a small size rectangular. On account of that reason, the value of four for a false negative was appeared. Described with the following image:
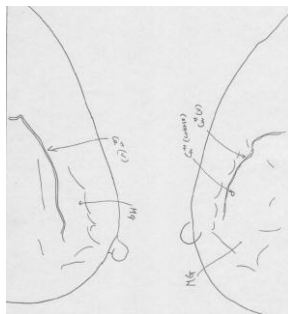

Figure 9.   Sketch image of different calcification shape.

We can obviously see the long shape denotes with $Ca^+(v)$. That shape also became a barrier for detector of clustered MC to detect the correct shape.

IV.   CONCLUSION

This study is developed exclusively to detect clustered MC. We have reasons why this system could not gain the perfect classification rate, first is the textural features that became an input of the network had a lack of proper characteristics to discriminate a clustered of MC and nonclustered MC, and second because we worked on small data only consists of 65 images. On the contrary, we realized to publish this kind of data should have a permit for their own information. However, from the gained results are adequately promising to be more developed in some areas, parallel with important thing for a human being is to help each other.

V.   FUTURE WORK

The future work that can be developed from this current progress is the detection of clustered MC to determine a mammogram image is included as benign or malignant. Conduct another localized and efficient method when forming contrast enhance image.

REFERENCES

[1] A. Tieudeu, C. Daul, A. Kentshop, P. Graebling, D. Wolf, "Texture-based analysis of clustered microcalcifications detected on mammograms,"Digital Signal Processing,Vol 22, pp. 124-132, 2011.

[2] Lucio F. A. Campos, A. C. Silva, A. K. Barros, "Diagnosis of breast cancer in digital mammograms using independent analysis and neural network," CIARP, pp. 460-469, 2005.

[3] M. H. Abdallah, A. A. Abubaker, R. S. Qahwaji, M. H. Saleh, "Efficient technique to detect the region of interests in mammogram image," Journal of Computer Science, Vol 4(8), pp. 652-662, 2008.

[4] Jong Kok Kim and Hyun Wook Park,"Statistical textural features for detection of microcalcifications in digitized mammograms," IEEE Transactions on Medical Imaging, Vol 18(30), pp. 231-238, 1999.

[5] Songyang Yu and Ling Guan,"A cad system for the automatic detection of clustered microcalcifications in digitized mammograms films," IEEE Transactions on Medical Imaging, Vol 19(2), pp. 115-125, 2000.

AUTHORS PROFILE

KOHEI ARAI received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a counselor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology from 1998 to 2000. He was a counselor of Saga University for 2002 and 2003. He also was an executive counselor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008.

INDRA NUGRAHA ABDULLAH was born in Bogor, Indonesia on June 18th 1987. Finished his bachelor degree in Bogor Agricultural University and graduated from Saga University for master degree in the field of Information Science on March 2011. He is currently pursuing to get Ph.D. Degree from the same university with specialization in image processing area. Leaf identification becomes his interest in his latest degree.

HIROSHI OKUMURA was born at Kyoto, Japan in 1964. He received B.E.S.E. and M.E.S.E. degree from Hosei University in 1988 and 1990, respectively, and Ph.D degree on environmental engineering from Chiba University in 1993. He became a research associate at Remote Sensing and Image Research Center, Chiba University first in 1993. Next, he became a research associateand a lecturer at the Department of Electrical Engineering, Nagaoka University of Technology in 1995 and 2000, respectively. He is now an associate professor at the Department of Information Science, Saga University. His research interests are in image and speech processing and remotesensing.