An Expert Image Processing System on Template Matching

Siti Noraini Sulaiman[1], Mohd Fauzi Alias[2], Nor Ashidi Mat Isa[3] and Mohamad Faizal Abd Rahman[4]

[1, 4] Faculty of Electrical Engineering,
Universiti Teknologi MARA, 13500 Permatang Pauh, Penang Malaysia.
[2, 3] School of Electrical & Electronic Engineering,
Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Penang Malaysia.

Summary
An expert image processing system on template matching is proposed to locate and determine the symmetrical value between breast mammogram. A new template matching algorithm using cross-correlation method is implemented in this work. Cross-correlation algorithm operates well on two-dimensional images and gives the best result for the matching process. We have developed a template matching algorithms for detecting similarity in grayscale mammograms recently. After employing the method, the detection of matching percentage becomes more reliable; more than 80% of matching percentage. Such result indicates that the new technique has improved the performance of our computer aided diagnosis system for mammographic breast cancer detection effectively.

Key words:
Breast cancer, computer-aided diagnosis, mammography, template matching.

1. Introduction

Breast cancer is the fifth most common caused of cancer death in the world after lung cancer, stomach cancer, liver cancer, and Colon cancer [1], [12]. In 2005, breast cancer caused 502,000 deaths (7% of cancer deaths; almost 1% of all deaths) worldwide [1], [12]. Breast cancer was the most common cancer amongst Malaysian women from the age of 20 years for 2002 and 2003. The breast cancer cases are mostly detected in the Chinese, followed by the Indians and then, Malays [2]. From the statistics published by the National Cancer Registry Reports 2002 and 2003, breast cancer formed 30.4% and 31% of newly diagnosed cancer cases in 2002 and 2003 respectively[2]. In 2003, 1 in 19 women in Malaysia possibly has a chance of getting breast cancer in her lifetime. The approximations of developing breast cancer of this possibility were 1 in 14 for Chinese women, 1 in 15 for Indian women and 1 in 24 for Malay women. In 2000, the International Agency for Research in Cancers (Globocan 2000) reported 3825 cases and 1707 deaths from breast cancer in Malaysia.

The normal breast is difficult to define and often confusion between physiologic and dystrophic changes [3], [7]. The densities of mammographic structures are low and there is an overlap between normal and pathological tissues [3]. This is because the breast is a three-dimensional structure and the mammographic image is the result of summation this structure into two-dimensional image. However, there are no certain rules and positioning in viewing mammographic image. Therefore, for this project, we included several useful techniques in viewing the mammographic image.
Template matching is an important problem with strong roots in detection theory [9]. It has obvious applications in computer and robot vision, stereography, image analysis and motion estimation [9]. Template matching in the context of an image processing is a process of locating the position of a sub image within an image of the same, or more typically, a larger size[4],[5],[15]. Template matching can also be described as a process to determine the similarity between two images. The sub image is referred to as the template image and the larger image is referred to as the search area (main image) [4], [5], [8], [15]. The template matching process involves shifting the template over the search area and computing a similarity between the template image and the window of the search area over which the template lies [4], [5], [6], [15]. These shifting and computing processes operate simultaneously and do the repetition until the template image lies on the edge of the search area.

For the image processing purpose, the most common technique for measuring the similarity between two-dimensional images is by using cross correlation method. Basically, the cross correlation method can be described as a correlation between two signals. Textbook presentations of correlation almost describe about the mathematical concepts and equations of correlation in

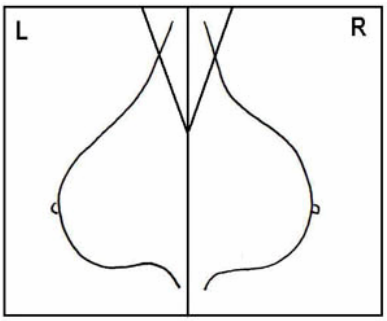
Manuscript received July 5, 2007.
Manuscript revised July 25, 2007.

the frequency domain using the Fast Fourier Transform. Unfortunately these correlation equations need to be changed into more simple and efficient expression for computer system usage. A correlation measure is determined between the template and respective windows of the search area to find the template position which has maximum correlation [4], [10]. Generally, for two-dimensional image, the correlation coefficient is computed from all area on the template image shifted on the main image. Therefore, a new template matching algorithm using cross-correlation method that applied on gray scale mammogram images is discussed in this paper.

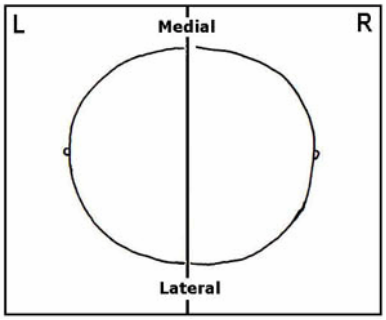
2. The Normal Breast and Mammogram Positioning.

The breast can be imaged from a number of angles. The most popular two different views in positioning the mammogram are mediolateral oblique (MLO) and cranio-caudal (CC) views as shown in Figure 2.1. For some cases, most of the researchers used MLO as their common view followed by the CC view [11].

In mammogram viewing process, it is important to know the exact orientation of the image. Generally, the breasts are one of the human's symmetric organs and best to be viewed in symmetric orientation. Comparison of the right breast to the left breast is done for evaluation of symmetry between both sides. Perceptual psychologists have shown that the eye can more easily perceive asymmetric densities when patterns are scanned in a mirror-image fashion rather than side by side [11], [18]. Therefore, the conventional method to evaluate mammogram is to place the MLO and CC views in mounted back to back. However, there is no certain position is ruled out to be followed but is a matter of individual preference.



(a)



(b)

Figure 2.1: The positions of the mammogram for viewing (a) Cranio-Caudal(CC) view (b) Mediolateral Oblique (MLO) view

3. Template Matching by Cross Correlation.

For a sample image (contains left and right breast images), the position, orientation and size need to be observed. These criteria are important in resulting the best output for the matching process. Unfortunately, almost of the images are not satisfied by these criteria. Therefore, some techniques are included for preprocessing purpose to get the satisfied images such as flip image, region of interest (ROI) and image binarization. These techniques are not the major work for this project. Therefore, these techniques are not discussed in this paper. We use a correlation method for this purpose which is including some other techniques such as features extraction, binary image, masking image and flip image. In statistic, the term cross-correlation is sometimes used to refer to the covariance cov(*X*, *Y*) between two random vectors *X* and *Y*, in order to distinguish that concept from the "covariance" of a random vector *X*, which is understood to be the matrix of covariance between the scalar components of *X* [13], [14]. The correlation method is implemented using the Fast Fourier transform and modified into discrete correlation expression for computer usage. Discrete Correlation is mathematically compared process by one image to another image discretely. The resulting image is a two- dimensional expression of equivalence. Generally, the equation for the two-dimensional normalized cross-correlation is given by [16], [17]:

$$CC(k,l) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} M(i-k, j-l)N(i, j)}{[\sum_{i=1}^{n}\sum_{j=1}^{n} M(i, j)^2 \sum_{i=1}^{n}\sum_{j=1}^{n} N(i, j)^2]^{\frac{1}{2}}}$$

Where;

n x n image, *M* is called a template image;

n x n image, *N* is called a search area (main image) with a window of a larger image.

Supposedly, the relation between template image and search area (main image) is defined as;

$$N(i, j) = cM(i, j)$$

$$CC = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} M(i,j)N(i,j)}{[\sum_{i=1}^{n}\sum_{j=1}^{n} M(i,j)^2 \sum_{i=1}^{n}\sum_{j=1}^{n} N(i,j)^2]^{\frac{1}{2}}}$$

$$CC = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} cN(i,j)N(i,j)}{[\sum_{i=1}^{n}\sum_{j=1}^{n} c^2 N(i,j)^2 \sum_{i=1}^{n}\sum_{j=1}^{n} N(i,j)^2]^{\frac{1}{2}}}$$

$$CC = \frac{c\sum_{i=1}^{n}\sum_{j=1}^{n} N(i,j)^2}{c[\sum_{i=1}^{n}\sum_{j=1}^{n} N(i,j)^2 \sum_{i=1}^{n}\sum_{j=1}^{n} N(i,j)^2]^{\frac{1}{2}}}$$

$$CC = 1$$

For some constant c, the correlation coefficient will result as a value 1, proving the 100% similarity between both images. This theory is applied together with the discrete correlation expression in this system. The equation for the two-dimensional discrete correlation is given by [5];

$$Out(k,l) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \ln 1(m,n) \ln 2(k+m, l+n)$$

Where *ln1* and *ln2* are the input images and *Out* is the output image. *M x N* is the dimension of both images and possibly they have different dimensions and sizes, however, it will filled in with certain pixel values to allow the matching process. The *ln1* and *ln2* are called the search area and the template image respectively. The output will reveal the resultant image and its position of the greatest match between the search area (main image) and the template.

4. Experiments and Results.

The effectiveness of the proposed template matching system has been evaluated using the sample of breast images. For each sample, there were left and right breasts for one patient. These images are declared as symmetrical breasts (normal breast). For each sample, one of the breast side's is set as the template and another side is the search area. The resultant output image will be in black and white pixels. The white pixels depict the similarity value while the black pixel shows the difference value between both images. Therefore, the matching accuracy becomes higher when the white pixel appears more than the black pixel in the output image. By using the ROI technique, two sample images are cut out form its original images and tested for the system accuracy evaluation.

The matching accuracy is evaluated by equation:

$$\text{Accuracy} \quad \frac{\sum \text{white pixels}}{\sum \text{pixels}} \times 100\%$$

Figure 4.1 shows our image samples for the experiment.

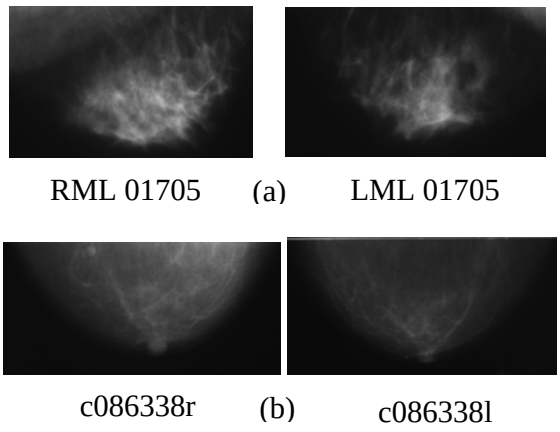


Figure 4.1: Input images for template and search area

The output images are shown in Figure 4.2 (a) and (b). The similarity of the corresponding image with its template indicated by the emergence of white pixel; i.e the more white pixel value, the more both image are alike and vise versa. The matching percentage for resultant images in Figure 4.2(a) and 4.2(b) are 94.89% and 82.82% respectively. These high percentages prove the images are nearly similar or identical.

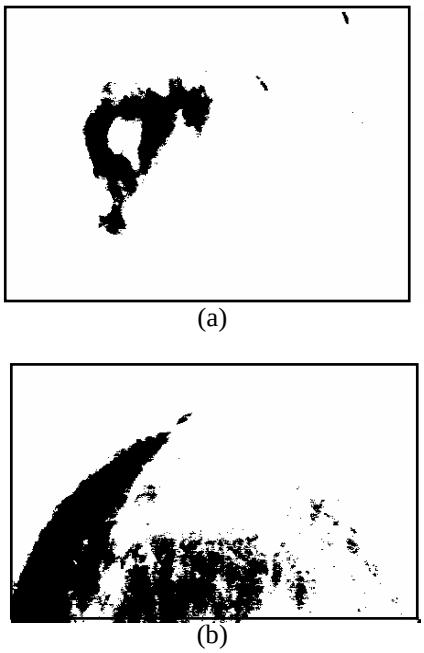
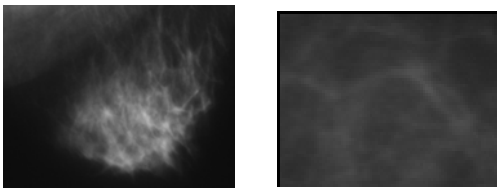


Figure 4.2 (a) The output image for RML 01705 and LML 01705 (b) The output image for c086338r and c086338l.

For the two-dimensional matching accuracy experiments, two samples of image are tested for this experiment. A specific area of interested region is copied from the search area to be declared as a template. Each of these copied images is compared to its original image. For sure, theoretically, the matching percentage for the matching accuracy experiment should be in a perfect

score (100%). Figure 4.3(a) and (b) show the input images for our matching accuracy experiment.



(a) (b)

Figure 4.3: The template images for matching accuracy experiment (a) An interested region from RML 01705 (b) An interested region from c086338l

The matching percentage for both images is 100%. The perfect score of matching percentage results a perfect accuracy in the system. Therefore, the matching system is 100% accurate included with the perfect match between the template and its search area (the output images are in white pixel). The proposed method gives a promising result, however, further study still need to be conducted with various cases of samples to verify the result.

5. Conclusion

In this paper, we have shown that the cross-correlation algorithm can be very effective in template matching method. An expert template matching system has been proposed to evaluate the symmetrical breast based on mammogram image. The system requires two input images to apply the cross-correlation algorithm for the symmetrical calculation. Then the output will show the resultant image (in black and white only) with their matching percentage between those images. As conclusion, this project has successfully carried out the following findings:

1. The proposed cross-correlation algorithm performs perfect subject in the template matching system.
2. The matching accuracy equation is totally reliable and result a perfect score in this project.

For this project, even though the results obtained so far are encouraging, more investigations need to be done in more detail on both theoretical and practical side for further study on matching process.

Acknowledgment

The work describe in this paper is supported by short term grant from IRDC UiTM entitled “Identification of Tumor Existence in Mammogram Image Using Image Processing Algorithm”.

References

1. Wikipedia, "Breast Cancer" Citing Internet Sources URL "http://en.wikipedia.org/wiki/Breast_cancer".
2. Radiologi Malaysia, College of Radiology Breast Health Information Centre, “Facts and Stats” Citing Internet Sources URL http://www.radiologymalaysia.org/breasthealth/About/FactsNStats.htm.
3. J. Stines and H. Tristant, “Breast Imaging: Radio histological Correlation”. 2005.
4. Yongmin Kim, Shijun Sun, HyunWook Park, “Template Matching Using Correlative Auto-predicative Search”. University of Washington, 2004.
5. Harley R. Myler, Arthur R. Weeks, “The Pocket Handbook of Imaging Processing Algorithms in C”. Department of Electrical & Computer Engineering, University of Central Florida, Orlando Florida. 1993.
6. Gregory S. Chirikjian and Alexander B. Kyatkin, “Engineering Applications of Noncommutative Harmonic Analysis: With Emphasis on Rotation and Motion Group”. 2001 by CRC Press LLC.
7. Abbas Samani1, Judit Zubovits2 and Donald Plewes3, “Elastic Module of Normal and Pathological Human Breast Tissues: An Inversion-Technique-Based Investigation of 169 Samples”. 2007.
8. Maria Kallergi, “Medical-Image Analysis and Methods”. Copyright 2005 by Taylor & Francis Group, LLC.
9. Mehdi Khosravi and Ronald W.Schafer, “Template Matching Based on a Grayscale Hit-or –Miss Transform”. 1996.
10. Dongming Zhao, “Image Processing and Feature Extraction”. 2005 by Taylor & Francis Group, LLC.
11. James Nguyen, Dr. David Fleiszer, “Interactive Mammography Analysis Web Tutorial”, McGill Faculty of Medicine. 2002.
12. WHO, World Health Organization, "Cancer" Citing Internet Sources URL http://www.who.int/mediacentre/factsheets/fs297/en/index.html.
13. John C.Russ. “The Image Processing Handbook, Fourth Edition”. 2002 by CRC Press LLC.
14. Kimmo Fredriksson and Esko Ukkonen, “Fast Template Matching Without FFT”. Department of Computer Science, University of Helsinki. 2001.
15. Rafael C.Gonzalez and Richard E.Woods, “Digital Image Processing Second Edition”. 2002.
16. Jana Kosecka, “More on template matching PPT. file”. Citing Internet Sources URL cs.gmu.edu/~kosecka/cs682/cs682-template-matching.pdf.
17. Jana Kosecka, “Binary Image Analysis PPT. file”. Citing Internet Sources URL

http://cs.gmu.edu/~kosecka/cs482/cs482-binary-analysis-2.pdf.
18. J.R. Parker, "Algorithms for Image Processing and Computer Vision". 1996.

[1]Siti Noraini Sulaiman obtained her B.Eng Hons in Electrical Engineering from Universiti Sains Malaysia in 2000 and MSc in Medical Imaging from the same university in 2003. She is currently lecturing at the Faculty of Electrical Engineering, UiTM Penang Branch Campus in Permatang Pauh. Her research interest includes intelligent system, image processing and their application to medical.

[2]Mohd Fauzi Alias received his B.Eng Hons in Electronic Engineering from Universiti Sains Malaysia in 2006. He is currently doing his MSc in Image processing for medical application at the same University.

[3]Nor Ashidi Mat Isa obtained his B.Eng Hons in Electrical Engineering from Universiti Sains Malaysia in 2000 and PhD in Image Processing and Neural Networks from the same university in 2003. He is currently lecturing at the School of Electrical and Electronics Engineering, USM Engineering Campus in Transkerian. He specializes in the area of intelligent system, image processing, neural networks for medical applications and algorithms.

[4]Mohamad Faizal Abd Rahman obtained his B.Eng Hons in Electronic Engineering from University of Manchester Institute of Science and Technology(UMIST) in 2001 and MSc in Electronic System Design from Universiti Sains Malaysia in 2005. He is currently lecturing at the Faculty of Electrical Engineering, UiTM Penang Branch Campus in Permatang Pauh