

GRAY LEVEL CO-OCCURRENCE MATRIX APPLICATION IN DETECTION OF ABNORMALITIES ON MAMMOGRAMS

ỨNG DỤNG MA TRẬN GLCM PHÁT HIỆN VÙNG BẤT THƯỜNG TRÊN ẢNH X QUANG VÚ

Nguyen Thai Ha, Nguyen Duc Thuan, Pham Manh Hung, Doan Dinh Nhan
Hanoi University of Technology

ABSTRACT

According to statistics at K hospital the rate of Vietnamese women having breast cancer is increasing and breast cancer is the most common type of cancer among women. Early detection of breast cancer plays an important role in the treatment and cure for patients. This paper has suggested a certain method for edge extraction in digitalized mammograms, especially in detecting microcalcifications and masses. This method applies spatial gray level co-occurrence matrices. The application of GLCM provides a new method of boundary extraction. When applied to mammograms, it gives satisfactory results. A medical image processing software is built based on this algorithm to help doctors in the analysis of mammographic images, diagnosis and detection of breast cancer. The software is still being researched and developed and will be packed into a perfect product to aid in breast cancer diagnosis at hospitals and in study as well.

TÓM TẮT

Theo thống kê tại bệnh viện K tỉ lệ phụ nữ Việt nam mắc bệnh ung thư vú ngày càng tăng và ung thư vú đang trở thành bệnh ung thư phổ biến nhất của phụ nữ. Việc phát hiện sớm bệnh thư vú đóng một vai trò rất quan trọng trong việc điều trị và cứu chữa bệnh nhân ung thư vú. Bài báo trình bày ứng dụng ma trận GLCM (Grey-Level Co-occurrence Matrix) tách biên ảnh và phát hiện các cụm vi canxi hoá và khối u trong ảnh X quang vú số hoá. Ứng dụng GLCM là một phương pháp mới trong việc tách ra đường bao giữa các vùng. Khi áp dụng trên ảnh X quang vú, nó cho kết quả tương đối khả quan. Dựa trên thuật toán này, một phần mềm xử lý ảnh được xây dựng nhằm hỗ trợ các bác sĩ trong việc phân tích ảnh X quang vú, giúp chẩn đoán và phát hiện ung thư vú. Phần mềm này vẫn đang được nghiên cứu và phát triển và sẽ được đóng gói thành một sản phẩm hoàn chỉnh giúp hỗ trợ việc chẩn đoán sớm ung thư vú tại các bệnh viện cũng như trong việc nghiên cứu.

I. INTRODUCTION

1.1 Overview

According to statistics at K hospital – the leading hospital in cancer treatment in the North of Vietnam – the rate of Vietnamese women having breast cancer is increasing and breast cancer is the most common type of cancer among women. These statistics also indicate that, Hanoi has the highest rate of cancer cases in Vietnam. This disease causes high rate of death, but if early detected, it can be cured completely [2].

Early detection of breast cancer plays an important role in the treatment and cure for patients. According to comment of specialists, cancerous cells in the breasts have the size of sand grains at the beginning so it is difficult to be detected. If they are detected at the first stage, the

chance of survival is up to 80%, at the second stage is 60%, at the third stage it is much lower and to the fourth stage usually the treatment is just to prolong patient's lives and ease painful symptoms rather than to cure the disease. Masses in the breasts at the first stage cannot be felt, and can be detected by screening mammography or biochemical tests. At stage 1, masses are smaller than 2 centimeters and there is no axillary node. At stage 2 masses are from 2 to 5 centimeters, and moving nodes on the same side appears. At stage 4 masses are of any size, nodes usually exist and spread to other organs in the body such as bone, lung, liver, the remaining breast, brain. The most effective diagnostic imaging method is mammography, which helps detecting of

abnormalities two years before they become malignant masses [8].

As we already knew, the breasts have a complex anatomical structure with a lot of lobes, tissues and superimposing components which are hard to be identified on X-ray films. However, not all abnormalities appearing on films are signs of masses and cancer. On a digitized mammogram, various cancerous or pre-cancerous signs can be detected such as: clusters of microcalcifications, visible injuries, circumscribed masses, illdefined masses, asymmetric structures... Many methods of digitized mammogram analysis have been researched and tested successfully in order to support doctors in the diagnosis [1,4,6]. Many basic image processing techniques are applied to mammograms to help doctors identify signs and suspicious areas on mammograms. A diagnosis-aiding system will segment and mark suspicious areas on the images and thus help early detection of abnormalities.

Currently in Vietnam, Hanoi K-Hospital is the leading hospital in cancer detection and treatment. However there are limitations in application of cancer-diagnosis-aiding tools. Some programmes have been used but they only have basic functions to help doctors interpret films more accurately such as brightness and contrast enhancement of films.... After surveying and researching at K-Hospital and some other hospitals, together with studying image processing algorithms, our group has developed a software which can detect abnormalities on mammograms to support doctors in the diagnosis. This software includes various functions such as reading, editing, deleting images, and some basic processing steps such as contrast and brightness enhancement, noise filtering... Additionally there are functions of supporting study and research such as histogram representation and modification, boundary extraction. In this paper, however, these functions of the software are not discussed, but a method of microcalcification detection using Grey-Level Co-occurrence Matrix (GLCM) which was suggested by Haralick [9,10].

1.2 The purpose of research and development of the detection aiding software:

There exists a fact that the number of patients of breast examination is big and the possibility of missing suspicious areas is high, may be up to 40-50%, moreover diagnosing results mainly depend on the quality of X-ray films, therefore the software is developed in order to support doctors in enhancing image quality, analysis and segmentation of suspicious areas on mammograms and thus make proper conclusions on the risk of having breast cancer of patients.

In addition, the software is also helpful in supporting research on medical image processing, especially radiographs.

II. PROGRAMME CONSTRUCTION

2.1 Programming language selection

For the purpose of study and research, we usually think of Matlab. This is a common language providing many supporting tools. However the limitation of Matlab is that the programme is heavy and hard to be packed due to copyright matter. Whereas C++ is also a strong language and usually used for image processing. Therefore we decided to select C#. This strong language is easily packed into a product and being built into a standard library for study and research activities.

2.2 Building algorithm

A Grey-Level Co-occurrence Matrix (GLCM) based texture measure have been the workhorse of image texture since they were proposed by Haralick. A GLCM is created by calculating how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j . By default, the spatial relationship is defined as instance between the pixel of interest and its adjacent pixel to the right in horizontal direction. Each element (i,j) in the resultant GLCM is simply the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image [9,10].

Figure 1 shows how graycomatrix calculates the first three values in a GLCM.

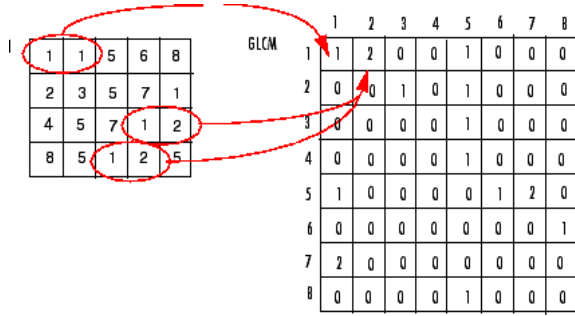


Fig. 1 An example of GLCM creation from left to right of a 3bit image.

In the output GLCM, element (1,1) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the same value of 1. GLCM(1,2) contains the value 2 because there are two instances where two horizontally adjacent pixels have the values of 1 and 2. Element (1,3) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 3. The input image is similarly processed for other pixel pairs (i,j) and the number of corresponding instances is assigned to each cell in GLCM.

Applying GLCM to edge extraction, segmentation, and cancrification detection in digital mammograms, there are several ways to calculate texture measures from the GLCM, such as: contrast, dissimilarity, and homogeneity.

Contrast group measures related to contrast use the weights related to the distance from the GLCM diagonal

$$U_c = \sum_{i,j=0}^{N-1} P_{i,j} (i-j)^2 \quad (1)$$

Where:

$P_{i,j}$ is the probability of a given outcome, for an i ro, and j column.

From formula (1), in term of contrast measure, weights (U_c) increase exponentially (0, 1, 4, 9, etc.) as one moves away from the diagonal

However in term of dissimilarity measure, weights (U_{dis}) increase linearly (0, 1, 2,3 etc.) (formular (2)).

$$U_{dis} = \sum_{i,j=0}^{N-1} P_{i,j} |i-j| \quad (2)$$

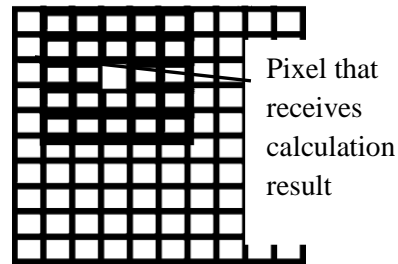
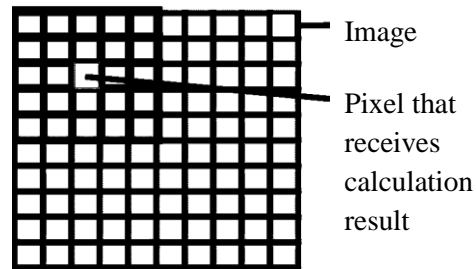
In these formulars, the probability of a given outcome $P_{i,j}$ is calculated by the formular (3).

$$P(i, j) = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}} \quad (3)$$

Where:

$V_{i,j}$: the number of times this outcome occurs
 $\sum_{i,j=0}^{N-1} V_{i,j}$: the total number of possible outcomes

Window 1:



Window2

Fig. 2 5x5 window

The result of a texture calculation is a single number representing the entire window (Fig 2). This number is put in the place of central pixel of the window, then the window is moved one pixel. The process is repeated of calculating a new GLCM and a new texture measure. In this way an entire image is built up of texture values. We can choose 3x3, 5x5 or 7x7,..., (2N-1)x(2N-1) windows (Fig. 2 is an example of 5x5 window).

The result of the calculation shows the difference between two neighbouring pixels in one window. The larger the gray level difference is, the greater the result will be and

vice versa. This feature is utilized to define edges of an image.

For higher accuracy, the results are calculated in two directions, from left to right and from top to bottom, and then summarized in the following formula (4):

$$U_{th} = \sqrt{U_D^2 + U_N^2} \quad (4)$$

Here:

U_{th} : stands for the summarized different gray level of two directions

U_D - different gray level from left to right

U_N - different gray level from top to bottom

2.3 Block diagram of the algorithm

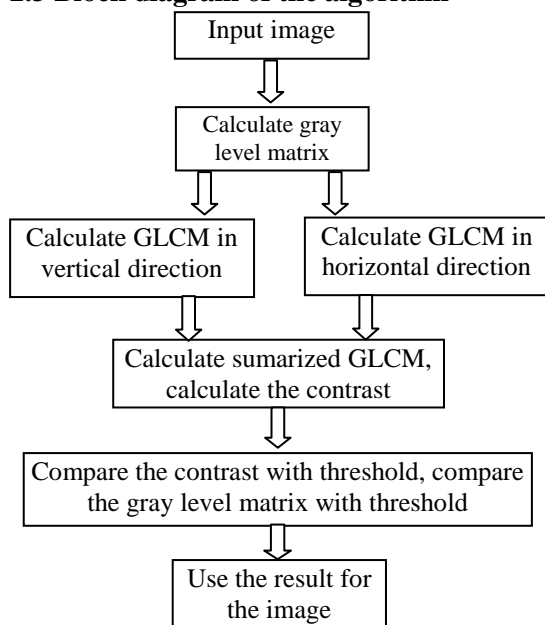


Fig. 3 The algorithm of calculating GLCM applied to software

III. RESULTS

3.1 Functions of the software

- Reading, editing, deleting and saving image.
- Representing and balancing image histogram.
- Filtering noise and smoothing image.
- Extracting image edge.
- Detecting cancrification and microcancification.
- Detecting masses.

3.2 Results of GLCM calculation and application

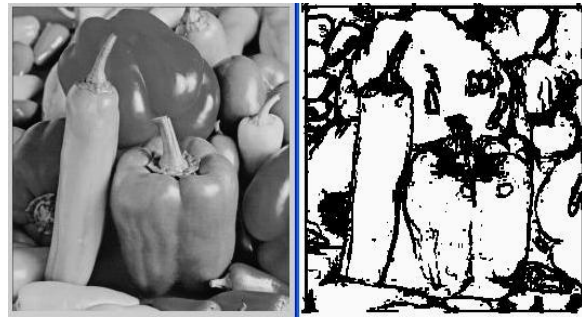


Fig. 4 Before and after thresholding image using GLCM (applied to normal image)

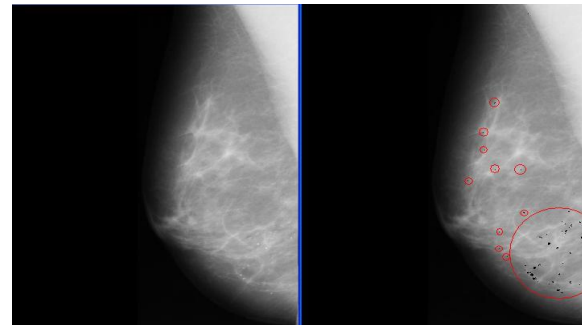


Fig. 5 Before and after thresholding image using GLCM for detecting microcalcification (applied to mammogram)

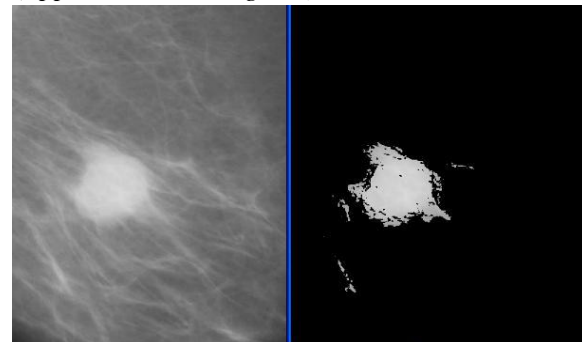


Fig. 6 The mass before and after extracting using GLCM

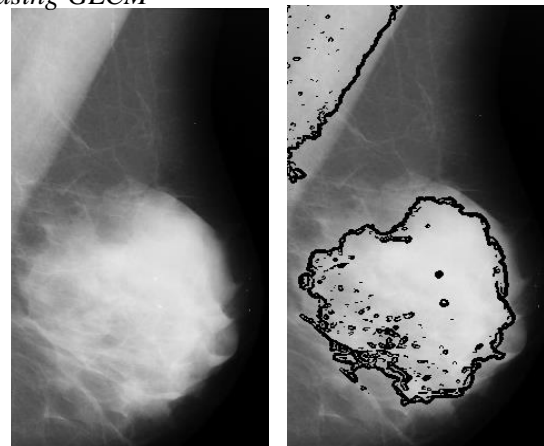


Fig. 7 Image of dense glandular tissue was extracted using GLCM

3.3 Result evaluation

As show in Fig.4, for normal images, the proposed method provides a clear extraction of boudary. For radiographs of non-dense glandular tissue, almost all microcalcifications and massed are detected if a proper threshold is chosen (the probability of correct detection is about 70%) (as shown in Fig. 5 and Fig. 6), and with images of dense glandular tissue, the error is relatively high (Fig. 7).

In this method, the threshold is selected manually. Due to properties of mammograms and methods of imaging, the brightness varies among images, and thus these thresholds cannot be applied to all images. A method of statistics or a method that automatically thresholds basing on image brightness is required.

This method can only segment the microcalcifications without indicate the

malignance or benigance of the microcalcifications. Therefore the algorithm continues to be developed with the purpose of feature extraction and recognition.

IV. CONLUSIONS

Due to properties of mammograms, conventional boudary extraction algorithms employing derivation, differentiation are not suitable in detecting and extracting boudaries of microcalcifications. The application features extracted from GLCM can be served as a new approach for boudary extraction. When applied to mammograms, it gives satisfactory segmentation results. It is rather hasty now to state the success of this method but the result is resonable. The algorithm is still being researched and developed and will be packed into a perfect product to aid in breast cancer diagnosis at hospitals and in study as well.

REFERENCES

1. John Terry Sample, Computer assisted screening of digital mammogram images; University of Southern Mississippi, 2003
2. <http://ungthu.net.vn/?a=readmore&id=CK8PARPQ1QNSIQP12>
3. *Ian T.Young, Jan J.Gerbrands, Lucas J. Van Vliet* (1998); Fundamentals of Image Processing, Delf University of Technology.
4. *Martin Masek*; Hierarchical segmentation of mammograms based on pixel intensity, The University of Western Australia, 2004.
5. Interactive Mammography analysis web tutorial
<http://sprojects.mmi.mcgill.ca/Mammography/anat.htm>
6. Mohammad Sameti – Detection of soft tissue Abnormalities in mammographic Images for early diagnosis of Breast cancer, University of Waterloo, 1994.
7. *Luong Manh Ba, Nguyen Thanh Thuy*; Introduction to digital image processing; Science and Technology Publishing House, 2003
8. *Bich Thuy, Hai Yen*; Handbook of breast and breast cancer; Medical Publishing House, 1997
9. <http://www.fp.ucalgary.ca/mhallbey/tutorial.htm>
10. <http://matlab.izmiran.ru/help/toolbox/images/enhanc15.html>
11. *Nguyen Thai Ha, Do Thanh Nam*; Software of quality control for X ray equipment; Proceeding of the second international conference on the development of Biomedical engineering in Vietnam, 7-2007.
12. *Nguyen Thu Van, Nguyen Tien Dung, Nguyen Thai Ha*; Computer aided diagnosis (CAD) system for detection of breast cancer in X-ray images; Proceedings of IEEE Second International on Communications and Electronics (ICCE 2008), Hoian 6-2008.
13. *Nguyen Thai Ha, Nguyen Duc Thuan, Nguyen Thu Van*; Building the mammographic image database, which supports diagnosing and saving digital mammograms in Vietnam hospitals; Journal of Science & Technology of Universities of Technology, No 67/2008.

Contact: Nguyen Thai Ha - Tel: (+844)3868.2099, Email: thaiha_bk@yahoo.com
Hanoi University of Technology - No. 1, Dai Co Viet road, Hanoi, Vietnam