Computer Aided Diagnosis for Breast Cancer Screening

Piyamas Suapang^{a,*}, Chadaporn Naruephai^b, Sorawat Chivapreecha^c

^a Biomedical Engineering Program, Department of Physics, Rangsit University, Pathumthani, 12000, Thailand

^{b,c} Department of Telecommunications Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand

* piyamas_suapang@yahoo.com

Abstract

Mass segmentation and density classification is an important for breast cancer screening. The purpose of this study is to develop the computer aided diagnosis for mass segmentation and density classification according to the fourth edition of BI-RADS criteria in digital mammography. The digital mammography was digitized high resolution in the image acquisition phase. After the digitization process, an active contour algorithm was applied for mass segmentation. Finally, percentage of mammographic density was calculated for density classification according to the fourth edition of BI-RADS lexicon. The study includes 100 digital mammography of women aged 33-81 years. The results show that the overall accuracy of computerized method classification is 88%.

Keywords: mammography, mass segmentation, density classification, BI-RADS.

1. Introduction

In current mammography imaging practice, there are basically two types of normal tissue distinguishable in the images. One is dense tissue, which is a two component mixture of stromal and epithelial tissue, appearing bright in the image and other is fatty tissue, which appears dark. The fundamental difficulty in either human or computerized breast image analysis is that dense normal tissue and abnormal tissue often have similar x-ray attenuations with respect to the x-ray spectrum in conventional imaging practice, which results in similar image intensities; also the textures are similar. A sample mammogram displaying the breast anatomy is shown in figure 1.

Fig. 1. Mammographic Breast Anatomy.

Mammographic density is an important predictor of risk of breast cancer. Typically, the radiologists classify mammographic density according to fourth edition of Breast Imaging Reporting and Data System (BIRADS) [1, 2], which developed by the American College of Radiology (ACR). BI-RADS descriptors are important factors for predicting malignancies that are assessed and provided by the radiologist. The fourth edition of BI-RADS criteria defined breast density in four categories:

- BI-RADS I: the breast is almost entirely fatty.
- BI-RADS II: there is some fibroglandular tissue.
- BI-RADS III: the breast is heterogeneously dense.
- BI-RADS IV: the breast is extremely dense.

The difference among experience and expertise of each radiologists causes and error of mammographic density classification. Discrepancies as high as 46% have been reported in mammographic density assessment by two different radiologists. Several researchers have pointed out the possibility to employ computer-aided diagnosis (CAD)

schemes to assist radiologists in the interpretation of mammograms. During the last decade CAD has been largely developed to improve the diagnostic accuracy in both the detection and classification of masses, that may indicate the presence of breast cancer [3, 4]. Furthermore, medical images segmentation is an important work to object recognition of the human organs, such as breast, lungs and ribs, and it is an essential pre-processing step in medical image segmentation [5, 6]. The work of the segmentation decides the result of the final processed image.

This paper presented here is an important component of an ongoing project of developing an mass segmentation and density classification system for breast cancer screening and diagnosis for digital mammography applications. Specifically, in this work the task of separating mass tissue from normal breast tissue given a region of interest in a digitized mammogram is investigated. This is the crucial stage in developing a robust system because the classification depends on the accurate assessment of the tumor-normal tissue border as well as information gathered from the tumor area.

2. Methodology

The element of the overall system scheme of computer aided diagnosis for mass segmentation and density classification according to the fourth edition of BI-RADS criteria in digital mammography is shown in figure 2. The mammograms are digitized in the image acquisition phase while the primary region of interest is located in the second phase called ROI. The next two phases: segmentation and classification cover the main processing steps of the mass segmentation and density classification system that is under development here at this facility.

Fig. 2. Block diagram for the element of the overall system scheme.

2.1 Digitization of the Mammography Images Module

The digitization of mammographic images module is image and data archiving in DICOM format from mammogram modality into database using Matrox Morphis (MOR/2VD/84*) capture card. This image acquisition is controlled through AVICAP Windows Class, which interface between Application and Device Driver. This is where patient information is entered in, the digitizer is accessed for scanning, and where any necessary changes can be done to the image signal before the digital image is saved.

Screening Mammography is a low-dose x-ray examination of the breasts in a woman who is asymptomatic. The Screening Mammograms are two x-ray views for each breast, typically cranial-caudal view, (CC) and mediolateral-oblique (MLO). These views that is shown in figure 3, are: Left Mediolateral Oblique View (LMLO), Right Mediolateral Oblique View (RMLO), Left Craniocaudal View (LCC), and Right Craniocaudal View (RCC).

Fig. 3. Typical Mammography views seen for each case (patient): (a). LMLO, (b). LCC, (c). RMLO and (d). RCC.

2.2 Mass Segmentation and Density Classification

Computer aided diagnosis for breast cancer screening according to the fourth edition of BI-RADS criteria in digital mammography is important distinctions between mass segmentation and density classification.

After the digitization process, it is valuable to develop a computer aided method for mass classification based on extracted features from the breast and region of interests (ROI) in digital mammography. This would reduce the number of unnecessary biopsies in patients with benign disease and thus avoid patients, physical and mental suffering, with an added bonus of reducing healthcare costs.

The ROI extracted from the breast region are shown in figure 4 and mass extracted from the ROI both for malignant and benign are shown in figure 5 that need to be classified and it is the main objectives of the paper. An active contour algorithm was employed for mass boundary segmentation with the equation (1); after the energy image on version of the image was smoothed.

Fig. 4. The breast border was initially determined snakes position for sample ROI extracted from breast region.

Fig. 5. The sample mass extracted from the ROI.

$$\frac{\partial v}{\partial t} = \alpha \frac{\partial^2 v}{\partial s^2} - \beta \frac{\partial^4 v}{\partial s^4} + k f_E + \lambda f_B \quad (1)$$

where $v(s, t)$ is the snake curve as a function of the arc length s and time t . The constants α and β control the elasticity and stiffness of the snake, and f_E is the external force. The balloon force f_B is perpendicular to the snake curve and is used to make the snake grow or shrink. The evolution equation is interested in a steady-state solution $v(s)$ for $t \rightarrow \infty$. To simplify tuning the parameters $\alpha, \beta, k, \lambda$, this paper take unit time steps $\Delta t = 1$ and assume that the external force f_E is normalized to have maximum magnitude equal to 1 (pixel/step). The balloon force f_B has magnitude one everywhere. The snake $v(s)$ at time t is represented by two vectors containing the x and y coordinates of a sequence of points on the snake curve. The distance between subsequent points is maintained close to 1 pixel. Snake combine image data and elasticity constraints to iteratively refine an initial, coarsely located contour. However, the elasticity constraints and the gradient based driving forces generally prevent snakes t enter into invaginated structures and correctly follow their contours.

2.3 Mammographic Density Classification

BI-RADS descriptors are important factors for predicting malignancies that are assessed and provided by the radiologist. The mass narratives include the overall shape description, the border region margin regularity, and the relative intensity of the mass region compared with the ambient normal tissue intensity. The BI-RADS also provides a four-category rating for assessing the overall breast tissue characteristic in terms of the fibroglandular composition. The intensity or the x-ray attenuation of the mass tissue region is described as density. The density here is the relative density, i.e. higher, lower or similar relative to the surrounding tissue. The density is rated on 4-point system.

The segmented mass area and total of breast tissue area were selected on images to obtain the number of pixels. The number of pixels in the segmented mass area and the number of pixels in the total of breast tissue area were calculated for the percentage of mammographic density, which are determined by the equation (2).

$$\%MD = \frac{NPS \times 100}{NPB} \quad (2)$$

where $\%MD$ is the percentage of mammographic density.

NPS is the number of pixels in the segmented mass area. NPB is the number of pixels in the total of breast tissue area. Finally, the percentage of mammographic density is calculated and classified with BI-RADS criteria.

3. Results and Discussion

One hundred cases of women aged 33-81 years (mean 52) have been digitized using the Matrox Morphis (MOR/2VD/84*) frame grabber board. All these cases have been diagnosed with cancer. However, the system also have the benign and normal images of some of these cases (that is screening mammograms done before the cancer had occurred). Hence, the system can compare normal, benign, and cancer images of the same case. The image quality of all the images is excellent and high resolution (RGB 32 bits per pixel with resolution 720 x 676 pixels). In the digitization of mammographic images module (Figure 6) has been developed using Borland C++ Builder programming language. The system takes as inputs pictures in most of the popular bitmap, gif, jpeg, tiff, png formats as well as DICOM. A preliminary viewer has been developed using Borland C++ Builder programming language. It can accept .bmp, .tif, .jpg, .gif, and .png image formats. Furthermore, it can perform basic image manipulation such as rotation, resizing, filter inversion, grayscale, contrast enhancement, sharpen, and zoom. Finally, work on the DICOM viewer is currently going on to improve its functionality. For example, the system wants to be able to draw a marker around a region of interest (such as a cancerous mass) on a DICOM image.

The segmented mass area (fibroglandular area) and total of breast tissue area were selected on images to obtain the number of pixels. The number of pixels in the segmented mass area and the number of pixels in the total of breast tissue area were calculated for the percentage of mammographic density, which are determined by the equation (2). Table 1 shows that the examples of percentage of mammographic density classification is calculated and classified with BI-RADS criteria.

Table 2 shows that the overall accuracy of computerized method classification is 88% (88/100). The BIRADS I reach 95.45% (21/22) correct classification, BIRADS II 96.29% (26/27), BIRADS III 81.81% (18/22) and IV 79.31% (23/29). In conclusion, the computerized method based on active contour segmentation would be useful as the radiologist assistant for mammographic density classification. Nevertheless, this method is suitable for fatty

breast, but mammographic density classification errors may occur in dense breast. The work on the segmentation and classification are currently going on to improve its functionality. For example, the system wants to be able to draw a marker around a region of interest (such as a cancerous mass) on a DICOM image. However, much more work needs to be done on the system in order to make it acceptable by radiologists for use in diagnosis.

Fig. 6 The digitization of mammographic images module.

Table 1. Sample of mammographic mass segmentation and density classification is calculated and classified with BI-RADS criteria.

Fibroglandular Tissue	The Percentage of Mammographic Density	BIRADS Criteria	Classifying Mammographic Density from Database
12.84	1	1
32.76	2	2
61.28	3	3
75.43	4	4

4. Conclusion

This study is to develop the computer aided diagnosis for mass segmentation and density classification according to the fourth edition of BI-RADS criteria in digital mammography. The digital mammography was digitized high resolution in the image acquisition phase. After the digitization process, an active contour algorithm was applied for mass segmentation. Finally, percentage of mammographic density was calculated for density classification according to the fourth edition of BI-RADS lexicon. The study includes 100 digital mammography of women aged 33-81 years. The results show that the overall accuracy of computerized method classification is 88%. The significance of the computerized method based on active contour segmentation is useful as the radiologist assistant for classifying mammographic density. Nevertheless, this method is suitable for fatty breast, but mammographic density classification errors may occur in dense breast. However, much more work needs to be done on the viewer in order to make it acceptable by radiologists for use in diagnosis.

Table 2. Mammographic density classification is obtained by radiologists and active contour algorithm.

		Mammographic Density Classification based on Active Contour Algorithm			
		BIRADS	BIRADS	BIRADS	BIRADS
		I	II	III	IV
Mammographic Density Classification by Radiologists	BIRADS I (<25%)	21	1	0	0
	BIRADS II (26-49%)	1	26	0	0
	BIRADS III (60-76%)	1	3	18	0
	BIRADS IV (>76%)	0	2	4	23

Acknowledgment

This work is partially supported by Rangsit Research Institute at Rangsit University, Department of Telecommunications Engineering at King Mongkut's Institute of Technology Ladkrabang, and Department of

Industry Physics and Medical Instrument at King Mongkut's University of Technology North Bangkok. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

- (1) J.N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern", *Cancer*, vol. 37, pp. 2486-2492, 1976.
- (2) American College of Radiology, "Illustrated Breast Imaging Reporting and Data System BIRADS", *American College of Radiology*, 3rd edition, 1998.
- (3) Yin, F. F., Giger, M. L., Doi, K., Vyborny, C. J., Schmidt, R. A., "Computerized detection of mass in digital mammograms: Automated alignment of breast images and its effect on bilateral subtraction technique", *Medical Physic*, vol 21, pp. 445-452, 1994.
- (4) Mendez, A.J., Tahoces, P.G., Lado, M.J., Souto, M., Vidal, J.J., "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms", *Medical Physics*, vol 25, pp. 957-964, 1998.
- (5) M.I. Rajab, M.S. Woolfson, and S.P. Morgan, "Application of region-based segmentation and neural network segmentation to skin lesions", *Computerized Medical Imaging and Graphics*, vol. 28, pp. 61-68, 2004.
- (6) H. Tang, E.X. Wu, Q.Y. Ma, D. Gallagher, G.M. Perera, and T.Zhuang, "MRI brain image segmentation by multi-resolution segmentation and region selection", *Computerized Medical Imaging and Graphics*, vol 24, pp. 349-357, 2000.
- (7) M. Kass, A. Witkin, D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Visio*, vol. 1, pp. 321-331, 1988.
- (8) G. Aubert and P. Kornprobst, "Mathematical Problems in Image Processing", *Germany: Springer Verlag*, pp. 153-179, 2001.
- (9) C. Xu and J. L. Prince, "Snakes, Shapes, and Gradient Vector Flow", *IEEE Transactions on Image Processing*, vol. 7(3), pp. 359-369, 1998.
- (10) M.I. Rajab, M.S. Woolfson, and S.P. Morgan, "Application of region-based segmentation and neural network segmentation to skin lesions", *Computerized Medical Imaging and Graphics*, vol. 28, pp. 61-68, 2004.

- (11) H. Tang, E.X. Wu, Q.Y. Ma, D. Gallagher, G.M. Perera, and T. Zhuang, "MRI brain image segmentation by multi-resolution segmentation and region selection", *Computerized Medical Imaging and Graphics*, vol 24, pp. 349–357, 2000.
- (12) Huertas, A. and Medioni, G., "Detection of intensity changes with sub pixel accuracy using Laplacian-Gaussian masks", *IEEE Trans. On Pattern Analysis and Machine Intelligence*, PAMI, vol. 8, pp. 651–664, 1986.
- (13) University of South Florida, "Digital Mammography Home Page", 2012. University of South Florida. Available: <http://figment.csee.usf.edu/Mammography/Database.html>.
- (14) D. Axmark and M. Widenius, "MySQL Reference Manual", 1995. MySQL. 15 Oct. 2011. Available: <http://dev.mysql.com/doc/mysql/en/index.html>.
- (15) G. Hojtsy, "PHP Manual", 1997, PHP. 15 Oct. 2011. Available: <http://www.php.net/manual/en/>.
- (16) R. J. Ferrari, R. M. Rangayyan, J. E. Desautels, R. A. Borges, and A. F. Frere, "Automatic Identification of the Pectoral Muscle in Mammograms", *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, 2004.
- (17) American College of Radiology. American College of Radiology, "Breast Imaging Reporting and Data System (BI-RADS)", 4th ed, Reston, Va: American College of Radiology, 2003.
- (18) Ho WT, Lam PWT, "Clinical Performance of Computer-assisted Detection (CAD) System in Detecting Carcinoma in Breasts of Different Densities", *Clinical Radiology*, vol 58(2), pp. 133-6, 2003.
- (19) Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E., "Percentage density: Wolfe's and Tabár's mammographic patterns: agreement and association with risk factors for breast cancer", *Breast Cancer Research*, vol 7, pp. 854-61, 2005.
- (20) Zhou C, Chan HP, Petrick N, Helvie MA, Goodsitt MM, Sahiner B, Hadjiiski LM., "Computerized image analysis: Estimation of breast density on mammograms", *Med. Phys.*, vol 28, pp. 1056–69, 2001.
- (21) Ciatto S, Houssami N, Apruzzese A, et al., "Categorizing breast mammographic density: Intra- and inter-observer reproducibility of BI-RADS density categories", *Breast J*, vol 14, pp. 269–75, 2005.
- (22) Blot L, Zwiggelaar R., "Background texture extraction for the classification of mammographic parenchymal patterns", *In MIUA*, pp. 145-8, 2001.
- (23) Oliver A, Freixenet J, Marti R, et al., "A Novel Breast Tissue Density Classification Methodology", *IEEE Transactions on Information Technology in Biomedicine*, vol 12(1), pp. 55-65, 2008.
- (24) Sheshadri HS, Kandaswamy , "A. Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms", *Computerized Medical Imaging and Graphics*, vol 31, pp. 46–8, 2007.
- (25) Yaffe MJ. "Mammographic density: Measurement of mammographic density". *Breast Cancer Research* 2008, vol 10, 2009.
- (26) Baker JA, Rosen ML, Crockett MM, Lo JY. "Accuracy of Segmentation of a Commercial Computer-aided Detection System for Mammography". *Radiology*, vol 235, pp. 385–390, 2005.