Density Based Breast Segmentation for Mammograms Using Graph Cut Techniques

Nafiza Saidin, Umi Kalthum Ngah, Harsa Amylia Mat Sakim, Ding Nik Siong , Mok Kim Hoe
School of Electrical and Electronic Engineering,
Universiti Sains Malaysia,
14300 Nibong Tebal, Pulau Pinang, Malaysia.
fizasep_98@yahoo.com, umi@eng.usm.my, amylia@eng.usm.my

***Abstract*—In this work we explore the application of graph cuts techniques to the problem of finding the boundary of different breast tissue regions in mammograms. The goal of the segmentation algorithm is to see if graph cuts algorithm could separate different densities for the different breast patterns. The graph cut is applied with multi-selection of seeds label to provide the hard constraint, whereas the seeds labels are selected based on user defined. Graph cuts have been explored on images of various imaging modalities but not on mammograms just yet. Therefore, this project is mainly focused on using graph cut algorithm to perform segmentation to increase visibility of different breast densities in mammography images. Segmentation of the mammogram into different mammographic densities is useful for risk assessment and quantitative evaluation of density changes. Our proposed methodology for the segmentation of mammograms on the basis of their region into different densities based categories has been tested on MIAS database.**

Keywords-image processing; medical imaging; image segmentation

I. INTRODUCTION

Breast cancer is the most common cancer in women in most parts of the world. In western countries, it is calculated that between one out of eight and one out of twelve women will develop breast cancer during their lifetime [1], [2]. This proportion is reduced in Malaysia, where it is estimated that one in twenty women will develop breast cancer during their lifetime [3]. However, breast cancer is the leading cause of cancer deaths in women in Malaysia. The most successful method for the early detection of breast cancer is screening mammography. Currently, mammograms are analyzed visually by radiologists. Because of the subjective nature of visual analysis, qualitative responses may vary from radiologist to radiologist. However, it is well known that expert radiologists can miss a significant proportion of abnormalities [4]. In addition, a large number of mammographic abnormalities turn out to be benign after biopsy [5]. Therefore, a computerized method for analyzing mammographic features would be useful as a supplement to the radiologist's assessment [5], [6] and commercial systems are also available [7], [8].

Previous research efforts in computer-aided diagnosis (CAD) for breast cancer detection mainly concentrated on detection and characterization of masses and microcalcifications on mammograms by using computer vision techniques [9], [10]. It has been demonstrated that an effective CAD algorithm can improve the diagnostic accuracy of breast cancer characterization on mammograms, which, in turn, may reduce unnecessary biopsies.

In Asia, there has been a rapid increase in the incidence of breast cancer and the disease may occur at relatively young age [3]. Young women have breasts that are dense and full of milk glands, sometimes making mammograms difficult to interpret. Breast density is an important factor in the interpretation of a mammogram. It has been shown that the accuracy of mammographic abnormality detection methods is strongly dependent on the breast tissue characteristics, where a dense breast drastically reduces detection sensitivity. In addition, breast tissue density is widely accepted to be an important risk indicator for the development of breast cancer [11]–[13]. Therefore, segmentation of the mammogram into different mammographic densities is useful for risk assessment and quantitative evaluation of density changes.

In this study, a new algorithm called graph cut was explored to evaluate its efficiency to segment different breast region according to their density in mammograms. Yuri Boykov and his co-researchers have improved a well-known graph based method called the graph cut. Graph cut technique enables objects in medical images to be reliably segmented by finding their precise boundaries. Precise segmentation would allow physicians to make accurate measurements, simplify visualization and consequently, make the diagnosis more reliable. Existing research on graph cut technique research papers have shown positive results in segmentation of medical images such as magnetic resonance imaging (MRI) and computed tomography (CT) scan images. Although the graph cut technique is showing very promising outcome, work on using graph cut technique on mammograms has yet to embark. Through this study, the advantages of graph cut techniques for segmentation of different breast regions can be established.

978-1-4244-4547-9/09/$26.00 ©2009 IEEE TENCON 2009

II. BACKGROUND

The first to have shown the relation between mammographic parenchymal patterns and the risk of developing breast cancer was Wolfe [11], who classified the parenchymal patterns in four categories. Breast Imaging Reporting and Data System (BIRADS) [10] also classified breast density into four categories. Miller and Astley [14] investigated texture-based discrimination between fatty and dense breast types. Byng *et al* [15] used measures based on fractal dimension. Bovis and Singh [16] estimated features from the construction of spatial gray level dependency matrices. Petroudi *et al* [17] used textons to capture the mammographic appearance within the breast area. Petroudi and Brady [18] described an algorithm to segment mammographic images into regions corresponding to different densities. The segmentation algorithm used textons in a Hidden Markov Random Field (HMRF). The results of the algorithm demonstrated close agreement to radiologist's segmentation and density interpretation. Torrent *et al* [19] compared two clustering based algorithm and one region based algorithm to segment fatty and dense tissue in mammographic images. The first algorithm is a multiple thresholding algorithm based on the excess entropy (EE), the second one is based on the Fuzzy C-Means clustering algorithm (FCM), and the third one Fisherfaces (FF) is based on a statistical analysis of the breast. The results showed that the region based approach allowed to obtain single and homogeneous regions. And there are several research that used intelligence system for density classification such as probabilistic Latent Semantic Analysis (pLSA), *k*-nearest neighbors (kNN) classifier, a decision tree classifier, and a Bayesian classifier. Bosch *et al* [20] proposed a new approach to model and classify breast parenchymal tissue using pLSA. Chatzistergos *et al* [21] worked aims at the classification of breast tissue according to Breast Imaging Reporting and Data System (BIRADS) using pLSA. Oliver *et al* [22] used of *k*-nearest neighbors (kNN) classifier, a decision tree classifier and a Bayesian classifier, based on the combination of the first two classifiers in their research.

The main goal of this study is to explore graph cut techniques developed by Yuri Boykov and his co-researchers to segment different breast tissue regions which correspond to the density in mammograms. Graph cuts have been explored on images of various imaging modalities but not on mammograms just yet. Therefore, this project is mainly focused on using graph cut algorithm to perform segmentation to increase visibility of different breast densities in mammography images. This would assist radiologists to further classify and characterize the breast tissue more accurately. And also the dense area using graph cuts algorithm can be used as input features for an intelligent system to classify the breast pattern according to BIRADs categories. Our proposed methodology for the segmentation of mammograms on the basis of their region into different densities based categories has been tested on MIAS database [23].

III. MAMMOGRAMS SEGMENTATION USING GRAPH CUTS

Graph cuts are segmentation techniques that divide the image into two parts, called "object" and "background". The minimum cut of the graph will determine the energy function to be minimized either locally or globally. Interactive segmentation requires a user to indicate certain pixels to be part of "object" or "background". This is the hard constraints imposed by the user which provide clues on the parts that needed to be segmented.

Graph cuts technique is based on combinatorial fact that a globally minimum cut of a graph with two terminals can be computed efficiently in low order polynomial time [24], [25], [26]. It should be noted that graph cuts are not newly developed techniques and have been used before in image segmentation. In the work done by Wu and Leahy, the image is divided in K parts to minimize the maximum cut between the segments [27]. However, the result of segmentation using this formulation is strongly biased to very small segments. Shi and Malik tried to solve this problem by normalizing the cost of cut [28]. Implementing the standard graph cuts in a straight forward manner can be quite slow. However, Boykov and Kolmogorov have compared several graph based methods in vision and also described a newer version of the max-flow algorithm that performs much better than the standard techniques [29]. Therefore, the implementation of the interactive segmentation method in this paper is based on the new graph cut algorithm from the same author.

The algorithm of graph cuts were developed using C++ programming language with the aids of software library developed by Olga Veksler [30, 31]. The software library is named as Multi-label MRF Optimization where it implements the Graph Cuts Energy Minimization methods described by Boykov *et al.* [32]. This library has been used by Google, Microsoft and others to solve problems (like stereo, segmentation, optical flow, and etc.) that have multi-label Markov Random Field formulations. The algorithms provided are called α-expansion and α/β-swap and were part of Olga Veksler's PhD thesis work [30]. The α-expansion algorithm is used in the code for energy minimization.

The equation is modified in relation to the software library and is represented as follows:

$$E(f) = \lambda \cdot \sum_{p \in P} \mathbf{Dp}(Lp) + \mu \cdot \sum_{\{p,q\} \in N} \mathbf{V}\{\mathbf{p},\mathbf{q}\} \cdot \delta(\, Lp \neq Lq) \qquad (1)$$

where $L = \{Lp \mid p \in \mathcal{P}\}$ is a labeling of image P. The first term on the right hand side of this equation is called datacost and it is also known as the regional properties term by Boykov and Funka-Lea [33]. $D_p(.)$ It is a data penalty function and it indicates individual label preferences of pixels based on observed intensities and pre-specified likelihood function [29]. While the second term is called smoothcost and it also called as boundary properties term by Boykov and Funka-Lea [33]. $V_{p,q}$ is an interaction potential and it encourages spatial coherence by penalizing discontinuities between neighboring pixels. Notice that there are two constants, λ and μ which correspond to datacost and smoothcost. The purpose of adding the two constants is to obtain the optimal segmentation for different types of image. Adjusting the value of λ and μ will affect the result of segmentation and are set by trial and error. Implementation of the graph cut algorithm is summarized in Figure 1. The most important part in this implementation is defining datacost and smoothcost. These two parameters affect the most on the result of segmentation and thus the equations have to be defined properly.

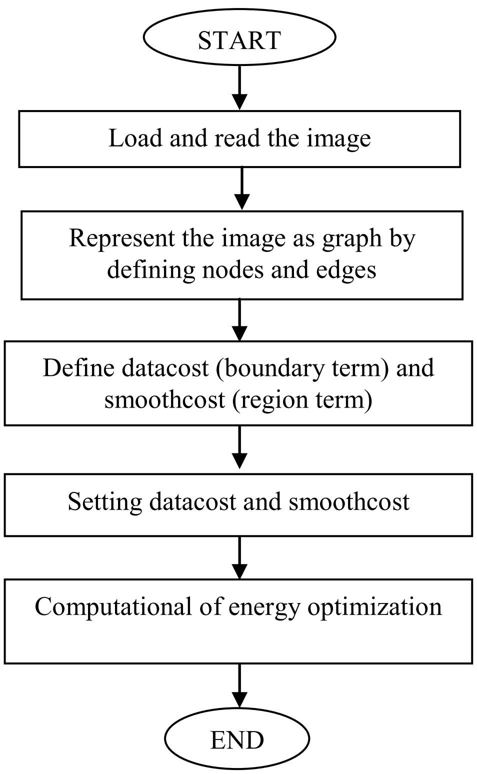


Figure 1. Flow Chart of Implementation of Graph Cuts algorithm

In this paper, we explore potential of graph cuts for breast region segmentation. This work also tries to colour-code each layer that represents different breast tissue. It is a very difficult task because mammograms have an inherent "fuzzy" or diffuse appearance. Users are required to mark near a certain region to provide hard constraints for the segmentation. Initially, the pixels of the image are represented as nodes of graph. Then, the image will be partitioned into two segments, “object” and “background” based on s-t cuts. Finally, to find and perform globally optimum segmentation, graph cuts are used for this purpose. Some results are presented in the context of breast mammography segmentation in section IV. This segmentation algorithm has the capability to segment an image up to ten different mark labels. But in this experiment, only five to six mark labels will be chosen depending on how many different breast tissues need to be highlighted. After marking, the graph cuts algorithm is compiled and run. The user can repeat the process by repositioning the seeds marks until the user is satisfied with the results. The user can repeat the segmentations until the abnormalities are clearly shown.

IV. EXPERIMENTAL RESULTS

The experiments have been done under different breast pattern. In segmentation using graph cuts; it is based on user-defined seeds label. The graph cuts is applied with multi-selection of seeds label, which the first label for finding the background boundary, the second label for finding the skin boundary, the third label for finding the fatty boundary, the fourth label for finding the dense boundary and the fifth label for finding the pectoral muscle boundary. We have performed a series of experiments investigating the use of graph cuts. For every test image, there are three output images. The first image shows the original image, while the second image shows marked seeds by the user. The third image shows segmentation output in color and the fourth image shows the segmentation output in the grayscale.

A. Experiment 1: Fatty Tissue

The first experiment deals with a mammogram which is predominantly comprised of fatty tissue. Figure 2 shows the result for MIAS image mdb028, which is malignant case and had well-defined masses. There are five seed labels selected for mdb028 image. The seed labels are marked by user.

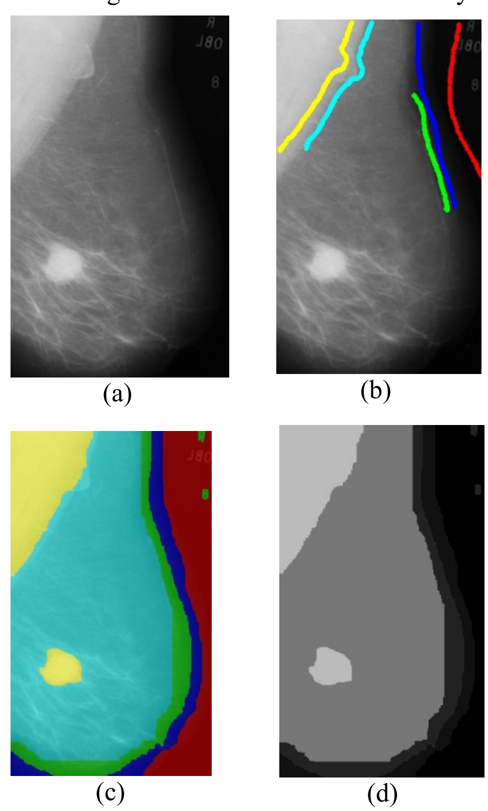
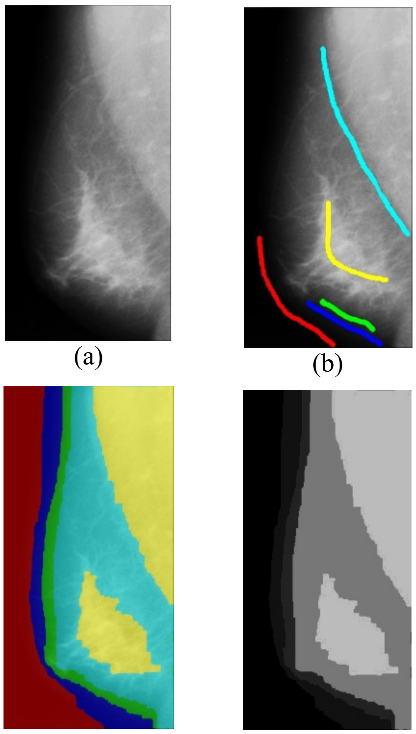


Figure 2: Segmentation of (a) original image mdb028, (b) using 6 marked seed labels and output segmented image with (c) color label and (d) grayscale label.

B. *Experiment 2: Fatty-Fibroglandular Tissue*

The second experiment deals with a mammogram which is predominantly comprised of glandular tissue. Figure 3 shows the result for MIAS image mdb233, which is malignant case. There are five seed labels selected for mdb233 image.

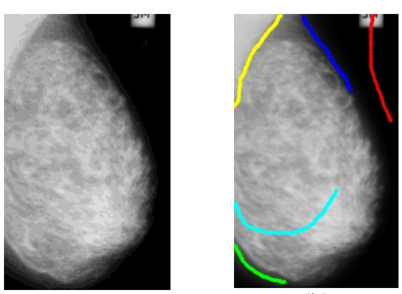


(a) (b) (c) (d)

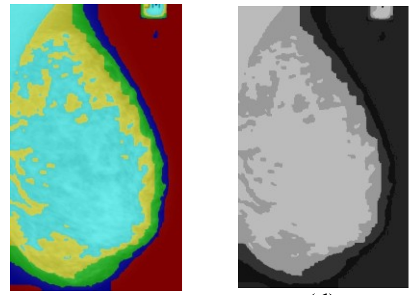
Figure 3: Segmentation of (a) original image mdb233, (b) using 5 marked seed labels and output segmented image with (c) color label and (d) grayscale label.

C. *Experiment 3: Dense -Fibroglangular Tissue*

The third experiment deals with a mammogram which is predominantly comprised of dense tissue. Figure 4 shows the result for MIAS image mdb130, which is malignant case. There are five seed labels selected for mdb130 image.



(a) (b)



(c) (d)

Figure 4: Segmentation of (a) original image mdb130, (b) using 5 marked seed labels and output segmented image with (c) color label and (d) grayscale label.

We have presented a method for segmenting the breast to areas of different density which investigates the use of graph cuts algorithm. The results of the segmentation are due to user defined seed labels based on the density feature to provide hard constraints for the graph cuts algorithm, as they combine tissue type and colour information. Experiment is doing and robustly, the result is promising. The experimental results indicate that the graph cuts technique can delineate the breast tissue layer in mammogram including the skin layer, the fatty, the fibroglandular and the dense region. This capability is very helpful for the practitioner or radiologists to analyze different parts of breast.

V. Conclusion and Recommendation

In conclusion, the graph cut segmentation algorithm performs well on mammogram. However, in this preliminary stage, the detection of the boundary is done semi-automatically where the user needs to define and mark the labels. An attempt will be made to automatically demarcate each portion according to the boundaries. Therefore, the next direction for this study is to create an automated segmentation algorithm without the need for the user to mark the seed points into the image. With this capability, the detection of mammogram abnormal structures will be more simplified and faster to run. One of the solutions would be to integrate the clustering technique into the graph cut algorithm. The objective of having the clustering technique is to divide the image into several different regions. These different regions can be used as labeling seeds for the graph cut technique. The result using graph cuts for the dense area also can be used as input feature for intelligence system, which can make classification of breast pattern according to BIRADs categories. This is important for risk assessment of breast cancer. More modifications to the algorithm would be explored. The presented method will overcome difficulties due to the breast variability achieving a good representation of region and density in the breast, thus providing an excellent base for density risk assessment.

References

[1] Eurostat, "Health statistics: Atlas on mortality in the European Union," Office for Official Publications of the European Union, 2002.

[2] F. Bray, P. McCarron, and D. M. Parkin, "The changing global patterns

of female breast cancer incidence and mortality," Breast Cancer Res., vol. 6, pp. 229–239, 2004.

[3] C. H. Yip, N. A. M. Taib and I. Mohamed, "Epidemiology of breast cancer in Malaysia," Asian Pacific Journal of Cancer Prevention, vol. 7, pp. 3690-374, 2006.

[4] R. L. Birdwell, D. M. Ikeda, K. D. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," Radiology, vol. 219, pp. 192–202, 2001.

[5] F. M. Hall, J. M. Storella, D. Z. Siverstond, and G. Wyshak, "Nonpalpable breast lesions: Recommendations for biopsy based on suspicion of carcinoma at mammography," Radiology, vol. 167, pp. 353–358, 1988.

[6] T. W. Freer and M. J. Ulissey, "Screening mammography with computeraided detection: Prospective study of 12860 patients in a community breast center," Radiology, vol. 220, pp. 781–786, 2001.

[7] R2 ImageChecker. (Jan. 2007) [Online]. Available: http://www.r2tech.com.

[8] iCAD Second Look. (Jan. 2007) [Online]. Available: http://www.icadmed.com.

[9] L. Tabar, T. Tot, and P. B. Dean, "Breast Cancer—The art and science of early detection with mammography," Stuttgart, Germany: Georg Thieme Verlag, 2005.

[10] American College of Radiology, "Illustrated Breast Imaging Reporting and Data System BIRADS," 3rd ed. Philadelphia, PA: Amer. College of Radiol, 1998.

[11] J. N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern," Cancer, vol. 37, pp. 2486–2492, 1976.

[12] N. F. Boyd, J.W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B. Miller, G. A. Lockwood, D. L. Tritchler, and M. J. Yaffe, "Quantitative classification of mammographic densities and breast cancer risk: Results from the Canadian national breast screening study," J. Nat. Cancer Inst., vol. 87, pp. 670–675, 1995.

[13] V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis," Cancer Epidemiol. Biomarkers Prev., vol. 15, pp. 1159–1169, 2006.

[14] P. Miller and S. M. Astley, "Classification of breast tissue by texture analysis," Image Vision Comput., vol. 10, pp. 277–282, 1992.

[15] J.W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J.Yaffe, "Automated analysis of mammographic densities," Phys. Med. Biol., vol. 41, pp. 909–923, 1996.

[16] K. Bovis and S. Singh, "Classification of mammographic breast density using a combined classifier paradigm," in Proc. Med. Image Understanding Anal. Conf., pp. 177–180, 2002.

[17] S. Petroudi, T. Kadir, and M. Brady, "Automatic classification of mammographic parenchymal patterns: A statistical approach," in Proc. IEEE Conf. Eng. Med. Biol. Soc., vol. 1, pp. 798–801, 2003.

[18] S. Petroudi and M. Brady, "Breast density segmentation using texture," Lect. Notes Comput. Sci., vol. 4046, pp. 609–615, 2006.

[19] A. Torrent, A. Bardera, A. Oliver, J. Freixenet, I. Boada, M. Feixes, R. Mart´ı, X. Llad´o, J. Pont, E. P´erez, S. Pedraza, and J. Mart´ı, "Breast density segmentation: A comparison of clustering and region based techniques," (eds.) IWDM 2008. LNCS, vol. 5116, pp. 9–16, Springer, Heidelberg, 2008.

[20] A. Bosch, X. Mu˜noz, A. Oliver and J. Mart´, "Modeling and classifying breast tissue density in mammograms," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2006.

[21] S. Chatzistergos, J. Stoitsis, K.S. Nikita, A. Papaevangelou, "Development of an integrated breast tissue density classification software system," IEEE International Workshop on Imaging Systems and Techniques, pp. 243 – 245, 2008.

[22] A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E. R. E. Denton, and R. Zwiggelaar, "A novel breast tissue density classification methodology," IEEE Transactions on Information Technology in Biomedicine, Vol. 12, pp. 55 – 65, 2008.

[23] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal and J. Savage, "The mammographic images analysis society digital mammogram database," Exerpta Medica. International Congress Series, vol. 1069, pp. 375-378, 1994. Available: http://www.wiau.man.ac.uk/services/MIAS/MIAScom.html.

[24] L. Ford and D. Fulkerson, Flows in networks. Princeton University Press. 1962.

[25] A. Goldberg and R. Tarjan, "A new approach to the maximum flow problem," Journal of the Association for Computing Machinary, 35(4): 921-940, October, 1988.

[26] Y. Boykov and M.P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," International Conference on Computer Vision, vol. I, pp. 105–112, July 2001.

[27] Z. Wu, and R. Leahy, "An optimal graph theoretic approach to data clustering: theory and its application to image segmentation,: IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(11):1101–1113, November, 1993.

[28] J. Shi, and J. Malik, "Normalized cuts and image segmentation," IEEE Conference on Computer Vision and Pattern Recognition, pages 731–737. 1997.

[29] Y. Boykov, and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), September, 2004.

[30] O. Veksler, "Efficient Graph-based Energy Minimization Methods in Computer Vision," PhD thesis, Cornell University, July 1999.

[31] Code. (December, 2008) [Online]. Available: http://vision.csd.uwo.ca/code/.

[32] Y. Boykov, O. Veksler, and R.Zabih, "Fast approximate energy minimization via graph cuts," IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(11):1222–1239, November 2001.

[33] Y. Boykov, and G. Funka Lea, "Graph cuts and efficient N-D image segmentation," International Journal of Computer Vision, 69(2) 109–131, September, 2006.