Review on Segmentation Methods of Mass Detection of Mammographic Images

Tushar P. Kele[1]

[1]Tushar P. Kele, Computer Science and Engg/ D.I.E.M.S/ Dr.BAMU University, Aurangabad, India
[1]kele.tushar@gmail.com

ABSTRACT

According to World Health Organization (WHO), breast cancer is most common cancer in woman in worldwide, becoming to one of the most fatal form of cancer. For breast cancer diagnosis mammography image analysis is a effective tool. And which is based on texture and shape analysis of mammary lesion. The grow-cut algorithm is a segmentation method based on cellular automata. And it able to perform relatively accurate segmentation through selection of internal and external seeds point. Now we proposed adaptive semi-supervised method of grow-cut algorithm based on modification of automaton evolution rule by adding a Gaussian fuzzy membership function in order to define non-defined borders. In our proposal manual selection of seed points is changed by semi-automatic stage, where only internal points are selected by a different evolution algorithm.

Index Term — Breast Cancer, Mass detection, Mammography Image analysis, Biomedical Image Segmentation, Grow-Cut Algorithm.

1. INTRODUCTION

Breast cancer has become an increasing problem for woman in world-wide according to world health organisation (WHO) It is the most common type of cancer in woman, with mortality, both for developed and underdeveloped countries [1]. Breast cancer survival rates can vary between 80% in high income countries, to below 40% in low-income ones [3]. The low income countries is related to lack of screening programs which assist in early detection of cancers. Early detection has an important impact on successfull treatment of cancer. But once medical treatment becomes harder in late stages. One of the most effective method for breast cancer is digital mammography [4].However, mammography visual understanding and analysis can be a hard task even to a specialist and such a procedure can be affected by image quality aspects, radiologist experience and tumor shape. However, mammography visual understanding and analysis can be a hard task even to a specialist and such a procedure can be affected by image quality aspects, radiologist experience and tumor shape. After the beginning of breast cancer, the period until tumors becomes palpable. i.e reaching a diameter around 1 cm, is about 10 years [5].During this period, breast imaging is essential, both for early detection and tumor monitoring. Tumor size takes an imp role in the planning of breast cancer treatment, avoiding mutilating surgeries, such as mastectomy [6]. Furthermore, image analysis and diagnosis are complex, mainly because of large variability of cases so for this MCAD has been playing an import role to assist radiologist and other related health professional im improving the accuracy of their diagnosis. The size of the segmented tumor is a determinant factor in mammogram diagnosis. It is very related to the malignancy of the tumor where a difference of just a few centimeters in the maximum diameter can determine whether is necessary to do a surgery or not. It can be a very difficult to detect contour of the tumor accurately depending on several factors, such as tumor shape, size, density, size, location and overall image quality [5].

The grow cut algorithm is an interactive supervised segmentation method by which users can obtain feasible results by selecting just a few points from inside and outside the region of interest. The grow cut is an interesting tool to segment objects of interest in medical images, because no other parameter are needed. The general purpose grow cut algorithm is based on cellular automata dynamics [4. These cellular automata are associated to pixels, which are previously labeled at the stage of selecting points inside and outside the region of interest It uses a bidimensional empirical mode decomposition method to segment mammography images achieving 83.96% of mean precision of segmenting masses of different classes of abnormality. The evolution of cellular automata depends on the selection of points internal and external to the object of interest, making possible to good segmentation of relatively difficult objects with complex borders in a process whose quality is highly dependent on user experience [7]. The higher the quality of points selection i.e the higher the user experience, the higher the quality of image segmentation results [8]. Hong proposes a Topographic Approach of segmentation, based on the fact that suspicious regions are usually brighter than neighbor regions, with uniform densities. However, for most of cases, regions of lesion do not have well-defined contours. Due to this fact, seed-based techniques, i.e. techniques in which users label initial seeds, achieve better quality in the final segmentation. GrowCut technique has been applied to successfully segment medical images, such as kidney and brain Cordeiro et al. apply the classical GrowCut to segment masses in mammograms, obtaining good results in terms of quality of segmentation. Zen et al. use a random-walk based segmentation, a method that also uses seeds provided by the user to achieve good segmentation results, but they do not provide quantitative analysis of the experimental results. Despite seed-based techniques have shown suitable performance for mass segmentation,they require a high level of specialist knowledge about the seed selection problem.

In this work, We propose an adaptive fuzzy semi-supervised version of grow-cut algorithm based on two basic modifications:

A) Automatic selection of internal points using the differential evolution optimizations algorithm. Maximizing the minimum distance between these points and the minimum gray level of the associated pixels, in order to minimize the need of human intervention.

B) Modification of cellular automata evolution rules by introducing Gaussian fuzzy membership function, in order to make the algorithm able to deal with complex and non-defined mammory lesion borders. Results were generated using MiniMIAS mammography image database.

2. STATE OF ART- RELATED WORK

The First demonstration of the use of multicellular neural network lesion localization in magnetic resonance mammography achieving good results. Nijad combines various enhancement methods and to segment breast region in order to obtain better visual interpretation analysis and classification of mammogram masses [7]. Oliver makes a review of state of art and shows that related works are divided into edge based ,region based and adaptive threshold methods. In edge based segmentation, it is difficult to determine the boundary of tumor due to some ill-defined lesion edges. Region based segmentation is considered more suitable for mass detection. Since region of tumor are usually brighter than their surrounding tissues, having almost uniform densities and fuzzy boundaries [8]. Lewis employ watershed to automatically segment tumor candidate regions, achiving an overall detection rate for mass tumors of 90%. However the used metric of analysis was based only on tumor location, not on the quality of segmentation Hong proposes a topographic approach of segmentation based on the fact that suspicious regions are usually brighter than neighbour regions with uniform densities [9]. For most of cases region of lesion do not have well defined contours. So for that seed based techniques in which users label initial seeds , achieve better quality in final segmentation.

Chakraborty et al. apply Multilevel Threshold [8] combined with region growing to perform segmentation for well-defined edge contours. However, both techniques present dif- ficulties at defining spiculated contours and ill-defined edges. Pereira et al. [7] uses an approach bases on Wavelet Analysis, achieving 79.2% of precision, for craniocauldal and mediolateral mammographic views. Berber et al. proposes a Breast Mass Contour Segmentation algorithm (BMCS), which uses an approach based on region growing, achieving 83.15 % of precision using the metric of Area Overlap Measure

Unsupervised and semi-supervised techniques try to reduce the required specialist knowledge on tumor region. Ramathi use Active contours to segment masses, achieving 86.85% of accuracy using an overlap measure between segment images and ground-truth. Pereira uses an approach bases on wavelet analysis, achieving 79.2% of precision, for craniocauldal and mediolaeral mammographic views.

3. PROPOSED SYSTEM

Our proposal is based on modification of the grow cut algorithm [7], a general purpose interactive segmentation algorithm able to segment objects with relatively complex borders by selecting adequate points internal and external to the region occupied by the object of interest.

3.1 A Simple Grow Cut

Growcut is based on cellular automata dynamics, represented by a grid of cells. Where each cell can assume a finite no of states varying according to neighbourhood pixels. The growcut technique uses the concept of seed pixels. Users initially label a set of pixels associated to different classes and taking into account the gray levels of these seeds , The algorithm tries to label all the pixels of the image. Each cell has a strength value and at each iteration the neighbours cells try to dominate this determined cell , changing its label. In case of defender cell has a higher strength than its dominators, then its label persist the same. Othewise this specific cell inherites the label of the dominant cell. This process continues until the algorithm reaches convergence i.e label remain same. The grow cut algorithm receives as input the image to be segmented and the label matrix with the labels provided by the user.The label provided by the user can be of two type: background label or object label , which corresponds to the region outside and inside of the lesion.When the user determines the object and background labels of chosen cell, the labelled cells start to interact with its neighbourhood. In grow cut as in the majority of seed based techniques, the quality of segmentation depends directly on the positon of the initial seeds. Therefore its depends on user knowledge to select appropriately seeds next to the edge of the object to be segmented [5].

The GrowCut technique uses the concept of seed pixels: users initially label a set of pixels associated to different classes and, taking into account the gray levels of these seeds, the algorithm tries to label all the pixels of the image. Each cell has a strength value and, at each iteration, the neighbor cells try to dominate this determined cell, changing its label. In case of a defender cell has a higher strength than its dominators, then its label persists the same. Otherwise, this specific cell inherits the label of the dominant cell. This process continues until the algorithm reaches convergence, i.e. label remain the same. . This process is performed with all cells of the grid. The Evolution of the cells is performed through iterations, which in each iteration all cells os the grid act as defender cells being attacked by its neighborhood. Within the same iteration this process is performed in parallel shows the evolution of a cells domain in the GrowCut technique, in different iterations of the evolution process.

3.2 A Classical Grow Cut

With our proposed modified growcut algorithm, we aim to reduce the need for initial specialist knowledge about the contour of the object of interest by reducing the effort of selection of seeds [6]. Morever the proposed algorithm aims to be fault tolerant, being able to recover from incorrect seed selection [7].

In classical Growcut, all the initial seeds selected by the user have maximum strength value, assigning high weights to seeds with incorrect labels. Unlike classical GrowCut , Our modified grow cut is based on the selection of seeds of only one class: to the object of interest [8]. Hence, we assign the maximum value to the cell of the center of mass, once we assume it has a higher chance of being correct labelled. The initialization is performed according to expressions :

$$¥p \,€\, P,\quad l_p = 0,\ \Theta_P=0,\ l_{cm}=l_{Obj} ,\ \Theta_{cm}=1$$

Where, p is a cell in space p of cells and l_p and are the labels and strengths of cell p, respectively . In our proposal, we also modified the update rule of Growcut cells in a way that the attack of each cell is based in a region modelled by a Gaussian fuzzy membership function. The Gaussian function is a probability density function of a normal distribution, which can be represented, in two dimesion based on parameters of average and standard deviation.

In the proposed approach, the initial localization of seeds is enough to estimate a fuzzy Gaussian function.

$$\Theta_{M,i} = \begin{cases} 1, \mu Bkg(i) > \mu Obj(i) \\ \Theta i, \mu Bkg(i) > \mu Obj(i) \end{cases} \quad (1)$$

$\mu_{Bkg}^{(i)} = 1- \mu_{Obj(i)},$ (2)

$$\mu_{Obj}^{(i)} = exp\left(-\left(\frac{Xi-Xm^2}{2\alpha_x s_x^2}\right)exp\left(-\left(\frac{Yi-Ym^2}{2\alpha_y s_y^2}\right)\right)\right) \quad (3)$$

where $\mu_{Bkg}^{(i)}$ is the fuzzy membership degree associated to the uncertainty of the ith cell belongs to image background, while $\mu_{Bkg}^{(i)}$ fuzzy membership degree associated to the uncertainty of ith belongs to the object of interest. These fuzzy membership functions are Gaussian functions whose variables xi and yi correspond to the coordinates of the ith cell in the grid, whereas xm and ym are the coordinates of the center of mass for the initially selected seeds; sx and sy are standard deviation of initial points, while ax and ay are weights of tunning of a Gaussian function.

The label of each qth cell, $l_{m,p,q}$ is updated according to the following expression of eq(4)

$$lm,p,q = \begin{cases} l_p, & \mu Bkg(q) > \mu Obj(q) \\ l_q, & \mu Bkg(q) \leq \mu Obj(q) \end{cases} \quad (4)$$

The main difference between the classical grow cut and our proposal is the consequent reduction of the effort to select seeds, once our approach require users to choose just the points internal to the object of interest, because background region is determined by the compliment of the Gaussian fuzzy membership function responsible to regulate the strength and label of each cell at the updating process. Our proposed modified evolution rule differs from original GrowCut because in growcut it depends only on strength of the attacking cell and difference of intensity values, while in a proposed method it also consider a fuzzy membership function to guide the update of cells, where regions with higher value of membership function related to background will have a different value of a strength compared to the lower ones.

3.3. Selection of Seeds Automatically

The selection of seeds consist of identifying initial pixels located in regions of tumor and non-tumor. In many seed based techniques such as Random Walks and graph cut Seeds are selected manually by a specialist. An important characteristic of the proposed algorithm is that it is not necessary to select non-tumor seeds, because the proposed algorithm can adjust its Gaussian fuzzy frontier based only on seeds of the tumor region.

In this work, we used the differential Evolution optimization algorithm to find automatically seeds in the region of the object of interest i.e the suspicious lesion.The solution is represented as a set of seeds, which explore the coordinates of image to find the best representation of initial seeds.Knowing that mass region frequently have higher intensity pixel values, the problem of finding an adequate set of seeds was converted into an optimisation problem.For this purpose, w used a multi objective fitness function, which evaluates both the distance between seed point and levels of intensity of the pixels related to them.

3.4 Adaptive Selection Of Parameters

The semi-supervised Growcut modification requires two parameters configurations: the no of points represented by a solution in the differential evolution algorithm and α value from fuzzy membership degree. However the best configuration parameters can vary according to the selected region of interest, as most of segmentation algorithms requires a specialist knowledge, which is dependent on professional experience and susceptible to human failure. Although the segmentation method is automatic, the selection of parameters still requires a configuration by a specialist.On the other hand, the process of selection of parameters can be automatized from the definition of an equation that maps the characteristics of the chosen region of interest and the best parameters of the modified GrowCut.

The Adaptive parameter equation consist of suggesting a configuration to the fuzzy Growcut algorithm based on the features of the ROI. To obtain the definition of adaptive parameter equation model, initially it was performed an exhaustive search in the fuzzy grow cut configuration space to find the best segmentation parameters for each ROI of the database.

3.5 Methodology

The proposed approach is called semi-supervised because it has a supervised and unsupervised step. Firstly, in the supervised step a specialist selects the region of interest, generating a sub-image as input for the segmentation method.

Afterwards, in the supervised step the haralick features are extracted and applied to Eq. 4 which will provide the algorithm parameters to the ROI. Next. The differential

Evolution optimization algorithm parameters algorithm automatically selects points internal to the probable mammory lesion. These seed points are located by an optimization process guided by a multi-objective fitness function to maximize the distance between points and the brightness of the pixels positions .Subsequently, our modified GrowCut algorithm performs the segmentation, generating a labeled image. Hence, once a region of interest , segmentation is performed automatically.

4. CONCLUSIONS

We proposed a new approach to segment masses in digital mammography images using an adaptive fuzzy modified version of the general purpose interactive grow-cut segmentation algorithm. Differently from the classical grow cut, in which user have to select points internal and external to the object of interest to be segmented, the intro of Gaussian fuzzy membership function allows users to select just internal points, trying to add our algorithm some robustness and fault-tolerance that do not exist in original approach. Also in order to minimize human intervention we used differential evolution algorithms to select automatic these internal points.

The main contribution of this consist of the modification of the evolution rule of grow cut cellular automata, introducing a Gaussian fuzzy membership function. The use of a Gaussian function combined with cellular automata evolution rule eliminates the need of selection of external seed points, which is not seen in state-of-art techniques, besides maintaining the quality of segmentation.

REFERENCES

[1] World Health Organization, WHO Position Paper on Mammography Screening, 2014.

[2] C.E. DeSantis, C.C. Lin, A.B. Mariotto, R.L. Siegel, K.D. Stein, J.L. Kramer, R. Alteri, A.S. Robbins, A. Jemal, Cancer treatment and survivorship statistics, 2014.

[3] M.P. Coleman, M. Quaresma, F. Berrino, J.M. Lutz, R. De Angelis, R. Capocaccia, P. Baili, B. Rachet, G. Gatta, T. Hakulinen, A. Micheli, M. Sant, H.K. Weir, J.M. Elwood, H. Tsukuma, S. Koifman, G.A. e Silva, S. Francisci, M. Santaquilani, A. Verdecchia, H.H. Storm, J.L. Young, Cancer survival in five continents: a worldwide population-based study (CONCORD), Lancet Oncol. 9 (8) (2008).

[4] S.K. Bandyopadhyay, I.K. Maitra, T.H. Kim, Identification of abnormal masses in digital mammography images, in: Proceedings – 2011 International Conference on Ubiquitous Computing and Multimedia Applications, UCMA 2011 (July 2015), 2012.

[5]] D.C. Allred, J.M. Harvey, M. Berardo, G.M. Clark, Prognostic and predictive factors in breast cancer by immunohistochemical analysis, Modern Pathol. 11 (2) (1998) 155–168.

[6] V. Raman, P. Sumari, H. Then, S.A.K. Al-Omari, Review on mammogram mass detection by machine learning techniques, Int. J. Comput. Electr. Eng. 3 (6) (2011).

[7] W.I. Suleiman, S.J. Lewis, D. Georgian-Smith, M.G. Evanoff, M.F. McEntee, Number of mammography cases read per year is a strong predictor of sensitivity, J. Med. Imaging 1 (1) (2014) 155–183, http://dx.doi.org/10.1117/1.JMI.1.1. 015503.

[8] D.X. Wang, F. Yuan, H. Sheng, An algorithm for medical imaging identification based on edge detection and seed filling, in: in: ICCASM 2010–2010 International Conference on ComputerApplication and System Modeling, Proceedings, vol. 15, 2010, pp. 547–548.

[9] S.Y.S.Ye, S.Z.S. Zheng,W.H.W. Hao,Medical image edge detectionmethod based on adaptive facet model, in: 2010 International Conference on Computer Application and System Modeling (ICCASM), vol. 3, 2010.

