Segmentation Methods of Mass Detection for Mammographic Images

Mr. Tushar P. Kele
PG student, Department of Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Aurangabad, India.

Mr. Pramod B. Bhalerao
Assistant Professor, Department of Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Aurangabad, India.

Abstract*—Breast cancer is the most common type of cancer in woman worldwide. Mammography image analysis is one of the most effective imaging technology for breast cancer diagnosis and this technology is based on texture and shape analysis of mammary lesion. The adaptive semi-supervised method of GrowCut algorithm is a segmentation algorithm where we modified automation evolution ruly by adding a Gaussian fuzzy membership function. Now manual selection of seed points of suspicious part of image is changed by semi-automatic stage, where we just select the internal points using differential evolution algorithms. Now we propose Grab Cut method which is used for separating object from background in image with given constraints. User is requested to mark a single rectangle around the object, using this we calculate the trimap. Using this trimap we can easily labelled the pixels i.e which label is background and which are foreground. Results we compare with the semi-supervised state of art of multiple methods and final result show that our method or algorithm achieves a better result for different lesions as compare to previous apporoaches.

Keywords— Breast Cancer, Fuzzy Grow Cut, Mass Detection, Mammography Image Analysis, Alpha Matting, Graph Cuts, Interactive Image Segmentation.

1. INTRODUCTION

Breast Cancer is found mostly in woman in world as per World Health Organisation (WHO). It is most common type of cancer in woman and cancer death happened in developed and under-developed countries [1]. As per report nearly 1.7 million new cases were analysis in 2012 year, and in that 12% were new cancer disease and 25% cancers were found in woman with the 522,000 deaths cases [2]. Breast Cancer are categories i.e 80% cases was come from high income countries and below 40% was come from low income countries [3]. For Breast Cancer early detection has an important factor to successful treatment of cancer. But once this stage go beyond then it is very hard to do the successful operation. The most effective method for breast cancer analysis is digital mammography but this mammography analysis is very hard. Even for specialist also to analysis such mammography images [4]. When the cancer is in beginning stage the breast imaging is essential so using this we can detect this breast as early as possible and then we can do tumor monitoring. The traditional techniques in image processing has been applied in medical field to make diagnosis less errors through accurate identification of anatomic anomalies. So tumor size is very important role to planning breast cancer treatment and also to avoiding multilple surgeries [5]. The Mammography image is very hard to analysis and diagnosis. So for this Mammography Computer Aided Diagnosis (MCAD) is useful for radiologist and other health professional to improve accuracy of their diagnosis [4].

The size of tumor is a important factor for diagnosis. Because this size shows it is related to malignancy of tumor or not, where just few centimetres in maximum diameter can determine whether it is needed to do a surgery or not. But it is very difficult to detect contour of the tumor. The contour of tumor accurately depending on several factors i.e tumor shape, density, size, location and overall image quality [5]. The challenges in tumor segmentation include low contrast images, intensity levels variation across multiple regions, poor image illumination, high noise, ill-defined contours and masses not obviously detected etc.

The Grow Cut algorithm is a supervised segmentation method. This algorithm by which users can obtain satisfied results by just selecting few points from inside and outside the region of interest. So this grow cut is a algorithm to segment the object of interest which are also use in medical images and for that we don't required any parameter. The grow cut segmentation method is based on cellular automata. These automata are associated to pixels which are label at the point of selection region and these automata are depends on selection of seeds. Using these selection make a possible good segmentation with difficult objects with hard borders. This process is highly dependent on user knowledge experience. If user has a higher experience then selection of seed is higher and then we get the good segmentation results [6].

The adaptive fuzzy semi-supervised method of grow cut is based on two modifications.

A) Automatic selection of internal points using different optimization algorithm i.e maximizing the minimum distance between these points and minimum grey level of associated pixels in order to minimize the need of human intervention.

B) Modification of automata evolution rule by introducing Gaussian fuzzy membership function. So this will help to algorithm to be able to deal with complex and non-defined borders.

The modified grow cut algorithms aim is used to reduce the need for initial specialist experience or knowledge. This knowledge is to select the proper seeds but this algorithm is fault tolerant so we can recover from incorrect seed selection of user.

This paper is organised as follows: in section 2 we present related state-of-art works: in section 3 we present a brief description of the Modified GrowCut algorithm, details of grab cut algorithm and method, a description of the image database MiniMIAS Database, in section 4 we show some experimental results we obtained in comparison with the state-of-art segementation method, in section 5 we discuss about the quality of the results. Finally in section 6 we performed general conclusion.

2. Related Work

Recent works have provided good accuracy in identifying the location of tumors (Liu et al., 2011) (Mohamed et al., 2009), however relatively little research has been done to verify the quality of segmentation. Oliver et al. (2010) makes a review of state of art and shows that related works are divided into edge-based segmentation, region-based segmentation and adaptive threshold.

In edge-based segmentation, it is difficult to determine the boundary of the tumor due to some ill-defined edges lesions. Region based segmentation are more suitable for mass detection, since regions of tumor are usually brighter than their surrounding tissue, have an almost uniform density and a fuzzy boundary (Raman et al., 2011).

Recent studies for tumor segmentation have been successfully applied to region-based techniques for tumor segmentation. Lewis and Dong (2012) uses Watershed to automatically segment tumor candidate regions, achieving an overall detection rate for mass tumors of 90%. However, the metric of analysis that was used and it is based on the location of the tumor and not on the quality of segmentation

In classical grow cut, all the initial seeds selected by the user which have maximum strength value. The modified grow cut is based on the selection of seeds of only one class i.e the object of interest [6]. We don't need background selection because from the seeds of object of interest we can estimate front region and from that we separating object and background. Therefore instead of assigning labels with strength, all the cells are initialized to zero strength except center of mass [7]. So we assign the maximum value to the cell of center of mass. The initialization is performed as below expression.

$$\forall_p \in P, l_p = 0, \theta_p = 0, l_{cm} = l_{obj}, \ \theta_{cm} = 1 \quad (2)$$

Where, p is cell in space p of cells and l_p *and* θ_p are the labels and strengths of cell p, respectively and label and strength of cell of center of mass of seeds are represented by θ_{cm} and θ_{cm}.

In this we also modified the update rule of grow cut i.e each cell is attacked based on region modeled by a Gaussian fuzzy membership function and it is a probability density function of a normal distribution which can represent in two dimensions which is based on these two parameters i.e average and standard deviation [8]. As we observed the region of tumor has a behavior similar to a Gaussian region. So we can say that a seed point model is not only based on the initial seed points but also on the membership function [8].

The region of tumor represents by a digital mammography image and it can also be represented by a Gaussian function [6]. It doesn't mean by contour of tumor is defined by expression and described by the function. So this function is represented the approximate where the tumor is located because the initial interest of the Gaussian function is in the region where tumor is located not in the definition of the edges [7]. So based on the Gaussian representation we can map by a membership function of the tumor presence [8].

The main difference between the classical grow cut and this modified grow cut is that the selection of seeds are minimize as compare to the grow cut once we choose the points inside the object of interest then region is determined by a complement of the Gaussian fuzzy membership function to update the process of next cell[6][7]. We determined by statistics of selected seed points and this algorithm is based on that our method which are fault tolerant and we see that our method is less dependent on user specialist [4].

The initial localization of seeds is enough to calculate fuzzy gaussian function. The strength of the model will be equal to 1 if the object cell to the background cell is higher. Otherwise, the strength of model is same as object cell [7].

This fuzzy membership function are Gaussian function whose variables xi and yi are coordinates of the ith cell in the grid, whereas xm and ym are the coordinates of the center of mass for starting selected seeds. The update evolution rule of grow cut algorithm. we propose pseudo code where $\Theta_{M,p}^t$ and $\Theta_{M,q}^t$ are model strength of cells p and q, respectively.

Algorithm 1. Modified Grow Cut Algorithm

```
procedure ModifiedGrowCut P, l
    l_cm = l_obj
    Θ_cm = 1

    for all p ϵ P do
        l_p^{t+1} ← l_p^t
        Θ_p^{t+1} ← Θ_p^t

        Calculate Θ_{M,p}^t

        for all q ϵ N(p) do

            Calculate Θ_{M,q}^t
        if g(||c_p⃗ − c_q⃗||_2).Θ_{M,q}^t > Θ_{M,p}^t then

            Calculate l_{M,p,q}^t
            l_p^{t+1} ← l_{M,p,q}^t
            Θ_p^{t+1} ← g(||c_p⃗ − c_q⃗||_2).Θ_{M,q}^t
        end if
      end for
    end for
        return l
end procedure
```

$$\Theta_{M,i} = \begin{Bmatrix} 1, & \mu_{Bkg}(i) > u_{obj}(i) \\ \Theta_i, & \mu_{Bkg}(i) \leq u_{obj}(i) \end{Bmatrix} \quad (1)$$

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

$$\mu_{Obj}(i) = \exp(-\frac{(x_i - x_m)^2}{2\alpha_x {s_x}^2}) \exp(-\frac{(y_i - y_m)^2}{2\alpha_y {s_y}^2}) \quad (3)$$

Where $\mu_{Bkg}(i)$ is the fuzzy membership degree which are associated to the uncertainity of the ith cell and this cell belongs to background image while $\mu_{Obj}(i)$ is fuzzy membership degree which are associated to the uncertainity of the ith cell which are belongs to the object of interest. s_x and s_y are the standard deviation of initial points , while α_x and α_y are the weights of tuning of the Gaussian function, and it is determined as per the problem of interest.

The label of each qth cell, $l_{M,p,q}$ is updated according to the below expression of eq.(5)

$$l_{M,p,q} = \begin{cases} l_p, & \mu_{Bkg}(q) > u_{obj}(q) \\ l_q, & \mu_{Bkg}(q) \le u_{obj}(q) \end{cases} \quad (4)$$

3. Proposed Method

3.1 Grab Cut

Grab cut is an iterative algortihm that combines statistics and graph cut in order to accomplish detailed 2D segmentation with very limited input. The work flows for the user goes as follows: The user selects a image which should be used to perform the segmentation on and draws a rectangle around the object of interest that should be segmented. That's all input that is needed by the grab cut algorithm to perform segmentation [9]. Afterward the user can do a final touch up of the image if needed or use the image again input by selecting a new rectangle.

3.2 Modelling foreground and background

The initial information given about the foreground and background are given by the user as a rectangular selection around the object of interest. Pixels outside this selection are treated as known background and the pixels inside are marked as unknown. From this data we want to create a model that we can use to determine if the unknown pixels are either foreground and background [9][10].

In the grab cut algorithm this is done by creating K components of multivariate Gaussian mixture Model (GMM) for the two regions. K components for the known background and K components for the region that could be foreground, giving a total 2K components [11].

3.3 Data Structure

Grab Cut requires four different bits of information for each pixel. Each image is stored in its own array. Each array is same size as that of the original image. The image is now obtained which consist of pixels z_n in RGB color space. So we use Gaussian Mixture Model (GMM) which are in practise already used for soft segmentation. Each GMM , one for foreground and one for background is taken to be a full covariance Gaussian mixture with K components. In order to deal with GMM traceably , in the optimization framework, an additional vector $K = \{k_1,,,,k_n,,,,k_N\}$ is introduced, with component index $k_n \in \{1,,,,,K\}$, assigning to each pixel, a unique GMM component , one component either from the background or the foreground the model, according to the $a_n = 0\ or 1^1$.

The gibbs energy for segmentation now becomes

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + v(\alpha, z) \quad (5)$$

Depending on the GMM component variables k. The data term is now defined, taking account of the color GMM model as

$$U(\alpha, k, \theta, z) = \sum D(a_n, k_n, \theta, z_n) \quad (6)$$

Where D(a_n, k_n, θ, z_n) = -log$(z_n|a_n, k_n, \theta)$ – log π(a_n, k_n) and p(.) is a gaussian probability distribution and π(.) are mixture weighting coefficients, so that therefore, the parameters of the model are now

$$\theta = \{\pi(\alpha, k), \mu(\alpha, k), \in (\alpha, k),\ \alpha = 0,1, k = 1 \ldots.. K\} \quad (7)$$

Where π represents the weighting components, μ represents the mean and ∈ represents the covariance of the 2k gaussian components for the background and foreground distributions.

The smoothness term is V is basically unchanged from the monochrome except that the contrast term is computed using Euclidean distance in color space.

$$V(\alpha, z) = \gamma \sum_{(m,n) \in C} [\alpha_n \neq \alpha_m] \exp - \beta ||z_m - z_n||^2 \quad (8)$$

The proposed methodology aim is to provide assistance to the specialist to find an accurate delineation of the mass. Therefore, it assumed that a region of interest was previously selected by a specialist and provided to the proposed system to perform a high quality segmentation. Therefore, the objective of the proposed method is not to segment the mass from a full mammogram, but to help the professional to identify the correct measure of the mass. Once the region on interest (ROI) is used as input, the segmentation task is performed automatically. The method is called semi-supervised because of the need of selection of the region of interest by a specialist. But once the ROI is given as input to the proposed system, the segmentation is performed automatically.

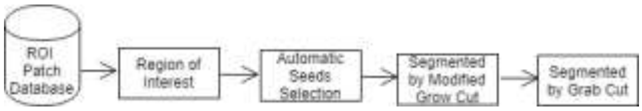


Fig 1. Flowchart of Proposed System.

3.4 The GrabCut segmentation algorithm:

This section describes the GrabCut hard segmentation algorithm which is iterative image segmentation in GrabCut Colour data modelling [1].

Algorithm:

- Initialize trimap T by supplying only T_B, the foreground is set to the $T_F = \Phi$; $T_U = T_B$, complement of the background.
- Initialize $\alpha_n = 0$ for $n \in T_B$ and $\alpha_n = 1$ for $n \in T_U$
- Background and foreground GMMs initialized from sets $\alpha_n = 0$ and $\alpha_n = 1$ respectively.

Iterative Minimisation Algorithm

1. Assign GMM components to pixels: for each n in T_U,
$$k_n = arg \min_{k_n} D_n(\alpha_n, k_n, \theta, z_n).$$
2. Learn GMM parameter from data Z:
$$\theta = arg \min_{\theta} U(\alpha, k, \theta, z).$$
3. Estimate segmentation : use min cut to solve
$$\min_{\{\alpha_n; n \in T_U\}} \min_{k} E(\alpha, k, \theta, z)$$
4. Repeat step 1 until convergence.

4. Results

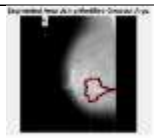
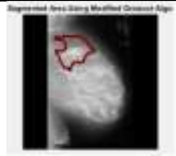
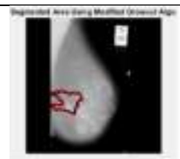
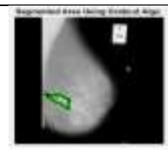
Input Image	Modified Grow Cut	Grab Cut
Mdb 13		
Mdb 30		
Mdb 32		
Mdb 58		

Fig 2. Segmentation with Modified Grow Cut and Grab Cut

Figure 2 shows the results of proposed algorithm and base algorithm. Each image in this figure shows that the bounding rectangle alone is sufficient for user interaction to enable foreground extraction to be completed automatically by GrabCut. Different examples are shown after simplification of the problem which is increasingly more problematic in terms of substantial amount of user interfaces.

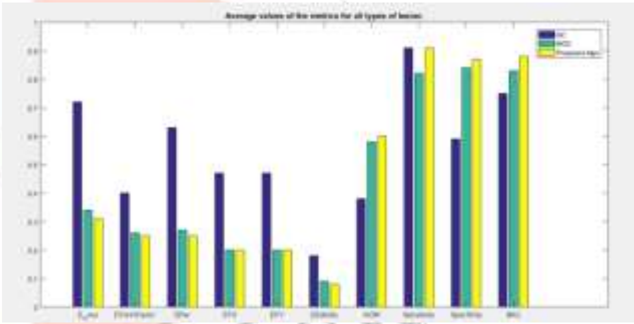


Fig.3 Average Values of the all the types of lesion.

From Fig 2 we can perceive that the proposed technique achieved the lowest average error, with lower median and lower range, for most of the metrics based on errors. Regarding AOM, Sensitivity, Specificity, and BAC, the proposed segmentation method reached a position among the techniques with best performances, achieving the best values for AOM and Specificity.

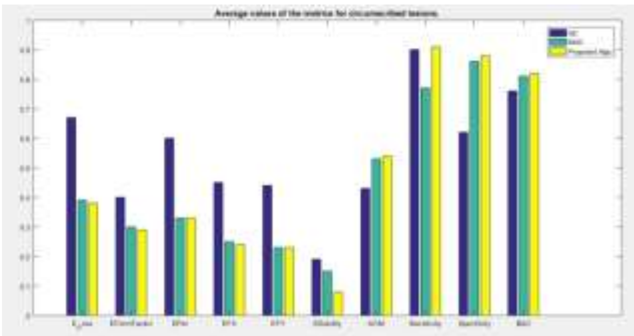


Fig 4. Average Values for Circumscribed lesions.

From the Fig3 we can notice that the proposed segmentation method achieved results with small standard deviations as well, for most metrics. Avarage results of metrics for circumscribed lesion image are shown in fig3.

From Fig 4, we can perceive that the proposed method assumed the lowest average error values, with the lowest standard deviations as well. For metrics AOM and BAC, the proposed technique was equivalent to the other state of the art approaches. Average results of metrics for indistinct type of masses are shown in fig 4.

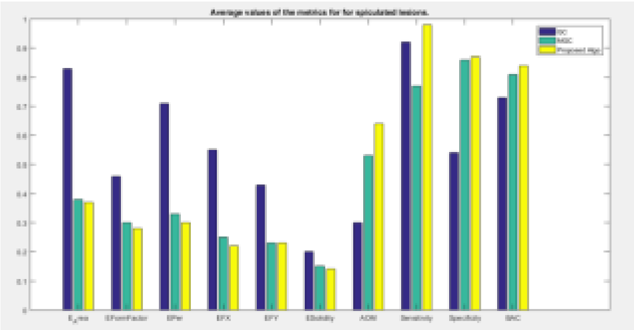


Fig 5. Average Values for spiculated lesions.

5. Comparison Between Modified Grow Cut and Grab Cut

Characteristic	Modified GrowCut	Proposed algorithm
Selection of seeds	Selection of seed only of object class.	Selection of seed only on one class.
Initialization	Only center of mass strength value equal to 1 and its showed by $\Theta_{cm} = 1$	All the cells are initialized with 0 except center of mass.
Segmentation	Based on knowledge and Gaussian function that separated bkg and obj region.	Based on user draws a rectangle on image to know bkg and obj region.
Evolution rule	Based on difference of intensity values and fuzzy membership function and it is represented by $\boldsymbol{\Theta_p^{t+1}} \leftarrow \boldsymbol{g}\left(\left\|\vec{c}_p - \vec{c}_q\right\|_2\right) . \boldsymbol{\Theta_{M,q}^t}$	Based on energy function. $\boldsymbol{E(\alpha,k,\theta,z)} = \boldsymbol{U(\alpha,k,\theta,z)} + \boldsymbol{v(\alpha,z)}$
Fault tolerance to seeds localization	High	High

Table1. Comparison between Modified Grow Cut and Grab Cut

6. Conclusion

This paper emphasis on modification of modified grow cut algorithm which is an algorithm that uses Guassian membership function and selected seeds. So based on these seeds the center of mass is identified but in proposed system we discard these selection of seeds where we just draw a rectangle on image object. So using this we got the foreground object and based on this foreground we can easily find out background object. To easily find out seed on images we use Gaussian Mixture Model (GMM) to get the exact seed on particular images. So as we see the fig 3, 4, and 5 we can see that our proposed system will gives better result as compare to previous algorithm.

7. References

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

[2] A. Oliver, J. Freixenet, J. Martí, E. Pérez, J. Pont, E.R.E. Denton, R. Zwiggelaar, A review of automatic mass detection and segmentation in mammographic images, Med. Image Anal. 14 (2) (2010) 87–110.

[3] S. Lewis,A. Dong, Detectionof breasttumor candidatesusingmarker-controlled watershed segmentation and morphological analysis, in: in: 2012 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), 2012.

[4] F.R. Cordeiro, W.P. Santos, A.G. Silva-Filhoa, Segmentation of mammography by applying growcut for mass detection, in: in: Studies in Health Technology and Informatics, vol. 192, 2013.

[5] D.C. Pereira, R.P. Ramos, M.Z. do Nascimento, Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm, Comput. Methods Programs Biomed. 114 (1) (2014)

[6] P. Sarkar, A brief history of cellular automata, ACM Comput. Surv. (CSUR) 32 (1) (2000).

[7] S. Vicente, V. Kolmogorov, C. Rother, Graph cut based image segmentation with connectivity priors, in: in: 26th

IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2008.

[8] B. kishore, R. Vijaya, R. Saha, S. Selvan, Article: using Haralick features for the distance measure classification of digital mammograms, IJCA Proceedings on International Conference on Advanced Computing and Communication Techniques for High Performance Applications ICACCTHPA 2014.

[9] Boykov, Y. And Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images.

[10] Dempster, A. Laird, M. And Rubin D. Maximum likelihood from imcomplete data via EM algorithm.

[11] Kolmogorov V. And Zabih R. What energy functions can be minimized via Graph Cuts.

[12] Mortensen E. And Barrett, W. Intelligent Scissors for image composition.