

Breast Density Segmentation based on Fusion of Super Pixels and Watershed Transform

Abdulrasaq Surajudeen
Center for Innovative Technology
Kwara State University, Nigeria

Zwiggelaar Reyer
Department of Computer Science
Aberystwyth SY23 3DB, UK

ABSTRACT

Breast density, defined as the proportion of fibroglandular tissue over the entire breast has been linked with a higher risk of developing breast cancer, in fact it has been suggested that women with a mammographic breast density higher than 75 percent have a four-to six-fold higher risk of developing breast cancer than women with little or no dense tissue. Therefore, automatic methods of measuring breast density could potentially aid clinicians to provide more precise breast cancer risk estimates. This paper proposes a novel method of segmenting breast density, which extracts objects with the same density using fusion of super pixels and a watershed based technique, this idea is based on the principle that both super pixel and watershed often results in over segmentation, for the later algorithm, over segmentation may be due to contours which have been suppressed according to similarity of contrast and topological measures, we took advantage of super pixel to consolidate space information and efficiently process the intensity non-homogeneity problem, afterward, re-introduced this contour with watershed transform to get a better segmentation.

General Terms

Segmentation

Keywords

Watershed, Super pixel, Mammograms, Segmentation

1. INTRODUCTION

Breast density is one of the strongest autonomous predictors of breast cancer risk [1], the denser the breast the higher the risk of developing breast cancer [2]. In fact it has been shown that women with a mammographic breast density higher than 75 percent have a four- to six-fold higher risk of developing breast cancer than women with little or no dense tissue [4]. Therefore, automatic methods of measuring breast density could potentially help clinicians to provide more accurate breast cancer risk estimates. This paper proposes a novel method of segmenting breast density, which extracts objects with the same density using fusion of super pixels and a watershed based technique. While so many approaches for mammography density segmentation exists [3], few have used super pixels towards mammography image segmentation. Although it is difficult to establish what constitute a good super pixels algo-

rithm; adherence to image boundary, speed, and efficient usage of memory are some of the desirable property of a good super pixels algorithm. One of the anticipated characteristics of super pixels is over segmentation which allows most important boundaries in the image to be found. Pei et al. [5] proposed a snake based model which is optimized by super pixels to segment the pectoral muscle. They use super pixel segmentation to pre-segment the original image into homogeneous regions with precise boundaries and then define the initial contour of the snake model from the super pixel edges and optimize it by using super pixel classification results. They were able to segment pectoral muscles of different sizes, shapes and intensities. Ji et al. [6] used a multistage segmentation method based on super pixels and fuzzy clustering (MSFCM) to segment MRI brain images, and their approach segment the image by merging the super pixels with the same classification label together. Massich et al. [7] proposed an energy based strategy to segment breast ultrasound (BUS) images based on discrete optimizations using super pixels and a set of novel features analogous to the elements encoded by the US BI-RADS lexicon, they map the image into a discrete set of elements and then formulate the segmentation in a metric labeling problem. Subsequently they re-map the labeling and obtain the final segmentation. Mahapatra et al. [8] proposed an automated method for segmenting cardiac right ventricles (RV) from magnetic resonance (MR) images, first the image was over segmented using super pixels, then each pixel was analyzed for the presence of the RV region using a random classifier. Any super pixels that was found to contain the RV became part of the region of interest (ROI) and this approach employed the use of semantic knowledge and context information. Prabha et al. [9] proposed a lazy random walk (LRW) approach to identify similar regions in an image and group these together based on the intensity of the image pixels.

2. MATERIAL AND METHODS

All 322 images obtained from the mini MIAS database [17] are used as experimental data. The spatial resolution of each image is fixed to 1024×1024 pixels. The database contain 208 and 114 normal and abnormal images. Among the abnormal cases, 63 are benign and 51 malign, image segmentation based on the fusion of super pixels and the watershed transform is proposed taken into account the fact that both super pixel and watershed often results in over segmentation, for the later algorithm over segmentation may be due to contours which have been suppressed according to similarity of contrast and topological measures, super pixel was used

to consolidate space information and efficiently process the intensity in homogeneity problem, afterward, re-introduce this contour with the watershed transform to get an improved segmentation. This method can be divided to four stages, which are detailed in subsequent subsections and an overview is shown in Fig 1.

2.1 Pre-processing Technique

The mammograms was pre-process to remove labels and identify the breast area, first the image geometry was exploited in order to detect the edges and find the contours. If the contour has four sides (since most mammography label and some artefacts are rectangular in shape) bitwise operation was employed to suppress it then the visual appearance of the image was improved using histogram equalization [16].

2.2 Rough segmentation

The primary aim of this step is to explore the semantic meaning of the images by locally grouping the pixels together as super pixel, and use this as a pre-processing step for the watershed segmentation algorithm.

Given the distance measure between a pixel i and the super pixel cluster center C_k (see sec 4, for details) as D , we can combine the intensity distance and spatial distance in to a single measure which is imperative to normalize color proximity and spatial proximity by their respective maximum distances within a cluster N_s and N_c , we can write D' as .

$$d_c = \sqrt{(L_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (1)$$

$$d_s = \sqrt{(X_j - x_i)^2 + (Y_j - y_i)^2} \quad (2)$$

$$D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \quad (3)$$

Given a sampling interval $N_s = \sqrt{\frac{N}{K}}$, (where S is the grid interval of super-pixels center, N is the total number of pixels in the image and K is a user-define pixel value) therefore, the maximum spatial distance expected within a given cluster should correspond to this sampling interval, determining the maximum color distance N_c is not so straightforward, as color distances can vary significantly from cluster to cluster and image to image. We can avoid this by fixing N_c to a constant m so that (3) becomes

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \quad (4)$$

Taken the square of both sides in (4) and multiplying both sides by m^2 .

$$D'(m)^2 = \sqrt{(d_c)^2 + \left(\frac{d_s}{N_s}\right)^2 m^2} \quad (5)$$

letting $D'm = D$, and taking the square root of both sides simplifies to the distance measure we use in practice.

$$D = \sqrt{(d_c)^2 + \left(\frac{d_s}{N_s}\right)^2 m^2} \quad (6)$$

we can adapt (6) for a greyscale image by setting $d_c = \sqrt{(L_j - L_i)^2}$ then we have

$$D = \sqrt{(L_j - L_i)^2 + \left(\frac{d_s}{N_s}\right)^2 m^2} \quad (7)$$

2.3 Improved segmentation

The aim of the watershed transform is to search for regions of high intensity gradients (watersheds) that divide neighboring local minima (basins) [13]. Prior to applying this algorithm the image filtered using a gaussian filter ($size = 5, 5, \sigma = 0$) to aid the thresholding step. then the Euclidian distance was computed and watershed transform was applied. In order to improve the final results, markers generated from the initial threshold was then used.

2.4 Label and extract object

In order to obtain the final results, it is essential to label the pixels. Every pixel value has a unique label value. Pixels with the same label value belong to the same object where each label corresponds to a unique object in the image. From there, all we need to do is loop over each of the labels individually and extract each object.

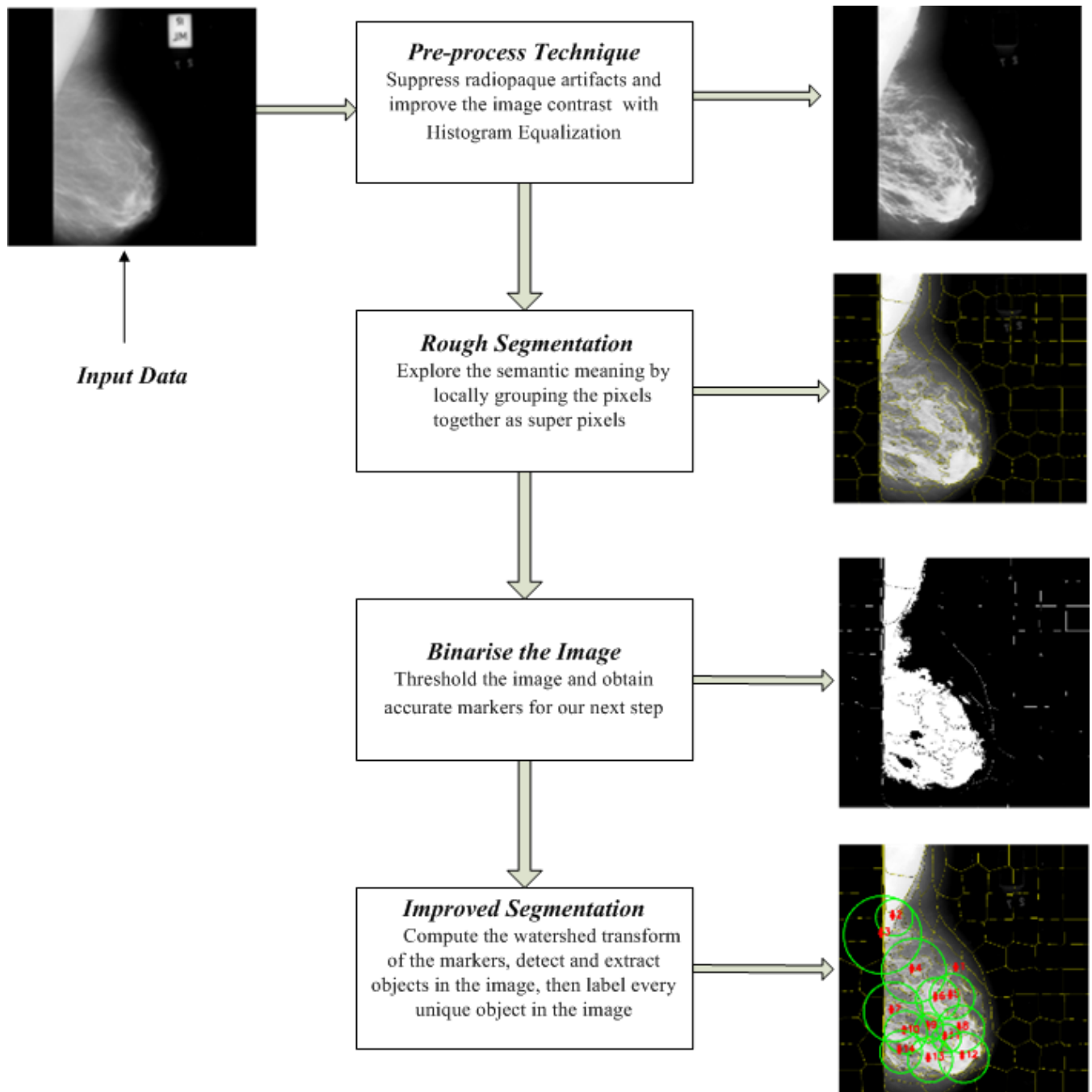


Fig. 1: Shows the block diagram for the propose Algorithm.

3. EXPERIMENTAL RESULTS

To verify the effectiveness of the algorithm, all 322 images obtained from the mini MIAS database [17] are used as experimental data. Figure 2 (a) shows the original image (mdb2651m) (b) after pre-processing, (c) rough segmentation of the image into super-pixels, (d) final image output after improved segmentation, the color code indicate differences in the tissue distribution while circles placed on

them gives us the idea on how tissues are distributed around the image. Figure 5: shows different mammography image of all the four BIRADS categories after applying the proposed approach, the first column shows the original images, the second column show the pre-process image, the third column after rough segmentation, fourth column shows the image after applying the proposed method, fifth column indicate the BIRADS class while the last column shows the average tissues composition.

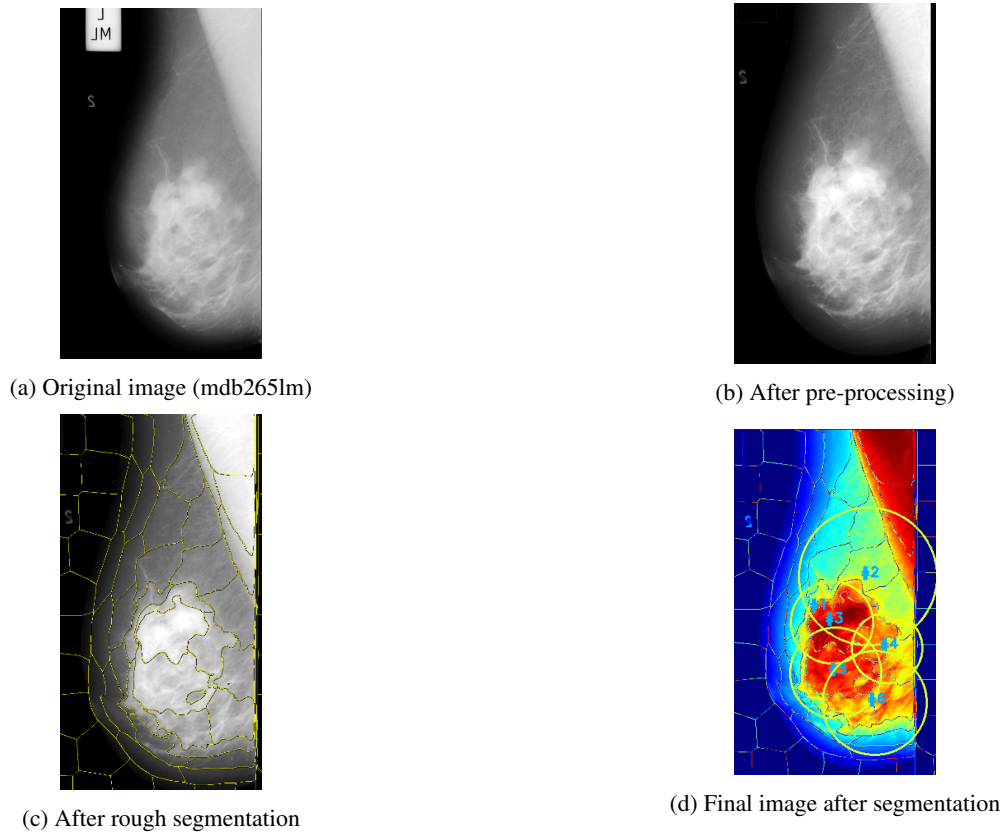


Fig. 2: shows mdb2651m indicating various steps in the segmentation process. (a) showing the original image, (b) showing the image after pre-processing, (c) after segmenting the image in to super-pixels, (d) final image output after improved segmentation, color coding shows the homogeneous region (red) the nodular (yellow), linear (green) and radiolucent (blue). The color coding indicate differences in the tissue distribution while circles placed on them gives us an idea on how tissues are distributed around the image

3.1 Comparison of results with moments based method

We compared the proposed approach using the same set of images. The experiment shows that the new method has an accuracy and robustness compared to moments based approach [18]. Figure 3: (left) (middle) and (right) shows the original image (mdb1051l) and the results of segmented mammographic image using both the geometric moments algorithm [18] and the proposed approach respectively, note the misclassification between nodular or homogeneous tissue and radiolucent tissue in (middle). Also figure 4: (left) (middle) and (right) shows the original image (mdb184rl) and the results of segmented mammographic image using both the geometric moments algorithm and the proposed approach respectively, also note the misclassification between nodular and homogeneous tissue in (middle).

4. THEORY

Super pixels are increasingly becoming popular in the field of computer vision since Ren and Malik [10] coined the term Super pixels in their 2003 paper, super pixels are group of pixels that are closely related to each other, the relationship may be due to color similarity or spatial proximity, the main goal of this algorithm is to bundle pixels together to form atomic regions with meaningful interpretation. Most super pixels algorithm results [8] in over

segmentation, which allow the most important boundary of the image to be extracted. This might be seen as a disadvantage because many insignificant boundaries are generated in the process. However the trade off to this disadvantage is the fact that no boundaries are missed. Super pixel are widely covered in the literature and are broadly categorized in either graph based or gradient-ascent method. Felzenszwalb and Huttenlocher [11] proposed a method that clustered pixels as nodes in a graph, this approach adheres well to image boundaries but produces irregular shapes. Levinshtein et al. [12] used level set geometric flow to dilate a set of seed location. Simple linear iterative clustering (SLIC) is a gradient-ascent method that adapted the k-means approach; it used color similarity of pixels and spatial information within images to generate compact and uniform super pixels. SLIC is very efficient and showed good adherence to image boundaries. In this paper, Achanta et al. [?] simple linear iterative clustering (SLIC) approach is used as a pre-segmentation technique.

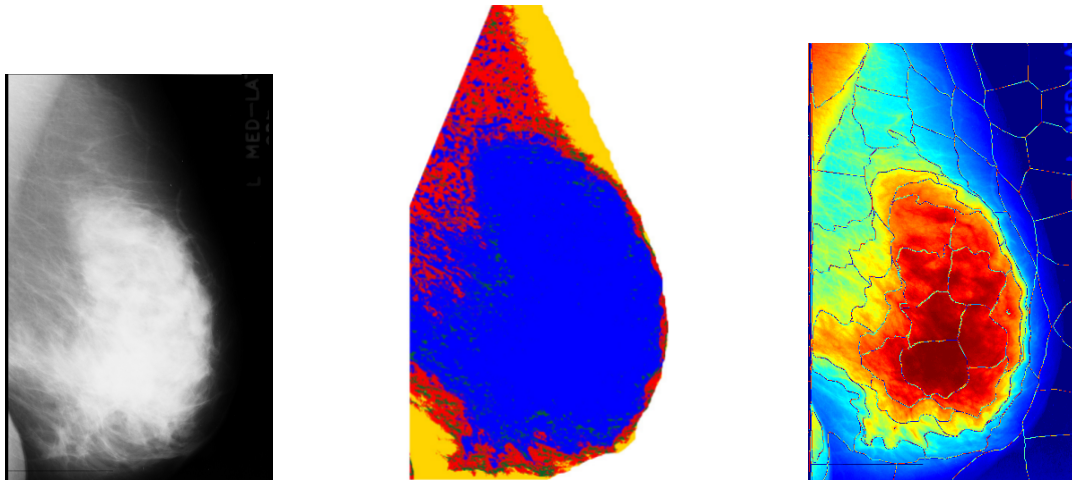


Fig. 3: Shows the original mamographic image mdb1051l (left) and the segmentation produced using the geometric moments algorithm [18] (middle) and the proposed approach (right), circles are removed for the sake of clarity in comparison. The tissue specific areas in the middle image (i.e geometric moments method) are: nodular (red), linear (green), homogeneous (blue) and radiolucent (yellow). The color coding for the proposed approach (right) is the same as used in Fig. 2. Note the misclassification between nodular or homogenous tissue and radiolucent tissue in the geometric moments (middle) method.

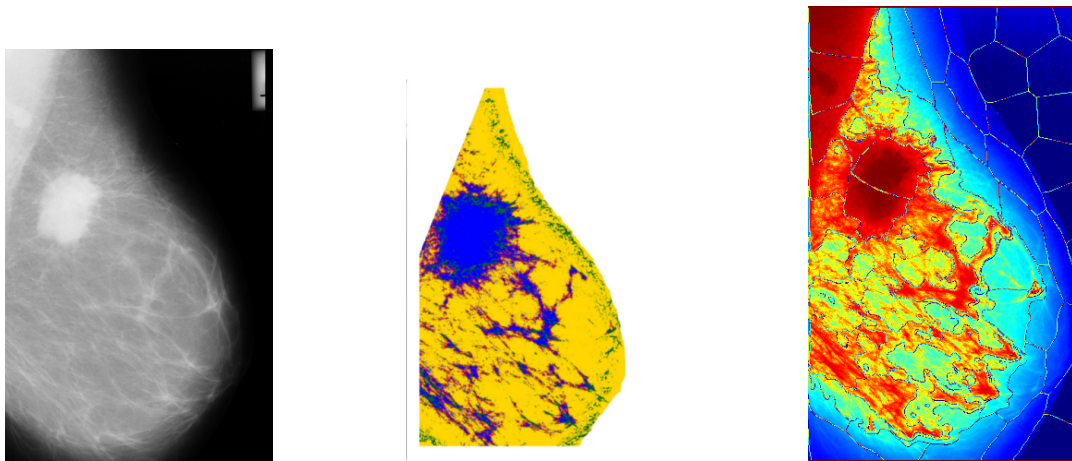


Fig. 4: Shows the original mamographic image mdb184rl (left) and the segmentation produced using the geometric moments algorithm [18] (middle) and the proposed approach (right), circles are removed for the sake of clarity in comparison. The tissue specific areas in the middle image (i.e geometric moments method) are: nodular (red), linear (green), homogeneous (blue) and radiolucent (yellow). The color coding for the proposed approach (right) is the same as used in Fig. 2. Note the misclassification between nodular and homogenous tissue in the geometric moments (middle) method.

4.1 Watershed Transform.

The watershed algorithm was originally proposed by [15] and later improved by [13]. The basic principle of watershed is as follow. Given an image f as a topographic surface and define the catchment basins of f and the watershed lines by means of a flooding process. Visualize that we pierce each minimum of the topographic surface, and that we drop this surface into a lake of water at a constant speed. The water entering through the holes floods the surface during the flooding, two or more floods coming from different minima may merge. We can avoid this event and we build a dam on the points of the surface where the floods would merge. At the end of

the process, only the dams emerge partitioning the landscape into regions or basins separated by dams, called watershed lines.

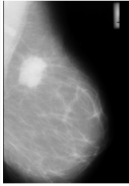
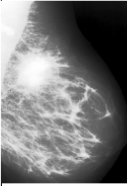
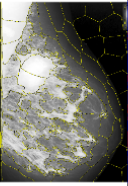
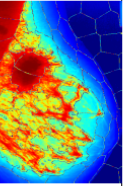
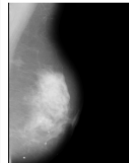
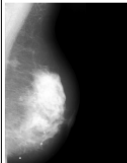
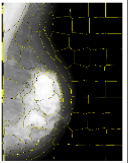
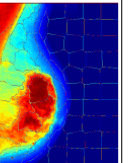
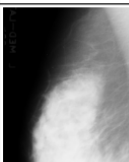
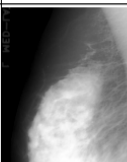
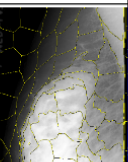
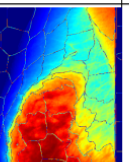
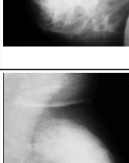
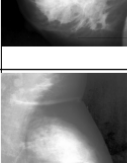
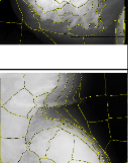
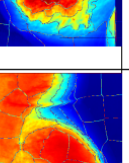
Original Image	Pre-process Image	Rough Segmentation	Improved Segmentation	Birads Class	Average Tissues Composition
				I	[8.3%,9.0%,17.7%,65%]
				II	[12.2%,9.1%,30.1%,48.6%]
				III	[19.3%,9.8%,53.3%,17.6%]
				IV	[9.2%,5.3%,83.4%,2.1%]

Fig. 5: proposed approach applied to different mammography image of all the four BIRADS categories, the first column shows the original images,the second column show the pre-process image, the third column after rough segmentation, fourth column shows the image after applying the proposed method, fifth column indicate the BIRADS class while the last column shows the average tissues composition. Color coding is the same as used in Fig. 2.

5. CONCLUSION

In this paper, a novel approach based on the fusion of super pixel and watershed technique is proposed. First we grouped the pixels together to form super-pixels; then detect and extract objects in the image using a marker base watershed transform. We explored the advantages of super-pixel to strengthening space information and successfully solved the intensity in-homogeneity problem.

We compared the proposed approach with segmentation obtained with moments based approach using the same set of image as experimental data and the results also show good segmentation accuracy, however, we note that the pectoral muscle is often misclassified as nodular or homogeneous tissue, which indicates that eliminating

the pectoral muscles from the mammogram is an important factor in density estimation so that breast analysis can be confined to the breast region alone.

6. REFERENCES

- [1] V.A. McCormack, I. dos Santos Silva. *Breast density and parenchyma patterns as markers of breast cancer risk: a meta-analysis*. Cancer Epidemiol Biomarkers Prev. 15: 1159-69, 2006
- [2] J.N. Wolfe. *Risk for breast cancer development determined by mammographic parenchymal pattern*. Cancer, 37:24862492, 1976.

- [3] W.He, A.Juette, E.R.E.Denton, A.Oliver, R.Mart, R.Zwiggelaar. *A review on automatic mammographic density and parenchymal segmentation*. International Journal of Breast Cancer 2015, Article ID 276217, <http://dx.doi.org/10.1155/2015/276217>, 2016.
- [4] N.F. Boyd, H. Guo, L.J. Martin, L.M. Sun, J. Stone, E.R. Fishell, R.A. Jong, G. Hislop, A. Chiarelli, S. Minkin, J.A. Yaffe. *Mammographic density and the risk and detection of breast cancer*. New England Journal of Medicine. 2007 356:227-236.
- [5] L.Pei, W.N Li. *Pectoral muscle segmentation in mammograms based on snake model optimized by super pixel*. Guangdianzi Jiguang Journal of Optoelectronics 26(8):1633-1638.
- [6] S.Ji, B.Wei, Z.Yu, G.Yang, Y.Yin. *A new multistage medical segmentation method based on super pixel and fuzzy clustering*. Computational and Mathematical Methods in Medicine 2014, Article ID 747549.
- [7] J. Massich, G. Lemaitre, J.Marti, F.Meriaudeau. *Breast Ultra-Sound image segmentation an optimization approach based on super-pixels and high-level descriptors*. International Conference on quality control and artificial vision (QCAV) 2015, 2015, LeCreusot, France.
- [8] J.Mahapatra. *Cardiac image segmentation from Cine Cardiac MRI Using Graph Cuts and Shape Priors* Society for Imaging Informatics in Medicine 2013. 227(6): 794804. Published online 2014 June doi:10.1007/s10278-014-9705-0, 2014.
- [9] R. Prabha, C. Kohila. *Lazy random walks for super pixel segmentation*. Karpagam Journal of Engineering Research (KJER) Volume No: II, Special Issue on IEEE Sponsored International Conference on Intelligent Systems and Control (ISCO15).
- [10] X. Ren, J. Malik. *Learning a classification model for segmentation*. Proceedings of the International Conference on Computer Vision, pages 10 17, 2003.
- [11] P.F.Felzenszwalb, D.P.Huttenlocher. *Efficient graph-based image segmentation*. D.P. International Journal of Computer Vision (2004) 59:167. doi:10.1023/B:VISI.0000022288.19776.77.
- [12] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, K.Siddiqi. *Turbopixels: fast super pixels using geometric flows*. IEEE Trans. Pattern Analysis and Machine Intelligence, 31,(12), pp. 2290-2297, 2009.
- [13] S.Beucher. *The Watershed Transform Applied to Image Segmentation*. Proceedings of the Pfefferkorn Conference on Signal and Image Processing in Microscopy and Microanalysis, pp. 299314, 1991.
- [14] R. Achanta, A.Shaji, K. Smith, A. Lucchi, P. Fua *super pixels compared to state-of-the-art super pixel methods*.IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (11), pp: 2274-2282 2012.
- [15] H. Digabel, C. Lantuejoul. *Iterative Algorithms*. Proceedings of the 2nd European Symposium Quantitative Analysis of Microstructures in Material Science, Biology and Medicine, 85-89, 1998.
- [16] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*.Third Edition, 2008.
- [17] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Sta-matakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, J. Savage *The mammographic images analysis society digital mammogram database*.in: Dance, Gale, Astley, Gairns (Eds.), Excerpta Medica. International Congress Series, Vol. 1069, Elsevier, 1994, pp. 375378
- [18] W.E.He, K.Stafford, E.R.E.Denton, R.Zwiggelaar. *Mammographic image segmentation and risk classification based on mammographic parenchymal patterns and geometric moments*. Biomedical Signal Processing and Control 6 (2011) 321 329.