

Automated Mammogram Segmentation Using Seed Point Identification and Modified Region Growing Algorithm

ABSTRACT

Segmentation is one of the prominent and crucial steps in any image processing applications. Segmentation subdivides the image into its constituent regions or objects. In this paper we propose a novel automatic segmentation algorithm for extracting portion of breast which contains tumor or abnormalities. The proposed algorithm consists of three different stages. In the initial stage, an automatic seed point identification method is used for locating the center pixel of the abnormal regions in the mammogram images. In the next stage, region of interest around the seed point is extracted using the modified version of region growing algorithm for aggregating pixels around the seed point. Finally, gradient operators are used for identifying boundaries of the segmented region. Using these boundaries, segmented region of the mammogram images are cropped and treated as ROIs that may constitute the tumor/abnormal regions. The segmented ROIs are well in agreement with the abnormality portions that are already identified and labeled by the Radiologists. Average time taken for extracting ROI of one mammogram image is 3.7393 seconds.

Keywords: mammogram, ROI, region growing, seed point

1.INTRODUCTION

Segmentation is one of the necessary tasks for any image analysis application for interpreting the objects or content in the image. Segmentation subdivides the image into its constituent regions or objects. The pixels in each partitioned region possess identical properties or attributes. These sets of properties of the image may include gray levels, contrast, spectral values, or textural properties. The segmentation is an iterative process and it stops when the objects of interest in an application have been isolated. The results of the segmentation is the number of homogenous regions each having unique label. An image is thus defined by a set of regions that are connected and non-overlapping, so that each pixel in the image acquires a unique region label that indicates the region to which the image belongs. The set of objects of interest in an image, which are segmented, undergoes subsequent processing such as object classification and scene description. Segmentation accuracy determines the eventual success or failure of any computerized analysis procedures. [1] [2] [3][4]

Segmentation has an important role in medical image processing. It includes detection of the coronary border in angiograms, multiple sclerosis lesion quantification, surgery simulations, surgical planning, measuring tumor volume and its response to therapy, functional mapping, automated classification of blood cells, studying brain development, detection of micro calcifications on mammograms, image registration, atlas-matching, heart image extraction from cardiac cineangiograms, detection of tumors etc.[5][1][6] [7] .

In medical imaging, segmentation is important for feature extraction, image measurements and image display. In some applications, it may be useful to classify image pixels into anatomical regions, such as bones, muscles, and blood vessels, while in others into pathological regions, such as cancer, tissue deformities, and multiple sclerosis lesions. In some studies the goal is to divide the entire image into sub regions such as the white matter, gray matter, and cerebrospinal fluid spaces of the brain [8], while in others one specific structure has to be extracted, for example breast tumors from magnetic resonance images [9].

A wide variety of segmentation techniques have already been proposed. However, there is no single standard segmentation technique that can produce satisfactory results for all imaging applications. The definition of the goal of segmentation varies according to the goal of the study and the type of the image data. Different assumptions about the nature of the analyzed images lead to the use of different algorithms [6] [8][9].

2.LITERATURE REVIEW

Segmentation of mass regions from the mammogram images proposed by Kom et.al [10] locally spots out masses which are denser than surrounding tissues using adaptive thresholding method. The results obtained by this method shows that 95.91% of sensitivity for mass detection. The receiver operating characteristic (ROC) analysis shows an area of 0.946 with enhancement of the mammogram and 0.938 without enhancement.

Another method for finding the breast edge using area enclosed by iso-intensity contours was proposed by Padayachee et.al [11] which improves the traditional thresholding methods for segmentation by incorporating spatial information into the segmentation. The results were evaluated by comparison to breast borders drawn by different radiologists. Results were generally good for those images which contain clear breast edges. Even though threshold based segmentation algorithms works well on clear breast edge but it do not provide information regarding the pattern or similarity measures of pixels in the images.

Dubey et al. [12] made a comparative study on mammogram segmentation using two different semi-automated methods viz level set and marker controlled watershed, which performs an accurate and fast segmentation of tumors in the mammograms. The robustness of the proposed method is demonstrated by considering a set of 17 mammogram images for the segmentation. These two methods seem to work well on the mammograms as seen from the exact boundary of the abnormal growth or lesions by demonstrating their comparative edge over other methods. Out of these two methods the marker controlled watershed segmentation shows better results than the level set approach.

A high percentage of digital mammograms have large proportion of pixels with no diagnostic information. An algorithm developed by Lou et al [13] automatically identifies the orientation of breast region as well as extract the breast region from mammogram images. Breast regions extracted from the digital mammograms reduce file sizes by three to five folds. During the extraction process, important parameters can be obtained such as breast height, width, and orientation. These parameters are particularly useful for correctly identifying the breast region from the original image. Furthermore, the extracted region contains only breast pixels that are useful and important in the process of image analysis to determine the gray value ranges for various breast tissue types.

An automatic image segmentation method based on improved watershed transform using prior information was proposed by Wei-Yen Hsu [14] to separate original mammogram image into the breast with tumor, the breast without tumor, and background. By doing so it reduces the volume of the data for processing and efficiently increases the performance of the system. In this algorithm breast regions are distinguished from background using canny edge detector. The main drawback of the watershed segmentation is that it produces over-segmented results. The watershed segmentation obtains catchment basin from the gradient of the image and results too many small regions. Moreover it is sensitive to noise and local variations in the image have a strong influence in the result.

A mammographic mass segmentation algorithm proposed by Wang et al [15] is used for extracting masses from mammogram images which is the most challenging task in mammogram segmentation. This is because of the low contrast with ambiguous margins, connected with the normal tissues, and of various scales and complex shapes. To detect the boundaries effectively they used a contour-based level set method which extracts the initial boundaries on the smoothened mammogram as the shape constraint. The relaxed shape constraint is then used to design a novel stopping function for subsequent vector-valued level set method. This method can effectively find ambiguous margins of the mass regions compared with existing active contours methods.

Region growing is one of the simplest as well as popular algorithms for segmentation. It is a technique for extracting connected region of the image based on some predefined criteria. It is merely a pixel aggregation of the images which satisfy some similarity measures among the pixels values. These criteria can be based on intensity information or edges in the image [16][17]. Region growing algorithms are very much useful for isolating features based on the texture patterns in the image. Tumors in mammogram images are always identified as a pattern distribution of gray values. So region growing algorithms are the best segmentation technique that can extract tumor patterns if any available in the mammogram images. In a paper proposed by Malek et.al [18] mammogram segmentation technique consisting of seed based region growing and boundary segmentation. Seed based region growing is used to identify an initial seed point automatically, which are very rare in mammogram applications. Starting with seed point the region will grow by appending to each seed those neighboring pixels that have properties similar to the seed.

A study on digital mammogram segmentation and tumor detection proposed by Rejani and Selvi [19] uses region splitting and region filling with the Discrete Wavelet Transform (DWT), artificial intelligence techniques and artificial neural networks. The fractal dimension analysis is used to find the roughness value, which locate the region suspicious for cancer in the mammogram. The dogs-and-rabbit algorithm initiates the clustering. Region splitting and fillings are used to segment the suspicious region. Finally the back propagation neural network is applied at the end to determine whether a given mammogram is suspicious for cancerous.

Most of the mammogram segmentation methods discussed above found to give good results and are being used in different mammogram analysis system. Nevertheless there have limitations largely owing to the complex structure as well as the fuzzy like nature of mammogram. Most of the methods discussed above are semi-automated. Few methods use Artificial intelligence related techniques and some of them are using frequency transformation methods for identifying the suspicious area in the mammogram. Normally tumor cells in the mammograms are labeled by high intensity values corresponding to the position of the tumor, but treated as noise with high probabilities during the preprocessing stage. So the identification of this portion is the most challenging tasks in any mammogram segmentation technique. The intensity patterns in a mammogram image may be the indication of different types of tumors. Therefore the intensity variation may be thoroughly analyzed for locating the exact tumor location in the image. The most important part of all the segmentation algorithms is the identification of the seed point in the image, which can be the center point of the tumor. In this paper we propose a mammogram segmentation method by incorporating some of the existing approaches. The algorithm has three stages. In the first stage, an automatic seed point detection. This is followed by the region growing algorithm for collecting the neighboring pixels around the seed point using some similarity measures. Finally the boundaries of the extracted regions are identified and labeled using gradient operators. From this boundary points suspected tumor areas are isolated and treated as ROIs of the mammogram for further investigation.

3.METHODOLOGY

Most of the mammogram image consists of background information which cannot provide any information regarding the breast or tumors in the breast. Removing such background information from the mammogram leads to considerable reduction in data size as well as the time for processing the data. The extracted gray levels of the mammograms then constitute Region of Interest (ROI) used for further analysis [20]. A fully automated mammogram segmentation method is presented below. It consists of three steps seed point detection, region growing and cropping of the images. The flow chart of the process is given in figure 1.

A standard mammogram image contains dark background and the breast tissue. Normally the breast tissue is brighter than the background. An easy way to locate the seed point in a mammogram image is to identify the centre row of the mammogram image. Normally this centre row of the mammogram image has high gray values which indicate breast tissue [21]. Neiber et.al [21] proposed an algorithm for identifying the seed point by subdividing the centre row into blocks of 100 pixels and the mean of the gray level of the each block is estimated. By using two specific thresholds the mean is associated to three gray levels conditions such as background, breast tissue and noise of the image. While computing the mean of each block, there may be a situation in which mean value is not associated to any gray level values in the image. Normally this situation is rare one and the method proposed in [21] produced good segmentation results for the images they used. So in this paper we proposed a modified version of the above algorithm by subdividing the centre row into consecutive blocks of 50 pixels and the median of the gray level in each block is calculated. Using median of each block, true gray value of the block is obtained. Using three specific thresholds, the median is associated to three gray level conditions for an 8-bit gray scale mammogram image. A gray level 0 – 89 is treated as background, gray level 90-230 as breast tissue and gray level greater than 230 is taken as unwanted or noise part of the image. The adjacent blocks with the same gray levels are connected to form chains. The longest chain with medium gray level in between 90-230 is attributed to a seed point that corresponds to the breast tissue. The threshold values are selected on empirically by choosing different threshold value to background, breast tissue and noise on the pre-labeled images from the standard Mini-Mias dataset. The entire process is depicted in figure 2.

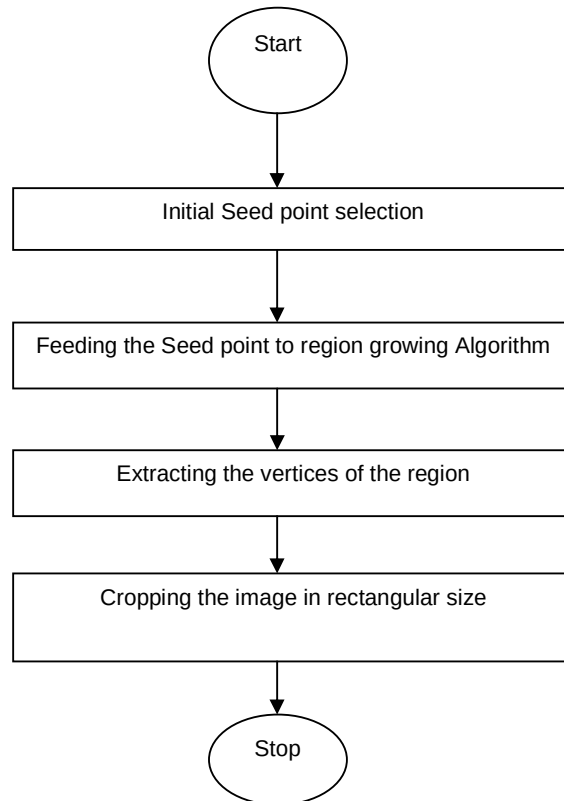


Figure 1: Flow chart of the automatic segmentation process

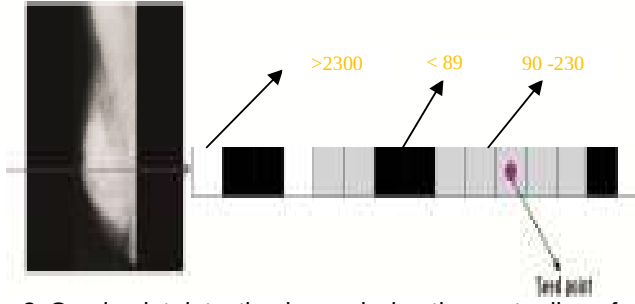


Figure 2: Seed point detection by analyzing the center line of a mammogram

In the next stage of the **algorithm**, pixel aggregation around the seed point is done by collecting all the 8-connected set of pixel values neighbors to the seed point based on the similarity measure which is defined on a threshold value such as:

$$\text{Threshold} = (\text{MaxIntensity}(\text{ROI}) - \text{MinIntensity}(\text{ROI})) * 0.05 \quad (1)$$

Starting from the seed point every 8-connected neighbors are checked whether their gray level satisfies the similarity property. Every neighbor pixel which satisfies the similarity condition is added to the segmented region and act as a new seed point for the next iteration. Thus a recursive process of pixel aggregation is continued till no further seed point which satisfies the similarity conditions in the image. After the pixel aggregation process, the boundaries of the segmented region are isolated using gradient operators. **This is done by identifying the four border pixel coordinates such left, right, top and bottom pixel coordinates of the aggregated pixels and then cropped the image using these four coordinates.** Finally from the boundaries of the aggregated region a rectangular portion of the image is cropped and treated as the ROIs which contain required gray values for further investigation.

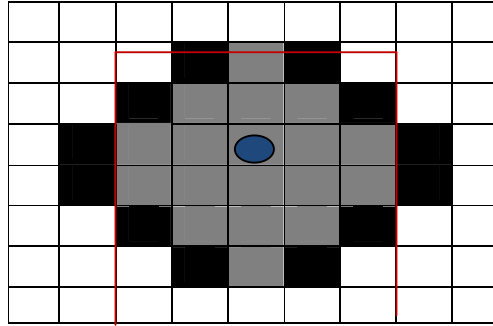


Figure 3: Sub region aggregation of 5x6 pixels from 8x9 pixel image.

4.RESULTS AND DISCUSSION

We used the mammogram images from Mini-MIAS[22] dataset and sample images obtained from the local hospitals for testing the **algorithm** suggested in this paper. All the images in the Mini-MIAS dataset consist of 256x256 pixels with 256 gray levels. The proposed segmentation **algorithm** segmented 64 benign and 52 malignant images in the Mini-MIAS dataset as well as 32 images from the local hospital in Kerala. Most of the extracted ROIs contain abnormality portions which are already confirmed by the radiologist. A sample of mammogram images in the Mini-MIAS dataset along with segmented images are shown in Figure 4. The digital mammogram images obtained locally are also have 8 bit per pixel that comprised [0-255] gray levels. We also segmented these mammogram images using the proposed method. It is clear that most of these images are segmented with proper abnormality identified and labeled by expert radiologist. Samples of images obtained from the local hospitals and its segmented portions are shown in Figure 5. Therefore we can have conclusion that this **algorithm** is an effective and most usable for segmenting abnormal portions of the digital mammograms.

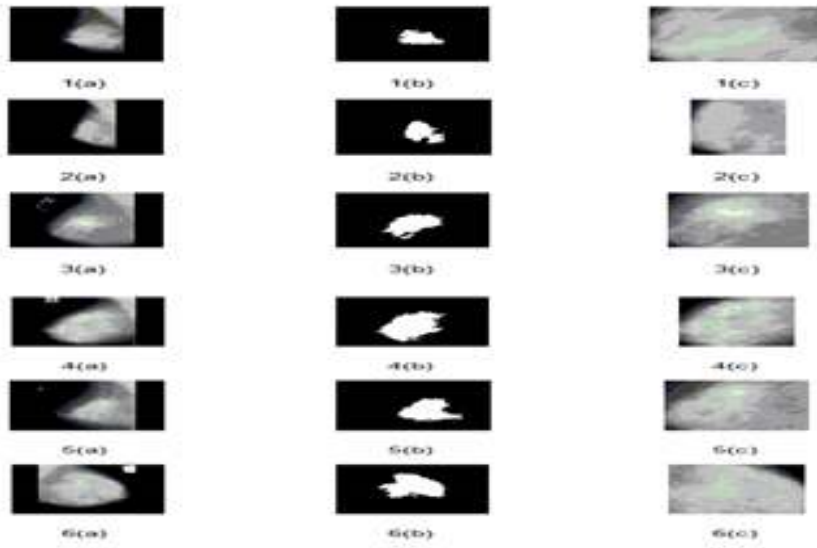


Figure 4: Segmented images – Mini-MIAS (a) Original image (b) Segmented Images (c) Extracted ROIs

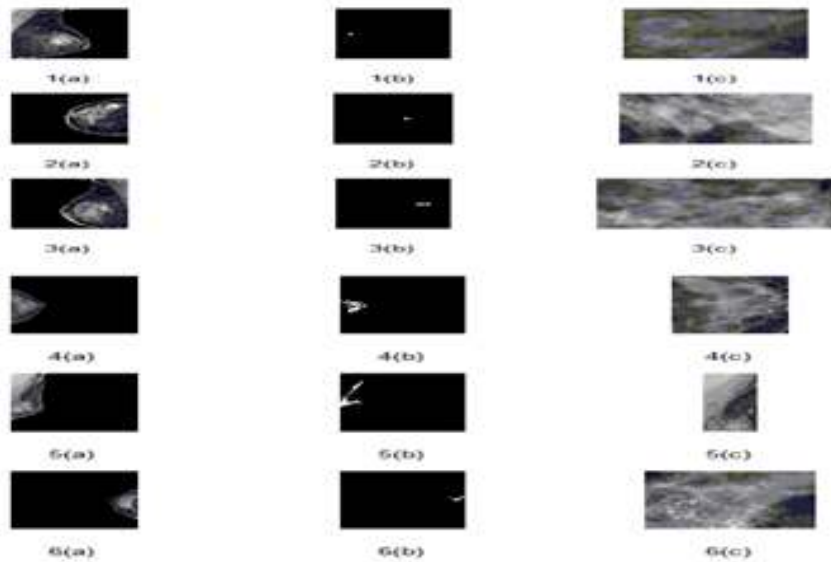


Figure 5: Segmented images -Local dataset (a) Original image (b) Segmented Images (c) Extracted ROIs

5.CONCLUSION

Digital mammograms are among the most difficult medical images to be analyzed due to the low tissue image contrast and slight perceptible differences. Hence the identification and locating abnormalities in a mammogram images are very challenging. In this paper we proposed a fully automated segmentation algorithm, which exploited the automatic seed point location using the median value of the predefined block of pixels in the image. The region of interest around the seed point is extracted using the modified version of region growing algorithm for aggregating the pixels around the seed point. Finally the gradient operators are used for identifying the boundaries of the segmented region. Using these boundaries, the segmented region of the images are cropped and treated as the ROIs of the mammogram that may constitute the tumor regions. The individual ROIs of the segmented images can be then used further for feature extraction and classification purpose. By inspecting the extracted ROIs, the proposed segmentation algorithm extracted 129 ROIs out of 149, which contain abnormality content exactly as defined and labeled by the Radiologist. Out of the remaining ROIs, two of them did not agree in any way to the information provided by the Radiologist. It is also computed the average time for extracting one mammogram image into its ROI as 3.7393 seconds.

REFERENCES

1. Tinku Acharya, Ajoy K. Ray, Image Processing Principles and Applications, John Wiley & Sons, Inc Publication, ISBN-13 978-0-471-71998-4, 2005.
2. A. Rosenfeld, A. C. Kak, Digital Image Processing, New York: Academic Press, 1982.
3. R.C. Gonzalez, An overview of image processing and pattern recognition techniques, Handbook of Geophysical Exploration pattern recognition and Image Processing, Vol. 20, Geophysical Press, New York, 1987.
4. W. K. Pratt, Digital Image Processing, 2nd edition, New York: John Wiley & Sons, 1991.
5. A. Rosenfeld and A.C. Kak, Digital Picture Processing, 2nd ed., Vol. 1, Academic Press, 1982.
6. Issac N. Bankman, Handbook of Medical Image Processing, Academic Press, Elsevier, ISBN 978-0-12-373904-9, 2009.
7. A. Mitiche, J. K. Aggarwal, "Image Segmentation by conventional and information-integrating techniques: a synopsis," Image and Vision Computing, 3(2), 50-62, 1985.
8. K. Lim, A. Pfefferbaum, "Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter," Journal Of Comput. Assist. Tomogram, 13, 588-593, 1989.
9. F. A. Lucas-Quesada, U. Sinha, S. Sinha, "Segmentation strategies for breast tumors from dynamic MR images," *JMRI*, 6, 753-763, 1996.
10. Guillaume Kom, Alain Tiedeu, Martin Kom, "Automated detection of masses in mammograms by local adaptive Thresholding", Computers in Biology and Medicine, 37, 37- 48, Elsevier, 2007.
11. J Padayachee, M.J Alport, W.I.D Rae, "Identification of the breast edge using areas enclosed by isointensity contours", Computerized Medical Imaging and Graphics, 31, 390-400, Elsevier, 2007.
12. R.B Dubey, M Hanmandlu, S.K Gupta, A comparison of two methods for the segmentation of masses in the digital mammograms, Computerized Medical Imaging and Graphics 34, 185-191 Elsevier, 2010.
13. S.L. Lou, H.D. Lin, K.P. Lin, D. Hoogstrate, "Automatic breast region extraction from digital mammograms for PACS and telemammography applications", Computerized Medical Imaging and Graphics 24, 205–220 Elsevier, 2000.
14. Wei-Yen Hsu, Improved watershed transform for tumor segmentation: Application to mammogram image compression, Expert systems with Applications, Vol. 39, 3950-3955, 2012. Second edition, Academic Press, Elsevier, ISBN 0-12-685875-6, 2003.
15. Ying Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, Bin Wang, "Mammographics mass segmentation: Embedding multiple features in vector-valued level set in ambiguous regions", Pattern Recognition, Vol. 44, 1903-1915, 2011, Elsevier.
16. Awcock G.J, Thomas R, "Applied Image Processing". London: The MacMillian Press Ltd. 1996.
17. Adam R, Bischof L "Seed region growing". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No.6, 641-647, 1994.
18. Aminah Abul Malek, Wan Eny Zarina Wan Abdul Rahman, Arsmah Ibrahim, Rozi Mahmud, Siti Salmah Yasiran, Abdul Kadir Jumaat, "Region and Boundary Segmentation of Microcalcifications using Seed-Based Region Growing and Mathematical Morphology", International Conference on Mathematics Education Research 2010.
19. Y. Ireaneus Anna Rejani, S Thamara Selvi, "Digital Mammogram Segmentation and Tumor Detection Using Artificial Neural Networks", International Journal of Soft computing 3(2): 112-119, ISSN: 1816-9503, 2008.
20. Dengler J, S.Behrens, J.F. Desaga, "Segmentation of Microcalcifications in Mammograms", *IEEE Trans. On Medical Imaging*, 12(4), 634-642, 1993.
21. Neiber H, Muller T, Stotzka R, "Local contrast Enhancement for the detection of microcalcifications", Proceedings for the 5th International workshop on Digital Mammography, M. J. Yaffee, Ed., Nijmegen, the Netherlands, 598–604, 2000.

295
296 22.[http://www.wiau.man.ac.uk/services/](http://www.wiau.man.ac.uk/services/MIAS/MIAS/mini.htm) MIAS/MIAS/mini.htm: The Mini-Mias dataset,2008.
297