

Image Enhancement and Edge-based Mass Segmentation in Mammogram

Yu Zhang, Noriko Tomuro, Jacob Furst, Daniela Stan Raicu
School of Computing, College of Computing and Digital Media
DePaul University, Chicago, IL 60604, USA
{jzhang2, tomuro, jfurst, draicu}@cs.depaul.edu

ABSTRACT

This paper presents a novel, edge-based segmentation method for identifying the mass contour (boundary) for a suspicious mass region (Region of Interest (ROI)) in a mammogram. The method first applies a contrast stretching function to adjust the image contrast, then uses a filtering function to reduce image noise. Next, for each pixel in a ROI, the energy descriptor (one of the Haralick descriptors) is computed from the co-occurrence matrix of the pixel; and the energy texture image of a ROI is obtained. From the energy texture image, the edges in the image are detected; and the mass region is identified from the closed-path edges. Finally, the boundary of the identified mass region is used as the contour of the segmented mass. We applied our method to ROI-marked mammogram images from the Digital Database for Screening Mammography (DDSM). Preliminary results show that the contours detected by our method outline the shape and boundary of a mass much more closely than the ROI markings made by radiologists.

Keywords: Mammogram, Image Enhancement, Mass Segmentation

1. INTRODUCTION

In 2008, it was estimated that 182,460 new cases of breast cancer would be diagnosed among women in the United States. Breast cancer is the second leading cause of cancer related deaths for women in the U.S. after lung cancer [1]. At present, the most effective method for breast cancer early detection is mammography screening [1]. Radiologists review mammogram images, detect and diagnose any abnormal growths of the breast. Based on the level of the radiologist's suspicion of the abnormality, a patient usually is recommend a routine follow up, or a biopsy to confirm the result [2].

The problem with mammography screening is that the error rate is high. At present, the sensitivity of most mammography reports is in the 68–92% range [3]. A study showed that among the women who had mammogram screenings and were recommended to have a biopsy for further confirmation, only 25% of the cases were found to have breast cancer [4]. If the accuracy of the diagnosis from radiologists on those abnormalities in mammograms can be improved, some breast biopsies could be avoided. Many Computer-Aided Diagnosis (CADx) systems have been developed as a second opinion to assist radiologists in their diagnosis [5]. Previous studies have shown that CADx systems can help radiologists analyze a suspected abnormality in a mammogram, and improve the diagnosis accuracy [6].

Mass and microcalcifications are the two most common types of abnormalities in mammogram images. The research presented in this paper is part of an ongoing project whose aim is to develop an image-based CADx system for classifying suspicious masses in mammograms as malignant or benign. The goal of the CADx system is to assist radiologists as a second opinion in their diagnosis of suspicious masses and to increase the diagnosis accuracy. The proposed CADx will include five parts: mass ROI extraction, mass segmentation, mass feature extraction, feature selection and classification (ensemble learning). At the classification stage, we plan to apply an optimal feature set to multiple classifiers, and construct an ensemble of classifiers. Figure 1 below depicts the schematic framework of our CADx system.

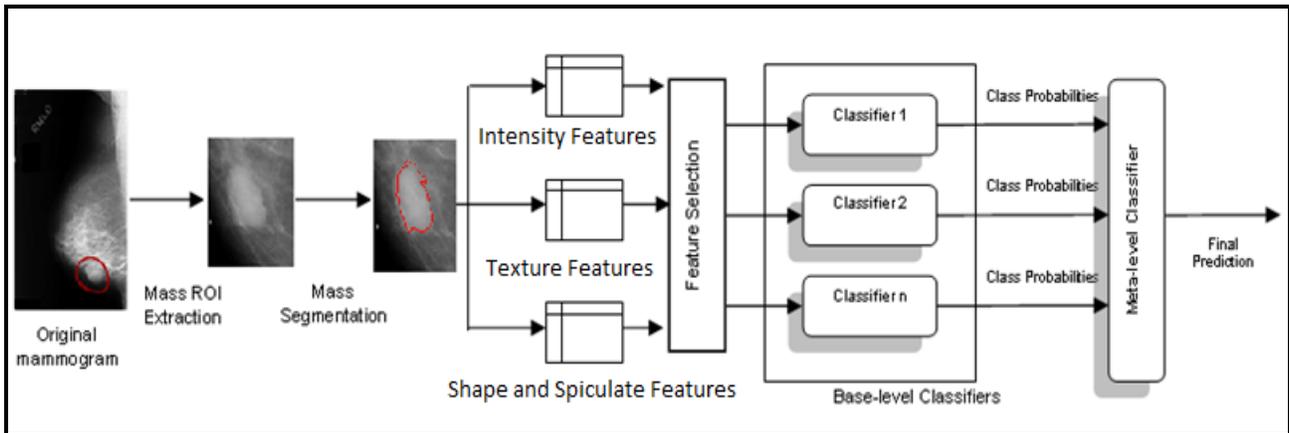


Figure 1. Framework of proposed Computer-Aided Diagnosis (CADx) system

For this paper, we focus on mass segmentation, which separates a mass from a ROI image and captures the boundary of the mass. In the research, a suspicious mass (ROI) is a rectangle image extracted from the original mammogram; it includes a suspicious mass and its surroundings, where the ROI has been detected and marked by either computer-aided detection (CAD) systems or radiologists.

In this paper, we present a novel, edge-based mass segmentation method for identifying the mass contour (boundary) for a mass ROI in a mammogram. In the experiment, we first applied image enhancement to a mass ROI image, and then computed the energy descriptor (one of the Haralick descriptors [7]) for each pixel of the ROI and built the energy texture image of the ROI. From the energy image, the edges were detected, then the possible mass regions were identified from the closed-path edges. Finally, the boundary of an identified mass region was used as the final segmented outline of the mass. This paper is organized as follows. In Section 2, we review related works of mass segmentation. In Section 3, we present the data and methodology. In Section 4, the preliminary results are presented and the image enhancement for mass segmentation is discussed. In Section 5, the conclusion and future work are discussed.

2. MASS AND MASS SEGMENTATION

Masses are thickenings of breast tissue which appear as lesions in mammograms. For radiologists, the shape and margin (spiculated level) of masses are two most important criteria in distinguishing malignant from benign masses. Usually a mass with poorly defined shape is more likely to be malignant than a well-circumscribed mass; and a mass with ill-defined margins or spiculated lesions is much more likely to be malignant than a mass with smoothed margins [2, 6]. In order to standardize the annotation scheme of mammogram reports, Breast Imaging Reporting and Data System (BI-RADS) has been developed by the American College of Radiology (ACR). BI-RADS includes a set of standardized lexicons describing mammography lesions. In BI-RADS, mass shapes are defined as round, oval, lobulated, irregular or architectural distortion; and mass margins are defined as circumscribed, obscured, microlobulated, ill-defined or spiculated [8].

In a CADx system, mass segmentation is an important step, which separates a mass from its background and captures the shape and boundary of the mass [6]. After mass segmentation, the contour of a mass is identified; and the shape features and spiculated level of the mass can be computed for classifying the mass as benign or malignant. Previous studies have shown that it is critical for a CADx system to extract accurate shape and boundary information for suspicious masses; and that improving the mass segmentation can significantly improve the accuracy of mass diagnosis [9, 10].

Many mass segmentation methods for mammogram images have been developed [9, 10, 11, 12, 13, 14, 15, 16]. There are two basic approaches: 1) region-based mass segmentation methods and 2) edge-based mass segmentation methods.

In region-based methods, mass regions are iteratively grown by comparing all neighboring pixels and including the pixels with similarity to the respective regions. The similarity can be measured with different types of properties, such as pixel intensity or computed texture features. In edge-based methods, segmentation is based on edge detection. Usually the images need to apply enhancement before segmentation in order to enhance relevant edges prior to the edge detection stage [14].

Using a region-based segmentation approach, Jiang et al. [9] developed a new mass segmentation method. Their method first applied a gamma correction and a Gaussian filter to a mass ROI to improve contrast and remove image noise. Then, using the principle of maximum entropy, optimal thresholds were selected to obtain initial segmented mass regions; and a morphologic dilation operation was applied to the binary image to obtain the initial contour of the mass region. Mencattini et al. [11] modified a region-growing mass segmentation procedure in their Computer Aided Detection (CAD) system. Their segmentation method included: removing artifacts based on the analysis of the ROI histogram, improving mass image contrast by applying a non-linear operator, and a region growing segmentation, where the center of a mass ROI was chosen as an initial seed position, and pixel intensity was used for similarity measurement. Xu et al. [16] developed a region-based mass segmentation algorithm. In their system, the suspicious areas were extracted using iterative thresholding. Then the suspicious area was decomposed by 3-level DWT (discrete wavelet transform) into four sub-bands and centers of the masses were located. Finally, the Canny edge information of each pixel in the ROI was computed, and the region growing technique was applied to merge adjacent regions.

Using an edge-based approach, Song et al. [12] developed a mass segmentation method. Using plane fitting and dynamic programming techniques, their method detected the "optimal" contour of a mass from the edge candidate points. Yuan et al. [13] developed a method for mass lesion segmentation on mammograms using a geometric active contour model. In their method, a radial gradient index (RGI)-based segmentation method was applied to yield an initial contour. Based on the initial segmentation, an automatic background estimation was applied to identify the effective circumference of the lesion, and a dynamic stopping criterion was implemented to terminate the contour evolution when it reached the lesion boundary.

3. DATA AND METHODOLOGY

3.1 Data description

In our experiment, we extracted mass ROI images from the Digital Database for Screening Mammography (DDSM) [17] from the University of South Florida. DDSM is the largest publicly available resource for the mammogram analysis research community. In DDSM images, suspicious regions (ROIs; including masses and microcalcifications) are marked by experienced radiologists, and BI-RADS information is annotated for each abnormal region [18]. However, in DDSM, most mass ROIs are marked by a circle for the mass location, and do not trace the detailed outline of the mass.

In DDSM, mammogram images are digitized by different scanners with different resolutions. In this research, for data consistency purposes, all mass ROI images are collected from the same type of scanner and resolution. We chose the scanner type LUMSYSYS because the largest number of cases are digitized by this type in DDSM. In our experiment, we first extracted all mass ROIs from the mammogram images digitized by LUMSYSYS. Then, we removed the mass ROI instances with extreme digitization artifacts (e.g. incorrectly ordered scan lines) and extreme large ROI images (size over 2000 x 2000 pixels). In addition, the mass instances with mixed BI-RADS descriptors and the ROI images which display only a portion of a mass were removed from our dataset. After removing those instances, a total of 525 mass ROI images were left for this study, where 263 instances were benign and 262 instances were malignant. Figure 2 and 3 show the distribution of the BI-RADS shape and margin features respectively.

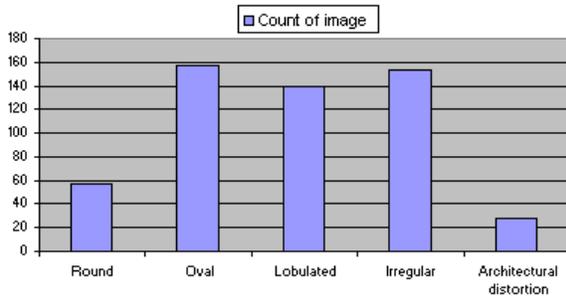


Figure 2. Mass BI-RADS shape distribution

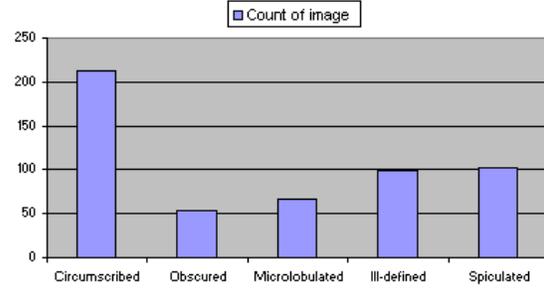


Figure 3. Mass BI-RADS margin distribution

3.2. Methodology

In DDSM, a suspicious mass region marked by radiologists was represented as a chain codes in an overlay file [18]. In our experiment, for each suspicious mass shown on a mammogram, a rectangle image was extracted as a ROI, which includes the suspicious mass and its surroundings. Then we applied the following four steps to segment a mass and obtain the mass boundary from the ROI image. All procedures were implemented by using MatLab in this study.

Step1: Image enhancement

Since most mammogram images have low intensity contrast, image enhancement processing such as contrast adjusting functions are often applied to mammogram images before performing segmentation. In our experiment, for image enhancement, we used a linear contrast adjusting function (gamma correction value = 1) to increase the image intensity contrast, then applied an averaging filtering with a size of 30 x 30 pixels to remove noise in the image. Figure 4 (a-c) shows an example of an original ROI image (a), the contrast adjusted image (b), and the contrast adjusted image after averaging filtering is further applied (c); and Figure 4 (d-e) displays the histograms of those images after contrast adjusting and averaging filtering.

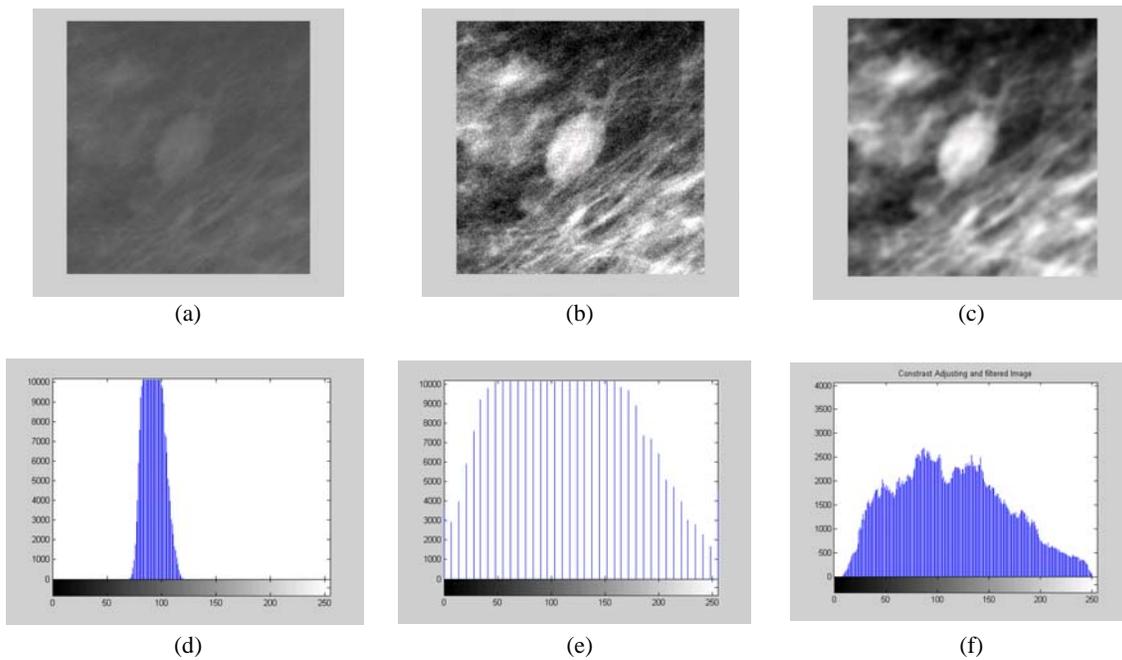


Figure 4. Original ROI image, enhanced images and their histograms

Step 2: Construct a texture image – pixel-based energy image

Haralick descriptors [7] are computed from a gray-level co-occurrence matrix of an image, and they are the most commonly used texture features in image processing. Energy is one of the Haralick descriptors, which measures the heterogeneity of an image. In our experiment, for each pixel in a ROI, we first computed a co-occurrence matrix with angle = 0 and distance = 1 using a neighborhood size of 7x7 pixels [18]. Then, we obtained an energy value for each pixel by computing the number of occurrences of the pairs of the same value within the co-occurrence matrix:

$$Energy = \sum_i^M \sum_j^N P_{ij}^2 \quad (1)$$

where M and N are the size of row and column of the matrix respectively, and P_{ij} is the probability of the pixel pair (i, j) in the matrix. Finally an energy texture image was constructed from the energy value of all pixels in the mass ROI image.

Figure 5 shows the energy texture images computed from the images in Figure 4 (a-c). As you can see, contrast in the energy texture images are gradually enhanced and with less noise -- the original mass ROI (5(a)) to the contrast-adjusted image (5(b)) to the contrast-adjusted and filtered image (5(c)).

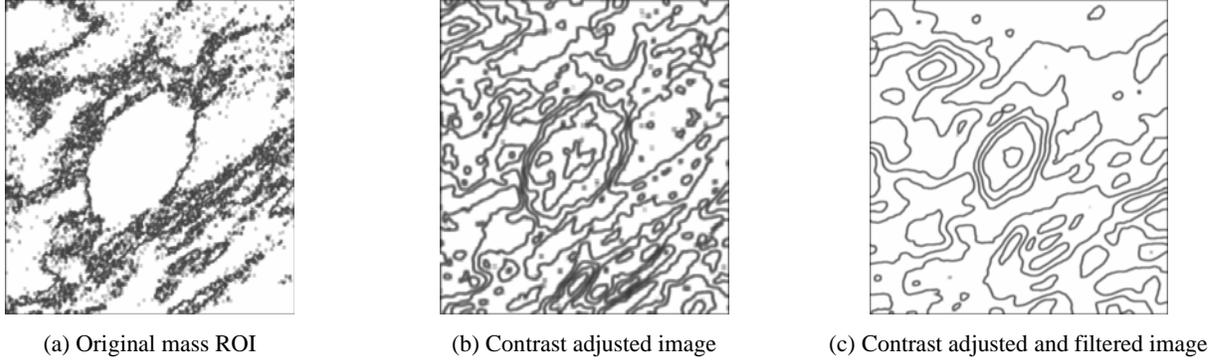


Figure 5. Energy texture images

Step 3: Find the edges and detect the possible mass regions

In the energy texture image (5(c)), we applied an adaptive thresholding method to detect edges. Then by setting a value of 1 to the pixels on the detected edges and a value of 0 to all other pixels, a binary image with detected edges was obtained. Figure 6 (a) shows a binary image with detected edges. Next, in the binary image, we applied a morphological operation (imfill function in the Matlab) to detect all closed-path edges, and used those edges as boundaries to find all possible mass regions (the white regions in Figure 6(b)). Finally, the region with the largest overlapping ratio over the ROI central area was identified as the segmented mass region (Figure 6(c)). In our experiment, the ROI central area (the red rectangle area in Figure 6(b)) was defined as the rectangle of half width and half length of the original mass ROI.

Step 4: Evaluate the mass segmentation

At this step, the segmented mass was evaluated. In our experiment, using a threshold of overlapping ratio of pixels of the segmented mass over the ROI central area, we determined whether or not the identified mass was successfully segmented. For a successfully segmented mass, the boundary of the identified region was used as the segmented mass contour. In Figure 6 (d), the green line is the successfully segmented mass contour identified by our algorithm, and the red line is the mass outline marked by the radiologist.

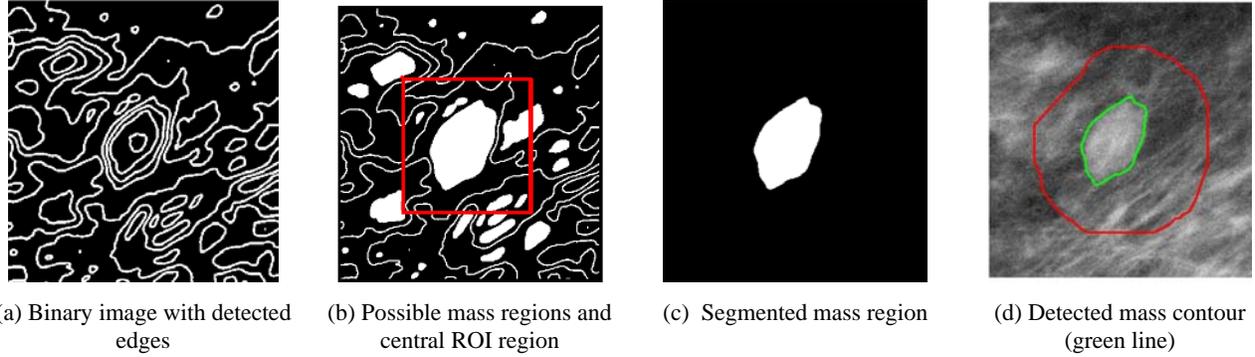


Figure 6. Mass segmentation

4. PRELIMINARY RESULTS AND DISCUSSION

4.1 Preliminary results

The results showed that, by using the aforementioned parameter values (the contrast adjusting function with $\gamma=1$ and the averaging filtering with size of 30×30 pixels), our segmentation method successfully segmented 355 (67.6 %) instances among the 525 mass ROIs with the threshold of 25% overlapping ratio over the central area.

Figure 7 (a-c) shows three examples of successfully segmented masses. In each image, a red line represents the radiologists' marked mass region, and a blue line represents the contour of a segmented mass obtained by our segmentation method. By visual inspection, the contours by our segmentation method outline the mass shapes and boundaries much more closely than the markings made by the radiologists. Figure 7 (d) on the other hand shows one example of an unsuccessfully segmented mass, where the segmented mass is not in the center area of the ROI. We will discuss more on the unsuccessfully segmented instances in the next section.

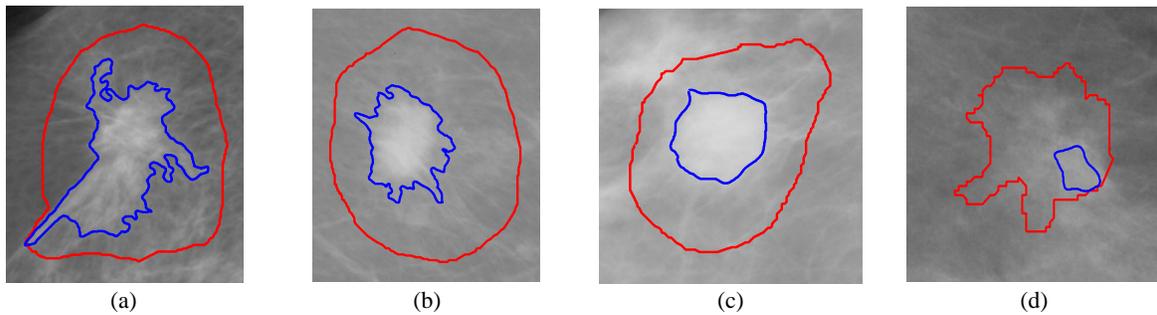


Figure 7. Examples of segmented mass instances

4.2 Image enhancement and mass segmentation

In our experiment, we found that image enhancement greatly affects the result of mass segmentation. We observed that there is no "one optimal" enhancement that can produce a best fit on all images for mass segmentation. Figure 8 displays the segmentation results of two mass ROI instances using six different image enhancements. For each instance, six different image enhancements are shown on the left - they were the combinations of three values used for the parameter in the contrast adjusting function (γ correction = 0.5, 1 and 2), and two values used for the parameter in the averaging filtering function (size 10×10 and 20×20 pixels). For the contrast adjusting function, the adjusting is weighted toward brighter values when $\gamma = 0.5$; it is weighted toward darker values when $\gamma = 2$; and it is

linear adjusting when $\gamma=1$. For the filtering function, a larger filtering size removes more noise, but some detail of the image is lost at the same time.

The six binary images on the right side are the results of mass segmentation extracted from those six enhanced images. For Instance 1, when the contrast adjusting with $\gamma = 2$ and filtering function with size 20×20 pixels, the instance obtained a better segmentation contour compared to other enhancements; while the best values for Instance 2 seem $\gamma = 0.5$ with either filtering size (10×10 or 20×20 pixels). As exemplified by those images, different mass instances may require different image enhancement functions to achieve a better segmentation.

The experiment results demonstrated that, depending on the image's contrast range and noise level, the "optimal" image enhancement for mass segmentation is different for each individual image. Generally speaking, for a darker image, the image enhancement should apply a contrast adjusting function with γ correction value less than 1 (such as 0.5) to increase the intensities of the dark areas; while for a brighter image, it seems more effective to use a contrast adjusting function with γ value greater than 1 (such as 2) to increase the contrast of the brighter areas in the image.

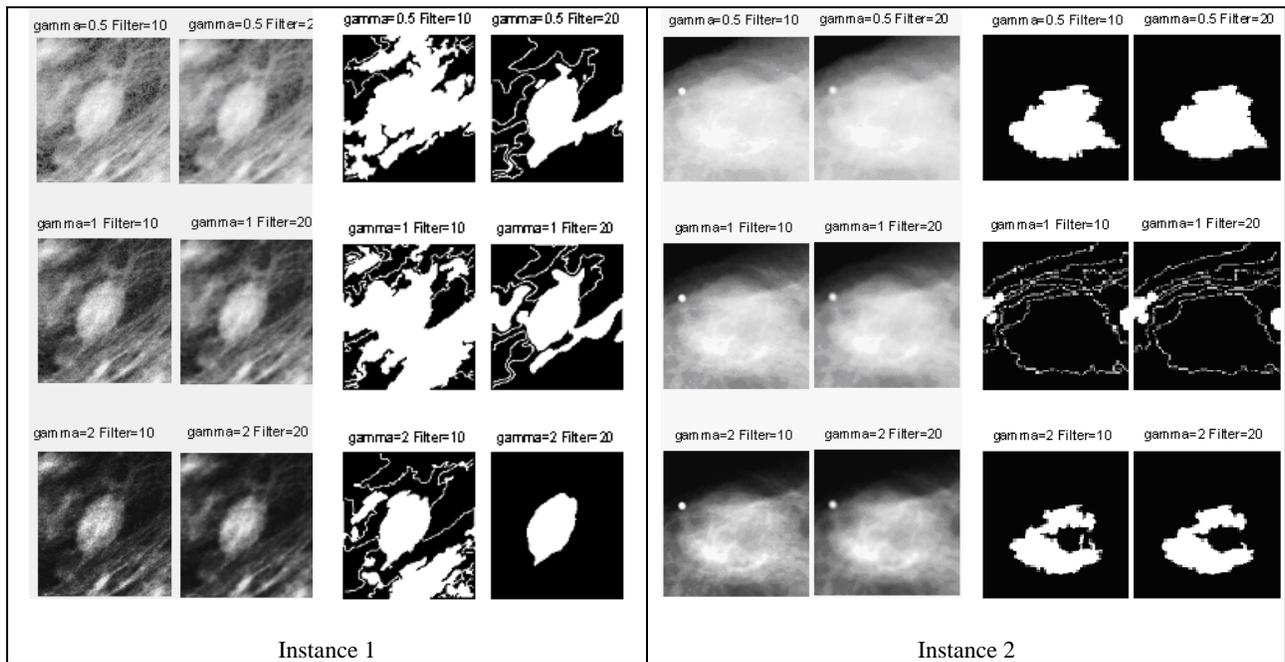


Figure 8. Mass segmentations with different image enhancement

5. CONCLUSIONS AND FUTURE WORK

Most mammogram images are in low intensity contrast, making them very difficult to detect the edges of a suspicious mass (a region of interest (ROI)). In this paper, we presented a novel, edge-based segmentation method for detecting the contour of a suspicious mass region in mammograms. Our method extracts edges from a pixel-level energy texture image of a mass ROI, and then detects closed-path edges to identify the mass region. Preliminary results showed that the mass contours captured by our segmentation outlined the boundary and shape of a mass much more closely than the radiologists' markings in DDSM.

In our experiment, we found that image enhancement greatly affects the quality of segmentation. Mammogram images have varied intensity contrast ranges and different noise levels. Also, different patients have various breast density levels. Those variability factors are making it difficult to select an/one optimal image enhancement which can fit all images for mass segmentation.

In future work, we plan to investigate a scheme where an image is processed by multiple segmentors where each segmentor uses a different set of image enhancement (by using varied gamma correction values for contrast adjusting functions and filtering sizes) and combining those segmentations in an ensemble to classify a suspicious mass as benign or malignant. We expect that the combination of multiple weak segmentors in an ensemble can produce better results (higher classification accuracy) than a single strong segmentor.

References

- [1] National Cancer Institute, "American Cancer Society Cancer Facts & Figures 2008" (<http://www.cancer.org>).
- [2] Winchester, D. J., Winchester, D. P., Hudis C. A. and Norton, L., [Breast Cancer(Second Edition)] , Springer (2007).
- [3] Yankaskas, B. C., Schell, M. J., Bird, R. E. and Desrochers, D. A., "Reassessment of Breast Cancers Missed During Routine Screening Mammography: A Community-Based", American Roentgen Ray Society, vol. 177, 535-541 (2001).
- [4] Ries, LAG, Harkins, D., Krapcho, M., et al., "SEER Cancer Statistics Review, 1975-2003", National Cancer Institute (2006).
- [5] Mu, T., Nandi, A. and Rangayyan, R., "Classification of Breast Masses Using Selected Shape, Edge-sharpness, and Texture Features with Linear and Kernel-based Classifiers", Journal of Digital Imaging, Volume 21, Number 2, 153-169 (2008).
- [6] Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P, et al., "Approaches for Automated Detection and Classification of Masses in Mammograms", Pattern Recognition (2006).
- [7] Haralick, R. M., "Statistical and Structural Approaches to Texture", Proceedings of the IEEE, vol. 67, pp.786-804 (1979).
- [8] D'Orsi, C. J., Bassett, L. W. and Berg, W. A., [Breast Imaging Reporting and Data System: ACR BI-RADS-Mammography" (ed 4)], American College of Radiology, Reston, VA (2003).
- [9] Jiang, L., Song, E., Xu, X., Ma, G. and Zhang, B., "Automated Detection of Breast Mass Spiculation Levels and Evaluation of Scheme Performance", Acad Radiol (2008).
- [10] Domínguez1, A. R. and Nandi, A. K., "Toward Breast Cancer Diagnosis Based on Automated Segmentation of Masses in Mammograms", Pattern Recognition, Volume 42 (2009).
- [11] Mencattini, A., Rabottino, G., Salmeri, M., Lojacono, R. and Colini, E., "Breast Mass Segmentation in Mammographic Image by an Effective Region Growing Algorithm", Advanced Concepts for Intelligent Vision Systems Conference (2008).
- [12] Song, E., Jiang, L., Jin, R., Zhang, L., Yuan, Y. and Li, Q., "Breast Mass Segmentation in Mammography Using Plane Fitting and Dynamic Programming", Academic radiology 16(7), 826-35(2009).
- [13] Yuan, Y., Giger, M. L. Li, H., Suzuki, K. and Sennett, C., "A Dual-stage Method for Lesion Segmentation on Digital Mammograms ", Med. Phys. Volume 34, Issue 11, 4180-4193(2007).
- [14] Marlagelada, O., "Automatic Mass segmentation in mammographic images" (Doctoral Dissertation), Universitat de Girona (2007).
- [15] Kupinski, M.A. and Giger, M.L., "Automated Seeded Lesion Segmentation on Digital Mammograms", IEEE Transaction on Medical Imaging, vol.12 (1998).
- [16] Xu, W., Xia, S., Xiao, M. and Duan, H., "A Model-based Algorithm for Mass Segmentation in Mammograms", Engineering in Medicine and Biology 27th Annual Conference (2005).
- [17] DDSM, The Digital Database for Screening Mammography, <http://marathon.csee.usf.edu/Mammography/Database.html>
- [18] Heath, M., Bowyer, K., Kopans, D., Moore R. and Kegelmeyer, W. P., "The Digital Database for Screening Mammography", Proceedings of the Fifth International Workshop on Digital Mammography, Medical Physics Publishing, 212-218 (2001).
- [19] Susomboon, R., Raicu, D.S. and Furst, J.D., "Pixel-Based Texture Classification of Tissues in Computed Tomography", DePaul CTI Research Symposium (2006).