A Hierarchical Pipeline for Breast Boundary Segmentation and Calcification Detection in Mammograms

Peng Shi[1¶*], Jing Zhong[2¶], Andrik Rampun[3] and Hui Wang[4]

[1] College of Mathematics and Informatics, Fujian Normal University, Fuzhou, Fujian 350117, China

[2] Department of Radiology, Fujian Cancer Hospital & Fujian Medical University Cancer Hospital, Fuzhou, Fujian 350014, China

[3] School of Computing and Information Engineering, Ulster University, Coleraine, Antrim BT52 1SA, United Kingdom

[4] School of Computing and Mathematics, Ulster University, Jordanstown, Antrim BT37 0QB, United Kingdom

[*] Corresponding author

E-mail: pshi@fjnu.edu.cn

[¶] These authors have contributed equally to this work.

Abstract

Breast cancer is one of the most common cancer risks to women in the world. Amongst multiple breast imaging modalities, mammography has been widely used in breast cancer diagnosis and screening. Quantitative analyses including breast boundary segmentation and calcification localization are essential steps in a Computer Aided Diagnosis system based on mammography analysis. Due to uneven signal spatial distributions of pectoral muscle and glandular tissue, plus various artifacts in imaging, it is still challenging to automatically analyze mammogram images with high precision. In this paper, a fully automated pipeline of mammogram image processing is proposed, which estimates skin-air boundary using gradient weight map, detects pectoral-breast boundary by unsupervised pixel-wise labeling with no pre-labeled areas needed, and finally detects calcifications inside the breast region with a novel texture filter. Experimental results on Mammogram Image Analysis Society database show that the proposed method performs breast boundary segmentation and calcification detection with high accuracy of 97.08% and 96.15% respectively. Calculation of Jaccard and Dice indexes between segmented breast regions and the ground truths are also included as comprehensive similarity evaluations, which could provide valuable support for mammogram analysis in clinic.

Key words: breast cancer; mammography; breast boundary segmentation; calcification detection

I. Introduction

Breast cancer is one of the most common cancers among women [1]. Early detection of breast cancer has been shown to be associated with reduced breast cancer morbidity and mortality [2]. In clinics, various breast imaging techniques are used in the early diagnosis and screening of breast cancer, including ultrasound [3], Computed Tomography (CT) [4], Magnetic Resonance Imaging (MRI) [5] and mammography [6]. Mammography is low-energy X-rays imaging of the breast [7], which generates high-resolution images with high bit-depth, and provides the possibility of discovering abnormalities masked by surrounding and overlapping breast tissue [8]. Breast tissues are shown as pixel clusters of different intensities distributed in mammogram images, which mainly include fibro-glandular, fatty, and pectoral muscle tissues. Meanwhile, abnormal tissues including some ill-defined masses or calcification may appear as indicators of breast cancer or other breast diseases, which may also be shown in mammography. Although having limitations, mammography has a sensitivity of 85 to 90% for breast cancer detection [9]. As a proven and effective imaging modality, mammography remains the key screening tool for the detection of breast abnormalities [10].

In a mammography based Computer Aided Diagnosis (CAD) system, features of breast need to be quantified automatically to provide clinical evidences for human experts. A key part of mammography CAD is image segmentation which estimates skin-air boundary and pectoral-breast boundary which together define the breast contours. Existing studies of image segmentation, focusing on breast boundary and pectoral muscle segmentation, could be generally classified into 5 categories: thresholding, region-growing, morphology-based, active contour (AC) and texture-based according to their segmentation approaches. Global [11] or adaptive [12] thresholding methods are usually used to get the skin-air boundary because of significant intensity differences between the background and

foreground tissues. However low contrast between pectoral muscle and the breast has limited the use of thresholding methods to get the pectoral-breast boundary. Chen et al. [13] proposes segmentation methods based on region growing by initializing 40 points along the mask boundary which is obtained by thresholding. Their results show the accuracy of pectoral-breast boundary detection is still lower than that of the skin-air boundary. Morphology-based methods [11, 14] use the natural shape features to build complicated models to fit the objects of breast. The main problem of these model-driven methods is that a generalized shape model is not able to cover all the complex shapes shown in mammography. AC is also a widely used method to segment the breast by initializing a boundary and letting the initialized boundary to approach the actual breast boundary based on minimizing energy functions. However, some of edge-based AC methods [15-17] dealing with mammography only detect the skin-air line with pectoral muscle left in the breast region. Texture-based methods extract textures from images based on texture filters such as wavelet [18] or Gabor filter [19], and determine boundaries between objects based on significant texture changes. Rampun et al. [22] developed a hybrid method by combining a model-based approach and region-based AC which produced very good results but still less accurate when estimating pectoral muscle boundaries with complex contours.

In addition to the image processing approaches, another approach takes image pixels as individual samples and performed segmentation by labeling pixels into different categories. Some researches [20, 21] firstly assign an initial label to each pixel in a mammogram image, and then adjust pixel labels based on energy changes in the Markov Random Fields (MRF) of pixels. The main difficulty of the pixel labeling approach is that they tend to generate unsmooth or discontinuous boundaries, which need post adjustment to find the natural smooth boundaries. Generally speaking, detection of breast boundaries is still a challenging task due to artifacts, homogeneity between the pectoral and breast regions, and low contrast along the skin-air boundary [22]. Abnormal tissues are usually shown on mammogram images with higher densities. Since density is one of the mammographic features that are related to breast cancer risk [23], the separation between glandular and other high density masses is important in characterizing breast tissues.

Furthermore, mammographic calcification is an important feature of invasive and *in situ* breast cancer [24], so the detection of calcification is another key step in mammographic analysis. Calcifications are always small in size and have no significant contrast to surrounding tissues in mammogram images, and the common idea of detection is using various filters to separate the inhomogeneous pixel clusters of calcification from surrounding tissues. A batch of filter-based approaches including gray level thresholding [25], Wavelet [26, 27] and filter banks [28], but it is still challenging for calcification detection because of the small size and low contrast in low quality mammogram images. Therefore further research considering shape and appearance features is needed in order to improve both accuracy and efficiency of filter-based calcification detection.

In this paper, we propose an automated image processing pipeline, based mainly on pixel clustering without training, to estimate breast boundaries and characterize breast tissues concurrently, including skin-air boundary estimation, breast segmentation, and calcification detection. The main contribution of our study includes a pixel-wise clustering scheme plus post processing to segment the breast boundaries precisely, and a novel filter to detect calcifications, which is suitable for most of calcification shapes and sizes in practice.

The rest of the paper is organized as follows. In Section 2, we firstly describe the dataset used to evaluate the proposed method followed by explanations of the technical parts of this study such as detection of skin-air boundary, pectoral muscle to breast boundary segmentation, and calcification localization. In Section 3, experimental results are presented to show the performance of the proposed methods which covers both quantitative and qualitative evaluations. Finally, discussions on the proposed methods and further improvements are presented in the last Section.

II. Materials and Methods

To deal with mammogram image segmentation more efficiently and robustly, an automated hierarchical pipeline including image pre-processing, breast segmentation, boundary detection and calcification localization is proposed, which is shown in Fig. 1. The spatial distributions of different tissues and abnormalities such as calcifications in breast can be extracted from the integrated pipeline, which makes it possible to focus on the region of breast and establish essential indicators for clinical references.

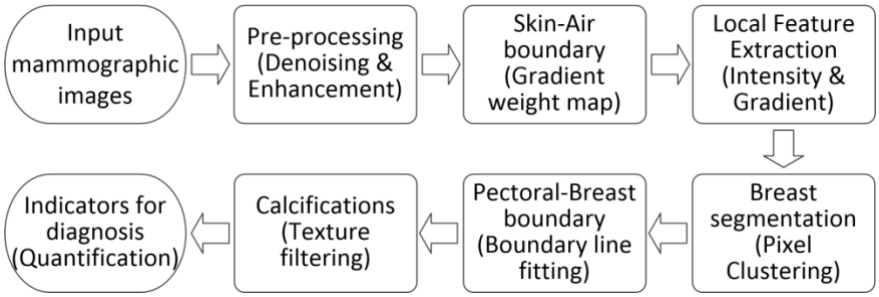

Fig. 1. Workflow of the proposed mammogram image processing and analysis pipeline

A. The Dataset

Generally two kinds of mammography are being used in clinic, including analog Screen-Film Mammography (SFM) and Full-Field Digital Mammography (FFDM) [8]. Although it has been claimed that FFDM has more benefits than SFM [29], analysis of digitally scanned SFM still has realistic significances because comparisons are always needed between prior SFM and current digitalized mammography in many cases, and the processing methods of SFM could be directly applied on FFDM images with few adjustments. On the other hand, most of the mammogram digital datasets nowadays do not contain ground truth or annotations from expert radiologist which make it difficult for quantitative evaluation.

As one of the first publicly released mammography datasets, the Mammogram Image Analysis Society (MIAS) database contains the original 322 images of digitalized SFM at 50 micron resolution [30], and is associated truth data of breast boundary, character of background tissue and especially the locations of calcifications and various masses provided by expert radiologists, in which 8-bits gray scale images are obtained from the original data format, representing the optical density of tissues. For comparison, FFDM images with ground truths of segmentation are also taken into account. 100 images from the 8-bits Breast Cancer Digital Repository (BCDR) database [31] and 201 images from the 32-bit INbreast database [32] are selected for our experiments, and the segmentation process of which are proved no different with that of SFM we mainly discuss.

B. Gradient Map for Skin-Air Boundary

In quite a number of digitally scanned SFM images from the MIAS database, various artifacts have similar high intensities as the breast. To obtain the accurate skin-air boundary, we firstly calculated the weight for each pixel based on the gradient magnitude at that pixel using a 3×3 cross window, and returns the weight array (*W*). The weight of a pixel is inversely related to the gradient magnitude combining gradients of both vertical and horizontal direction at the pixel location, which separated higher intensity areas from low intensity background because sharp gradient magnitude changes occurred on the edges between foreground and background. Then, a horizontal line fitting strategy was applied to remove artifacts such as labels, markers, scratches, and even tapes which adhesive to the top of breast boundary as shown in Fig. 2f.

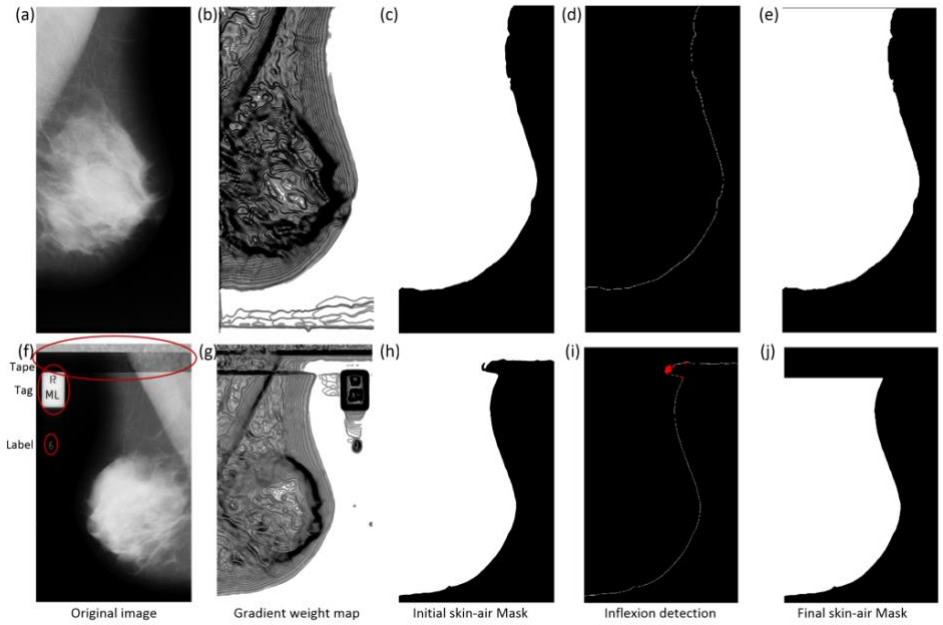

Fig. 2. Skin-air boundary detection with or without artifacts, in which (a) an original image without artifacts, (b) gradient weight map, (c) initial mask of skin-air boundary, (d) skin-air boundary with no inflexion point detected, and (e) final mask of skin-air boundary. (f) an original image with multiple types of artifacts, (g) gradient weight map, (h) initial mask of skin-air boundary, (d) inflexion point detected on skin-air boundary, and (j) final mask of skin-air boundary after with tape artifacts removed.

As shown in Fig. 2, significant differences are along the skin-air line in the image without artifacts. Minor gradient weight noises could be easily removed by image erode techniques [33] as illustrated from the second column to the third, which erased thin lines and kept the main body of gradient weights. Only the largest body was kept as the breast plus pectoral region, and artifacts with smaller sizes were removed. Meanwhile, usually artifacts adhere to the breast break the smooth outline of breast boundary. A 2D curvilinear structure detection method [34] was used on the initial boundary line, which detected the inflection points with sharp changes of curvature, shown as red dots along the initial skin-air line in Fig. 2(i). Then, the tape could be cut off by a horizontal line from the lowest inflexion point based on the natural curves of breast. Closed line as the skin-air boundaries was then acquired as shown in the last column of Fig. 2. Adjusted mask of breast plus pectoral muscle could then be defined for further segmentation.

C. Breast segmentation based on pixel-wise clustering

In order to segment different tissue structures from each other, we used *k*-means clustering with city blocks as a metric measure to find the breast boundary. A median filter [35] for enhancement with five pixels in window radius was applied in de-noising before pixel-wise feature extraction and contrast enhancement between tissues with different densities. Most of the scatter noises were eliminated and small pixel aggregates such as calcifications were enhanced at the same time.

Single image pixels within the skin-air boundary are treated as individual samples in the clustering and aggregates are composed by pixels sharing the same label. To simplify the feature set and accelerate the clustering speed, a two-dimensional vector including normalized ($I_{(x,y)}$, $W_{(x,y)}$) was projected to a 2D feature space in the clustering, in which $I_{(x,y)}$ represented the intensity and $W_{(x,y)}$ was the gradient weight of pixel P(x,y).

$$d_{cityblock}(p_1, p_2) = \sum_{j=1}^{n} \left| p_{1j} - p_{2j} \right| = \left| I_1 - I_2 \right| + \left| W_1 - W_2 \right| \quad (1)$$

In Eq. 1, Cityblock distances were taken as metrics between sample feature vectors in calculating distances between pixels, where P_1 and P_2 are two pixels inside the skin-air boundary respectively. P_{ij} means the *j*-th feature of P, and n=2 here because of the two-dimensional feature vector were being used in the clustering.

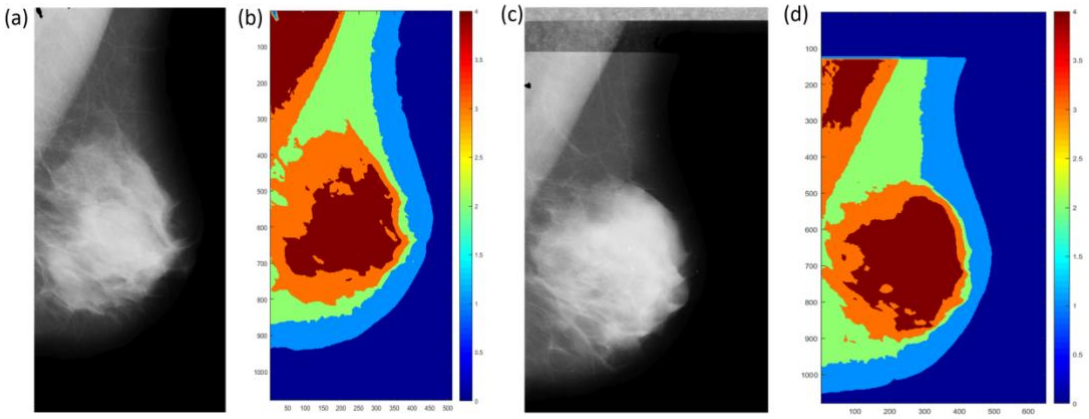

Fig. 3. Pixel-wise clustering, in which (a) and (c) original mammogram images with and without artifacts respectively, (b) and (d) pixel labeling map after clustering, where five different colors represent as different categories including the background, and dark red represents highest intensity as shown in the color bar.

Pixels in the feature space were labeled as four classes following an unsupervised *k*-means clustering strategy, in which to initially make four cluster centers given at random and calculate the barycenter of each sample cluster into new cluster center, and iterated until the displacement distance of cluster center was less than a given value. It was specified that one of the classes contains the areas of glandular, which has comparatively higher intensities and lower gradient weights comparing to other clusters. All four clusters were well separated as shown in Fig. 3, representing the accuracy of the unsupervised clustering based on the proposed features, where five different colors represent as different categories including the foreground tissues (dark red, amber, green and light blue) and background (dark blue). Outline of pectoral muscles were then extracted based on the cluster including part of breast with similar pixel intensities and gradient weights, which was determined by

sorting cluster center values and selecting the center with the highest intensity.

In practice, usually glandular tissues inside the breast and part of pectoral muscles have similar high intensities, which are two separated pixels clusters with highest value shown in Fig. 3(b) and 3(d). Therefore, the parts with highest class labels inside the breast were considered as glandular, and the morphologic quantification could be performed for classification based on glandular proportions.

D. Pectoral to breast boundary line detection

Besides the breast segmented in pixel-wise clustering, part of pectoral muscle is always included in the same cluster because the muscle always has similar strong signals to the glandular in the images. Therefore, further segmentation was required to determine the accurate boundaries between pectoral muscle and breast.

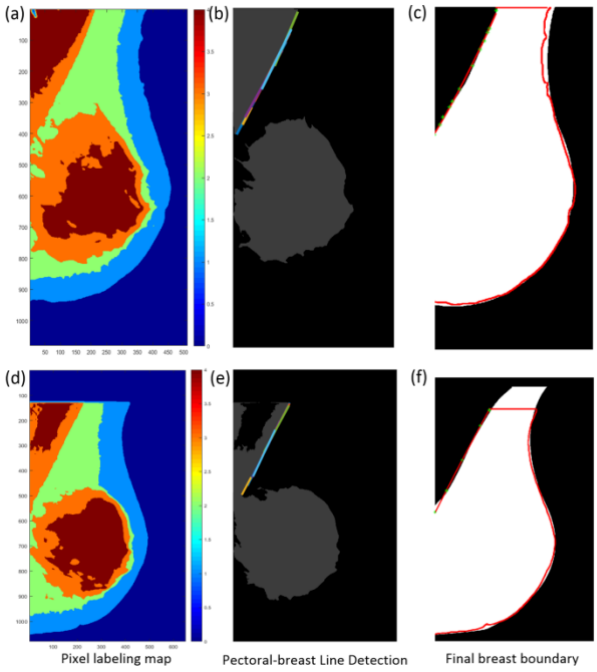

Fig. 4. Pectoral-breast boundary detection, in which (a) and (d) pixel labeling map after clustering, where the amber color regions were retained including the boundaries between pectoral muscles and breast, and morphological operations including image fill and erosion were applied to form regions with smooth edges in the next column, (b) and (e) initial straight lines detected along pectoral-breast boundaries, where lines are shown in different colors, (c) and (f) projected skin-air boundaries (red lines) and fitted pectoral-breast boundaries (red lines with green stars) on the ground truth of breast mask area (white area).

In practice, usually the boundary between pectoral muscle and breast is a smooth line at the upper left corner of the whole image, and it should be along the natural trend of pixel cluster extracted above. A series of built-in Matlab functions were used in smoothing edges of the mask, finding initial points and line fitting. Firstly, morphological operations were needed to adjust segmented pixel cluster including erasing too small regions using an area threshold of 100 pixels, eroding and then dilating bit-wise image mask with a diamond-shape morphological structural element with 3×3 pixels wide, and filling holes and gaps based on connectivity of 4 neighborhood pixels. Then, to get the smooth pectoral-breast boundary, a line fitting approach was applied on the edges of pixel cluster

area according to Hough transform [36], which detected straight line sections with angles between 10° to 80° to the x coordinate, and all initial line segments of the same trend along the edge were marked. Finally, taking the end points of marked straight line sections, the final boundary was obtained by smoothing with a polynomial curve fitting approach [37] to fit the control points, which calculated a polynomial of degree 2 fitting the intial points. As shown in Fig. 4(c) and (f), the green stars are initial end points of the marked line segments, and the red line is the fitted curve along the natural boundary between pectoral muscle and breast. Both sides of the breast were finally defined as the Region of Interest (ROI) in breast cancer diagnosis and screening.

E. Calcification detection for abnormality quantification

Mammographic calcification is an important feature of invasive and *in situ* breast cancer [24], and usually represent as small bright pixel clusters scattered inside breast. However, calcifications are difficult to be detected because of the small size, or to be separated from surrounding high density tissues. A combined LAW's texture filter [38] was applied for the further segmentation of calcifications inside the breast, which had sliding windows as shown in Fig. 5.

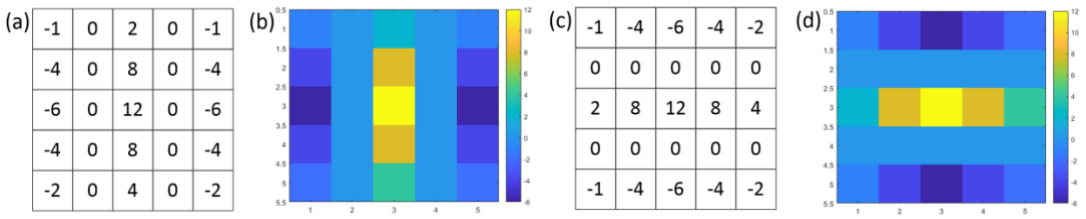

Fig. 5. Local texture filter for calcifications detection, in which (a) and (c) are filter windows along vertical and horizontal directions, and (b) and (d) are heat maps of filters respectively, where warmer colors represent higher weights in filtering.

During the filtering, the filter covered the neighborhood area of a certain pixel (i,j) and calculate mean value of product between pixel intensities and the filter weights, where v(i,j) is the filtered value at pixel (i,j), n is the number of element in the filter window from upper to lower, left to right, w_{nv} is the weight in the filter window in vertical direction and w_{nh} is in horizontal direction, and I_n is the intensity of the pixel at the *n*-th position of the filter.

$$v(i,j) = \sum_{n=1}^{25} (w_{nv} \times I_n + w_{nh} \times I_n) \Big/ 50 \tag{2}$$

In pratice, most of pixel clusters of calcifications have similar size to the 5×5 pixels Law's texture filter, which enhance the centralized pixels with high intensities inside the calcifications by higher weights and waken the surrounding pixels effects by lower weights. By filtering from both vertical and horizontal direction, the local small rounded maxima of pixels were enhanced while the surrounding breast tissues were not enhanced because of the combined filter's circle structure.

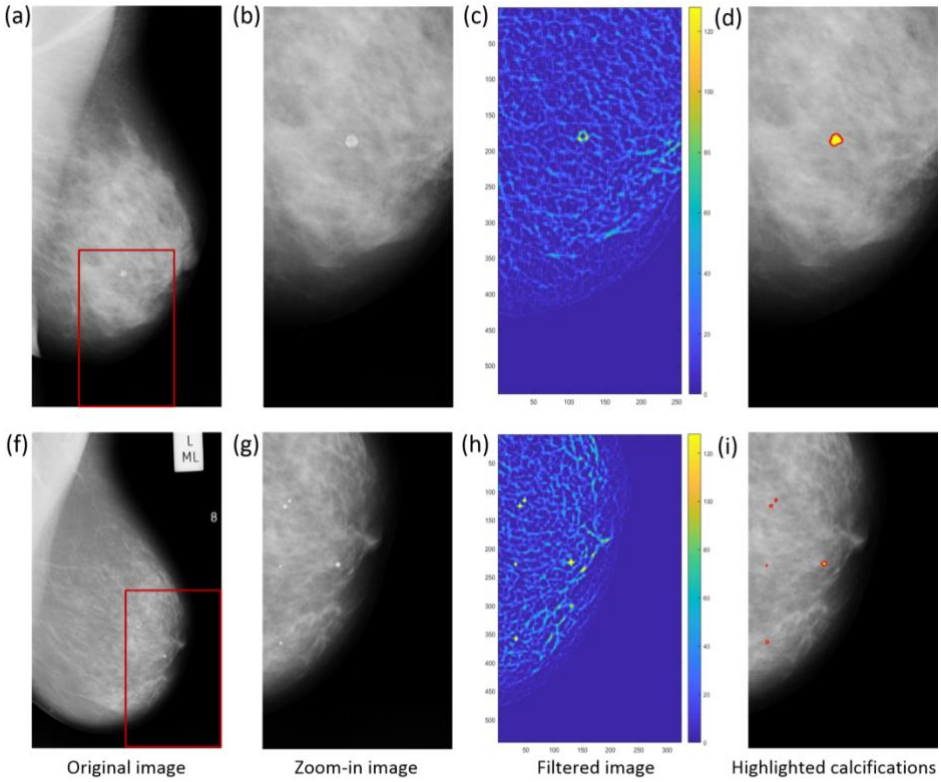

Fig. 6. Calcifications detection based on local texture filter, in which (a) and (f) are original images with calcifications, (b) and (g) are zoom-in areas, (c) and (h) are heat maps of filtered images where calcifications have warmest colors, and (d) and (i) are images with calcifications highlighted with red edges respectively.

Significant contrasts between calcifications and the surrounding tissues could be seen after the combined filtering as shown in Fig. 6. Global maxima of the filtered images were then extracted as candidate calcifications, and the final highlighted calcifications were acquired by traversing the candidates with suitable areas (≦60 pixels) and eccentricities (≦0.8) according to the natural morphological features of calcifications from experiments.

III. Results

To validate the proposed algorithm, we used the dataset including 322 mammogram images in MIAS dataset v1.21, which has resolutions vary from 400×1080 to 1000×1300 in pixels. Our method was implemented in MATLAB and deployed on a PC with 3.0GHz CPU and 8G RAM. In practice, it took 2.28 seconds in average for this algorithm to process one image in the test. The low time cost of proposed pipeline is fully competent to meet the real-time requirements in clinical use, which also saved manpower and avoided subjective errors.

A. Breast boundary segmentation defining region of breast

In mammogram image analysis, regardless of the pectoral muscle, the ROI was focused on breast as the basis of further quantifications, such as the glandular and fat proportions. Noises, artifacts and pectoral muscle should be removed in sequence, and the region within the detected boundaries was validated in multiple ways to prove the consistency of the proposed segmentation method.

According to the nature of mammography, the breast region is defined as the area between skin-air

line and pectoral-breast line. Therefore, we evaluated the performance using the strategy that segmented Breast region (Br) comparing to Background (Bg) plus Pectoral Muscle region (Pm) (Br vs (Bg + Pm)), which measured the performance of the proposed method in separating the breast region only from the pectoral and background regions. Breast masks and labels provided in support information of MIAS database were taken as ground truth for comparison.

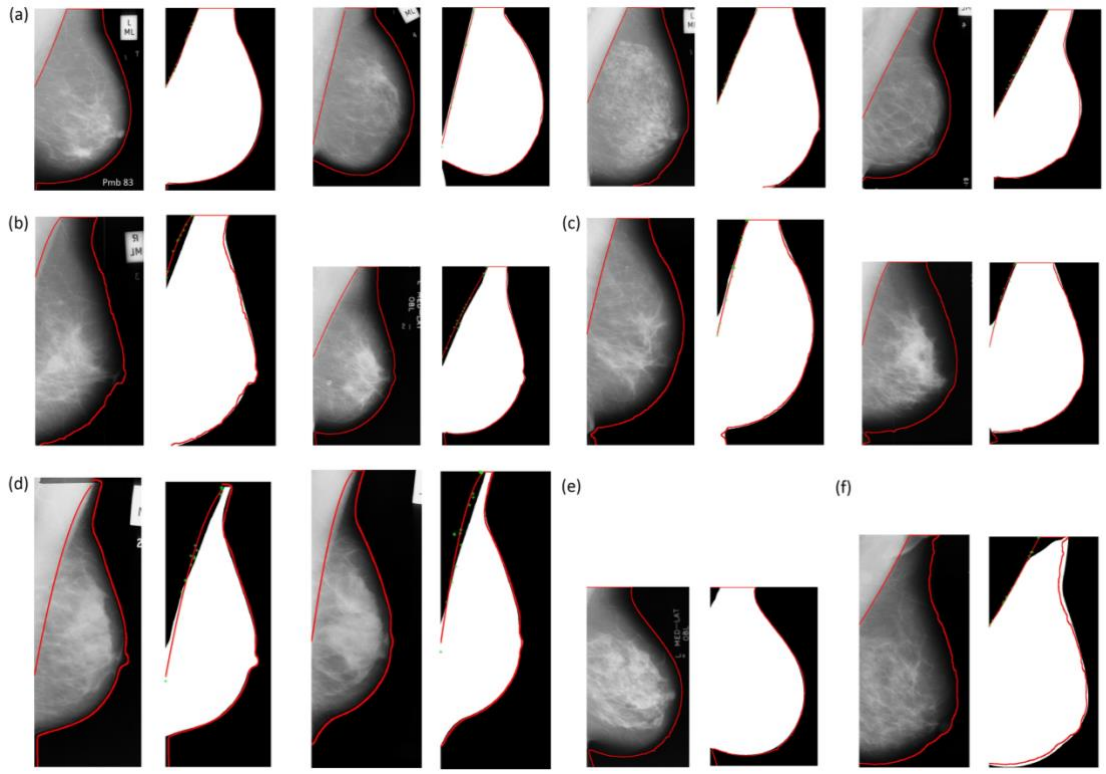

Fig. 7. Examples of breast segmentation results of SFM images in MIAS, in which the left of each pair includes breast boundaries marked on original images, and the right one contains boundaries marked on ground truth of breast region, where all the images are shown in the original ratios of height to width. (a) samples in the first row are results of good qualities images, (b) two samples with over-segmentation of breast, (c) two samples with under-segmentation of breast, (d) two samples with both over-segmentation and under-segmentation of breast, (e) one sample with no pectoral-breast boundary detected, and (f) one sample with top breast boundary misestimated.

As shown in Fig. 7, background, breast tissues and pectoral muscles have stratified intensity differences and the separation were performed well on good quality images. The skin-air boundaries in all tested images were detected quite similar to that of the ground truth. Meanwhile, almost straight lines fitted by the initial line segment dots shown as green stars in the image, and the trend of lines was along the natural boundaries as the ground truth showed. However, quite a number of images had various quality problems including high density artifacts and unclear boundary between pectoral muscle and dense breast. Irregular boundary lines were usually detected after pixel clustering on these areas, and line fitting algorithms are applied to get the smooth boundaries as illustrated respectively in Fig. 7(b), (c) and (d). Those problems generally included over-segmentation and under-segmentation of breast along the pectoral-breast boundaries, which reflected the minor errors occurred between different classes during the pixel-wise clustering. Besides, some image problems also caused segmentation errors in clustering. A very small area of pectoral muscle was

included in the category of breast in clustering and no pectoral-breast boundary was detected in Fig. 7(e). Fig. 7(f) shows an irregular top boundary of breast, which is an oblique line comparing to the horizontal straight lines in most of images.

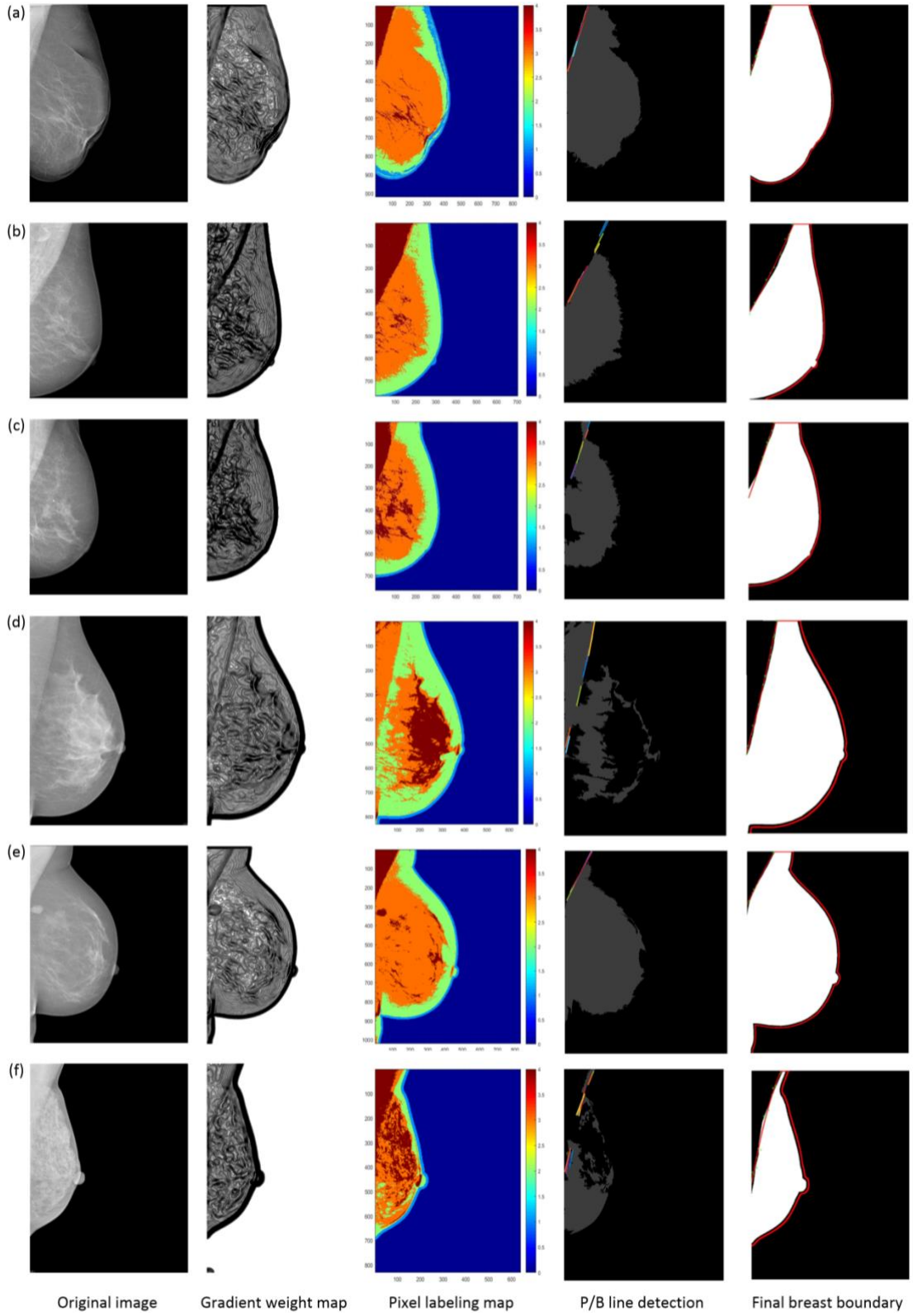

Fig. 8. Examples of breast segmentation process of FFDM images from BCDR and (d-f) INbreast databases, from left to right in each row there are the original image, gradient weight map, pixel

labeling map, P/B line detection map and the final breast boundary with ground truth, in which (a-c) and (d-f) are well segmented, under segmented and under segmented separately.

Besides SFM images, the proposed segmentation methods could be directly applied on FFDM images because both SFM and FFDM are gray-scale images with only difference of bit depth. The segmentation process of FFDM images shown in Fig. 8 is the same as that of SFM images, and the only difference could be found in detecting skin-air boundary based on gradient weight map, in which no artificial in FFDM images needs to be removed as shown in column 1 and 2.

In a comprehend evaluation strategy, four metrics were defined to generate the comprehensive accuracy of segmentation, including TP, TN, FP and FN are true positive, true negative, false positive and false negative respectively, in which the metrics were defined as follows, where Bs was the area of segmented breast based on the proposed method, Bt was the ground truth area of breast, and I represented the whole image.

$$\mathrm{TP} = (Bs \bigcap Bt)/I \tag{2}$$

$$\mathrm{TN} = [(I - Bs) \bigcap (I - Bt)]/I \tag{3}$$

$$\mathrm{FP} = [Bs \bigcap (I - Bt)]/I \tag{4}$$

$$\mathrm{FN} = [(I - Bs) \bigcap Bt]/I \tag{5}$$

All four metrics were combined for the accuracy calculation, in which

$$\mathrm{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Furthermore, Jaccard index [39] and Dice coefficient [40] were also considered to fully evaluate the performance of the proposed method, in which Jaccard index measured the similarity and diversity between the breast segmentation results and the ground truths, and Dice coefficient mainly represented the similarity between them.

$$\mathrm{Jaccard} = \frac{|Bs \bigcap Bt|}{|Bs \bigcup Bt|} \tag{7}$$

$$\mathrm{Dice} = 2\frac{|Bs \bigcup Bt|}{|Bs| + |Bt|} \tag{8}$$

Table 1. Quantitative results over 322 images from MIAS database, in which the accuracy is presented as percentages with standard deviations, and other metrics are presented as average values only.

	Well Seg.	Over Seg.	Under Seg.	Over + Under Seg.	No P/B boundary	Misestimated top breast boundary	Overall
Number of images (% percentage of images) in	227 (70.50)	39 (12.11)	31 (9.63)	16 (4.97)	3 (0.93)	6 (1.86)	322 (100)

each category							
TP	56.52%	55.39%	54.88%	53.06%	54.55%	54.97%	56.01%
TN	41.57%	39.18%	41.13%	39.92%	37.88%	38.60%	41.07%
FP	0.64%	3.86%	1.37%	4.26%	6.03%	4.89%	1.41%
FN	1.26%	1.54%	2.63%	2.84%	1.57%	1.51%	1.49%
Accuracy	98.09± 1.66%	94.57± 3.44%	96.01± 2.86%	92.88± 5.25%	92.43± 6.12%	93.57± 5.95%	97.08± 2.28%
Jaccard	96.22%	92.28%	93.79%	86.15%	92.13%	91.30%	94.89± 6.77%
Dice	97.3%	92.7%	96.3%	95.6%	91.5%	89.6%	96.36± 5.28%

In terms of separating the breast region from the pectoral and air regions, TP represented the overlapped areas of segmented breasts of the proposed method and the ground truth, and TN included the overlapping backgrounds of both methods. In practice, the standard deviations of TP and TN showing the big shape changes between different images in the dataset were high (≈10%). Those deviations have no direct relationships with the final accuracies of segmentation, which were only related to the precision of breast segmentation on each image. Therefore, standard deviations of TP, TN, FP and FN in each category of images were not listed in Table 1.

Generally, the proposed method performed well in segmentation of most breast areas from the background and pectoral muscle, while error rate of segmentation was composed of True Negative rate (TN=1.49%) and False Positive rate (FP=1.41%). In this case, TN means the missing parts of segmented breast, and FP usually includes parts of pectoral muscle or artifacts that are falsely labeled as breast.

In the first row of Table 1, slight higher FN than FP showed the skin-air line always a little smaller that the ground truth boundary, which made the elongated edge of breast wrongly classified as background by the gradient weight map. For over and under segmented cases, the pectoral-breast lines were fitted outside and inside the edges of ground truth respectively, showing the errors of polynomial curve fitting method caused by some imprecisely detected line segments along the pectoral muscles edges after pixel-wise clustering. Besides, in some cases with very small pectoral muscles projected in the images, the pectoral-breast lines might not be detected in the experiment, which made the FP rate (⩾6%) extremely higher than normal. Misestimated top breast boundary was an exception case in MIAS dataset, in which the top edges of breast in ground truths were not given as horizontal lines but in some inclined angles instead. The proportion of this unusual case is very low (⩽2%), and the increase brought to FP was small, so the proposed method was not designed to deal with this case.

Besides, the proposed method produced overall similarity ratios of Jaccard = 94.89% and Dice = 96.36%, which were slightly lower than the metric of segmentation accuracy showing those two metrics were more sensitive to shape changes between the segmentation results and ground truths than only calculating the overlapping area ratios. All Jaccard indexes decreased in not well segmentation cases in Table 1, and the accuracies of 'Over + Under segmentation' had the lowest Jaccard index because of the largest shape changes comparing to the ground truths. Since the union area of segmentation results and ground truths were relatively stable in different cases, Dice

coefficients didn't change much as that of Jaccard index, which suggested that the proposed method was robust and capable in most of image qualities.

B. Calcification detection for abnormality quantification

Other validations were focusing on detecting calcifications based on the proposed textural filter. There are 25 images were marked as 'CACL' in the ground truth of 3rd column in MIAS consensus, suggesting calcification dots found in the original images. Locations and sizes of those marked calcifications were listed afterward. However, not all the calcifications were listed in the illed-masses consensus of MIAS, so if the marked calcifications in ground truth could be found by our method, then the validation of calcification localization in one image was considered as success.

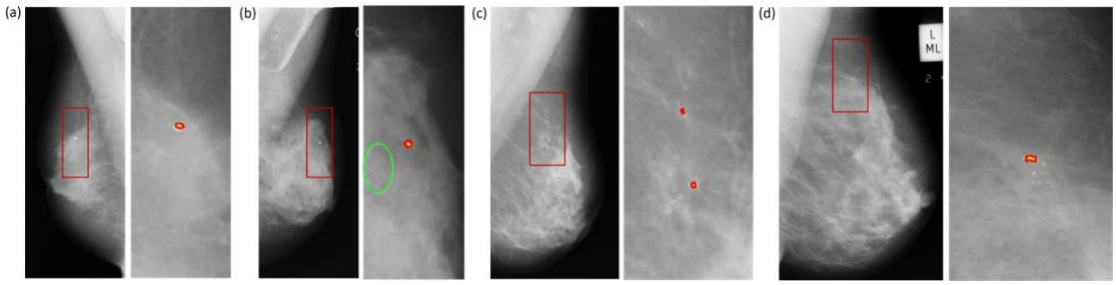

Fig. 9. Examples of calcification detection in MIAS, in which the left of each pair include original images, and the right one contains marked boundaries of detected calcifications in the zoom-in images, where all the images are shown in the original ratios of height to width.

The full list of experimental results comparing to the ground truth is listed in Table 2. Only 25 images in MIAS were marked as 'CALC', and in total 26 coordinates of centre and approximate radius (in pixels) of a circle enclosing abnormality were marked as ground truth in those 25 images [30]. In total 25 of 26 calcifications were detected by our method as shown in Table 2, which strongly suggested the effectiveness of the proposed texture filter in calcification detection. The only one missing detection happened in image numbered as 'mdb223ls' shown as Fig. 9(b), where the calcification inside the green circle was in very tiny size and the contrast was much lower comparing to the detected one besides it. Other calcifications with normal sizes and contrasts could be detected successfully for further morphological quantifications based the precise boundaries marked.

Table 2. Experimental results on calcification by the proposed method

CALC image number	Calcifications listed	Calcifications unlisted	Detected calcifications listed	Accuracy
25	26	5	25	96.15%

In total, 25 images in MIAS were marked as 'CALC', and 26 calcification with position were listed as ground truth in the introduction of dataset. As shown in Table 2, 25 detected calcifications in the experiment had exact positions as the list, with five more detected calcifications which were not listed in the introduction of dataset. Besides the images with calcification listed in MIAS, a batch of candidate calcifications were also detected from the high intensity pixel clusters with suitable sizes, which were not listed in the ground truth, but might also give the full reference for breast cancer diagnosis or screening as suspects.

IV. Discussions

Comparing to most of the existing mammogram image processing methods, the proposed method executes multiple tasks in a hierarchical way to acquire good performance in each step, which deals with specific morphologic characteristics of different objects for detection

A. **Qualitative and quantitative comparisons**

Table 3 shows a comparison with some of the existing studies in the literature, which involves the most common approaches on the segmentation of both skin-air and pectoral-breast boundaries.

Table 3. Mammographic breast segmentation approaches and corresponding accuracies, in which S/A is skin-air boundary and P/B is pectoral-breast boundary

Segmentation Approach	Detected Boundary	Method Proposed by	Dataset	Accuracy
Thresholding	S/A	Wei et al. [11]	DDSM	94.9%
	S/A	Raba et al.[12]	MIAS	98%
	S/A Pectoral-only	Czaplicka et al. [41]	MIAS (300 images)	98% 98%
Region Growing	P/B	Chen and Zwiggelaar [13]	MIAS	92.8%
	P/B	Maitra et al. [42]	MIAS	95.7%
Morphology-based	S/A	Wei et al.[11]	DDSM	94.9%
	S/A	Yapa and Harada [14]	MIAS (100 images)	99.1%
Active Contour	S/A	Ferrari et al. [15]	MIAS (84 images)	96%
	S/A	Martí et al.[16]	MIAS (65 images)	97%
	S/A	Wirth and Stapinski [17]	MIAS (25 images)	97%
Texture-based	P/B	Mirzaalian et al. [18]	MIAS (90 images)	N/A
	S/A	Casti et al. [19]	mini-MIAS and FFDM	99.6%
Pixel-wise Clustering	P/B	Ours	MIAS BCDR (100 images) INbreast (201 images)	97.08% 97.61% 97.38%

In comparison with other methods reviewed in [8, 22], the accuracies of breast segmentation are mainly divided into two parts, the skin-air line segmentation and the pectoral-breast segmentation. By summarizing performances of representative approaches from the literatures, methods only detecting skin-air (S/A) boundaries with pectoral muscle included archived the highest accuracy of 99.1% on selected images from MIAS dataset. In methods further detecting pectoral-breast boundaries (P/B), Maitra et al. [42] archived 95.7% of segmentation accuracy based on region growing strategy. Some studies only focus on breast pectoral segmentation and did not separate the breast boundary from the air background. For example, accuracies the pectoral-only detection of 98% was proposed by Czaplicka et al. [41], and the S/A boundary detection accuracy is also proposed as 98% in the same paper, so its combined accuracy should be 98%×98%≈96% as the highest overall accuracy

from the literatures.

As shown in both Table 1 and Table 3, the segmentation accuracy of our method is over 97% along with high similarity rates of Jaccard (94.89%) and Dice (96.36%) indexes in MIAS database. These results on proper breast segmentation provide solid foundation for further researches with focus on the region of breast. In the meantime, our algorithm achieved a little higher segmentation accuracy on FFDM databases because no more artificial such as tapes or labels in FFDM, which improved the accuracies on S/A line detection than SFM.

B. Furture improvements

Breast segmentation is the fundamental steps in the study of mammography, which defines the boundaries of the region for further quantifications. The performance of breast segmentation was highly determined by pixel-wise clustering and line fitting and smoothing methods. Segmentation errors could be further avoided by improving these two core algorithms. Within the image processing pipeline, both *k*-means clustering and line segment detecting and fitting methods are flexible functional modules and could be further improved without changing the general pipeline. The more precise boundaries of breast and inside glandular detected, the better view of mammographic diagnosis could be obtained.

The Breast Imaging Reporting and Data System (BI-RADS) [43], which defines diagnostic categories, defines a six-point scale from negative to highly suggestive of malignancy based on mammography. A set of indicators such as glandular density, calcifications and specific masses have been proved to be related to the risk of breast cancer [44, 45]. The morphological features characterizing those indicators, as derived from results of image segmentation and calcification detection, may conduct preliminary screening for doctors to save manpower and avoids subjective errors.

By processing images and generating effective indicators in mammographic analysis, this study can also been applied to other breast imaging modalities such as MRI. Based on the robustness of clustering image segmentation methods, further improvements may focus on enriching detection indicators of detailed structures of illed-masses. The pipeline will be integrated into a web released software for public use, and the clinical application of the proposed framework will be helpful for mammography reading in breast cancer diagnosis and screening.

V. Conclusions

In mammography based breast cancer diagnosis and screening, correct segmentation and labeling of different tissues are important image processing steps, which are needed, e.g., before the morphology of labeled tissues is quantified for further analysis. In this paper, we have presented an image processing pipeline to analyze mammogram images, including a breast and pectoral muscle segmentation method and a calcification detection method. Skin-air boundary is firstly detected based on gradient weight map, and pectoral-breast boundary is then estimated by clustering pixels in the foreground into different layers. Next, breast segmentation is performed in the pixel-wise classification to get the initial boundary between pectoral muscles and breast region based on line detection and polynomial curve fitting. Finally, calcification detection is performed by a well-defined texture filter as one of the biomarkers for breast cancer risk. Experimental results showed that the proposed pipeline produced comparable results as manual breast outline segmentation and calcification detection but has higher efficiency, which could also directly applied on FFDM images

generated in clinic with few adjustments.

Generally two main advantages made the proposed methods capable to find the breast boundaries with high accuracy. First, gradient weight map clearly defines skin-air boundary between foreground and background, which is more efficient than calculating an adaptive threshold in traditional intensity histogram-based methods. Second, since pectoral muscles are not always clearly visible in mammograms, especially in dense breasts, the unsupervised learning approach of kmeans clustering took pixel as the study object, overcame the inherent defect of low-quality image of the traditional method based on geometrical morphology, and no areas needed to be selected by human intervals before the segmentation. Furthermore, our method is adaptable in dealing with images with unusual pectoral muscle shapes against the breast by clustering pixels without taking shape models into account. A well designed model of pectoral muscle region [22] may be helpful in P/B boundary detection, which can be used as a solid foundation for future research on pectoral muscle segmentation of our work.

Beside the effectiveness in breast, The pipeline has high speed and accuracy to be fully competent to meet the real-time requirements in clinical use mainly because that the pectoral muscle is estimated in the same step of pixel-wise cluster through pectoral-breast segmentation, in which the proposed algorithm converted the segmentation task into a classification task. This simplified the calculation and improved the efficiency of the process. Furthermore, the texture filter considers local correlations of single pixels within a suitable neighborhood region. Therefore, it can detect calcifications efficiently even with small noises or inside large masses, which could provide valuable support for mammogram analysis in clinic.

Acknowledgement

This work is a joint research project undertaken by Fujian Normal University and Ulster University. The authors would like to thank every laboratory member involved in this cooperation. The project is financially sponsored by the National Science Foundation of China (61501121), the Chinese Scholarship Council of the Ministry of Education ([2016]3035), the Returned Overseas Chinese Scholars, State Education Ministry ((2015)1098), and Fujian Provincial Department of Science and Technology (2015J05145 and 2016Y0018).

Author Contributions

P.S. and H.W carried out the conception and design, revised the manuscript, and approved the final version to be published. J.Z. participated in the conception and design, and analysis and interpretation of the data. A.R. carried out the part of computerized methods design and experiments. All authors read and approved the final manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

Ethics and consent statements: The study was performed on the public released database with no personal information provided.

REFERENCES

1. DeSantis, C., Ma, J., Bryan, L., et al. (2014). Breast cancer statistics, 2013. CA: A cancer journal for clinicians, **64**(1), 52-62.
2. Oeffinger, K C., Fontham, E. T. H., Etzioni, R., et al. (2015). Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA, **314**(15), 1599-1614.
3. Jalalian, A., Mashohor, S.B.T., Mahmud, H.R., et al. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clinical Imaging, **37**(3), 420-426.
4. Chen, B., and Ning, R. (2002). Cone - beam volume CT breast imaging: Feasibility study. Medical Physics, **29**(5), 755-770.
5. Mann, R.M., Kuhl, C.K., Kinkel, K., et al. (2008). Breast MRI: guidelines from the European Society of Breast Imaging. European Radiology **18**(7): 1307-1318.
6. Olsen, O., and Gøtzsche, P.C. (2001). Cochrane review on screening for breast cancer with mammography. The Lancet, **358**(9290), 1340-1342.
7. Johns, P.C., and Yaffe, M.J. (1987). X-ray characterisation of normal and neoplastic breast tissues. Physics in Medicine and Biology, **32**(6), 675-695.
8. Mustra, M., Grgic, M., and Rangayyan R.M. (2016). Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. Medical & Biological Engineering & Computing, **54**(7): 1003-1024.
9. Dromain, C., Boyer B., Ferre, R., et al. (2013). Computed-aided diagnosis (CAD) in the detection of breast cancer. European Journal of Radiology **82**(3): 417-423.
10. Oliver, A., Freixenet, J., Marti, J., et al. (2010). A review of automatic mass detection and segmentation in mammograic images. Medical Image Analysis, **14**(2), 87-110.
11. Wei, K., Wang, G., and Ding, H. (2006). Segmentation of the breast region in mammograms using watershed transformation. 27th annual international conference of the Engineering in Medicine and Biology Society. IEEE-EMBS, 6500-6503.
12. Raba, D., Oliver, A., Marti, J., et al. (2005). Breast segmentation with pectoral muscle suppression on digital mammograms. Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), **3523**, 471-478.
13. Chen, Z., and Zwiggelaar, R. (2012). A combined method for automatic identification of the breast boundary in mammograms. 5th international conference on biomedical engineering and informatics (BMEI). IEEE, 121-125.
14. Yapa, R.D., and Harada, K. (2008). Breast skin-line estimation and breast segmentation in mammograms using fast-marching method. International Journal of Biological, Biomedical and Medical Sciences, **3**(1), 54-62.
15. Ferrari, R.J., Rangayyan, R.M., Desautels J.E.L., et al. (2004). Identification of the breast boundary in mammograms using active contour models. Medical and Biological Engineering Computing, **42**(2), 201-208.
16. Martí, R., Oliver, A., Raba, D., et al. (2007). Breast skin-line segmentation using contour growing. Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), **4478**, 564-571.
17. Wirth, M.A., and Stapinski, A. (2003). Segmentation of the breast region in mammograms using active contours. Visual communications and image processing. International Society for Optics and Photonics, Lugano, 1995-2006.
18. Mirzaalian, H., Ahmadzadeh, M.R., Sadri, S. (2007). Pectoral muscle segmentation on digital mammograms by nonlinear diffusion filtering. IEEE Pacific Rim conference on communications,

computers and signal processing, 2007. IEEE, 581-584.

19. Casti, P., Mencattini, A., Salmeri. M., et al. (2013). Estimation of the breast skin-line in mammograms using multidirectional Gabor filters. Computers in Biology Medicine, **43**(11), 1870-1881.
20. Adel, M., Rasigni, M., Bourennane, S. et al. (2007). Statistical segmentation of regions of interest on a mammograic image. EURASIP J Advances in Signal Processing, **1**, 1-8.
21. Wang, L., Zhu, M.L., Deng, L.P., et al. (2010). Automatic pectoral muscle boundary detection in mammograms based on Markov chain and active contour model. Journal of Zhejiang University Science C, **11**(2), 111-118.
22. Rampun, A., Morrow, P. J., Scotney, B. W., et al. (2017). Fully automated breast boundary and pectoral muscle segmentation in mammograms. Artificial Intelligence in Medicine, **79**, 28-41.
23. McCormack, V.A. and dos Santos Silva, I. (2006). Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis. Cancer Epidemiology and Prevention Biomarkers. **15**(6), 1159-1169.
24. Evans, A. (ed.). (2002). Breast calcification: a diagnostic manual. Cambridge University Press.
25. Davis, D.H., and Dance, D.R. (1990). Automatic computer detection of clustered calcifications in digital mammograms. Physics in Medicine and Biology, **35**(8), 1111.
26. Strickland, R.N., and Hahn, H.I. (1996). Wavelet transforms for detecting microcalcifications in mammograms. IEEE Transactions on Medical Imaging, **15**(2), 218-229.
27. Yu, S., and Guan, L. (2000). A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. IEEE Transactions on Medical Imaging, **19**(2), 115-126.
28. Nagaiah, K., Manjunathachari, K., & Rajinikanth, T. V. (2015). Efficient Image Enhancement Techniques for Micro calcification Detection in Mammography. IJEEE, 7(02).
29. Hofvind, S., Hofvind, S., Skaane, P., et al. (2014). Mammographic Performance in a Population-based Screening Program: Before, during, and after the Transition from Screen-Film to Full-Field Digital Mammography. Radiology, **272**(1), 52-62.
30. Suckling, J., Parker, J., Dance, D., et al. (2015). Mammograic image Analysis Society (MIAS) database v1.21.
31. López, M. G. A., de Posada, N. G, Moura. D. C. et al. (2012). BCDR: A Breast Cancer Digital Repository. International Conference on Experimental Mechanics, 1065-1066.
32. Moreira, I. C., Amaral, I., Domingues, I., et al. (2012). INbreast: toward a full-field digital mammographic database. Academic Radiology, 19(2), 236-248.
33. Gonzalez, R. C., Woods, R.E., and Eddins, S.L. (2009). Digital Image Processing Using MATLAB: and Mathworks, MATLAB Sim SV 07, Gatesmark Publishing.
34. Lowell, J., Hunter, A., Steel, D., et al. (2004). Measurement of retinal vessel widths from fundus images based on 2-D modeling. IEEE Transactions on Medical Imaging, **23**(10), 1196-1204.
35. Sun, T.N., and Neurvo, Y. (1994). Detail-preserving median based filters in image processing. Pattern Recognition Letters, **15**(4), 341-347.
36. Fernandes, L.A.F., and Oliveira, M.M. (2008). Real-time line detection through an improved Hough transform voting scheme. Pattern recognition, **41**(1),99-314.
37. Lancaster, P., Salkauskas, K. (1986). Curve and surface fitting: an introduction. Academic press.
38. Tencer, L., Reznakova, M., and Cheriet, M. (2012). A new framework for online sketch-based image retrieval in web environment. 11th International Conference on Information Science,

Signal Processing and their Applications (ISSPA), IEEE, 1430-1431.
39. Tan, P.N., Steinbach, M., and Kumar, V. (2005). Introduction to data mining. 1st ed.
40. Dice, L.R. (1945). Measures of the amount of ecologic association between species. Ecology, **26**(3), 297-302.
41. Czaplicka, K., and Wodarczyk, H. (2011). Automatic breast-line and pectoral muscle segmentation. Schedae Informaticae, **20**, 195-209.
42. Maitra, I.K., Nag, S., and Bandyopadhyay, S.K. (2012). Technique for preprocessing of digital mammogram. Computer Methods Programs Biomedicine, **107**(2), 175-188.
43. American College of Radiology: Breast Imaging Reporting and Data System (BI-RADS) Atlas. (2007). Reston, VA, American College of Radiology.
44. Boyd, N.F., Guo, H., Martin, L.J., et al. (2007). Mammographic density and the risk and detection of breast cancer. New England Journal of Medicine, **356**(3), 227-236.
45. Vachon, C.M., Van Gils, C.H., Sellers, T.A., et al. (2007). Mammographic density, breast cancer risk and risk prediction. Breast Cancer Research, **9**(6), 217-226.