Applied Medical Informatics
Vol. 32, No. 2 /2013, pp: 1-13

A Preprocessing Algorithm for the CAD System of Mammograms Using the Active Contour Method

Farhan AKRAM[1], Jeong Heon KIM[2], Inteck WHOANG[3], and Kwang Nam CHOI [1,*]

[1] Department of Computer Science & Engineering, Chung-Ang University, Seoul, Korea
[2] Korea Institute of Science and Technology Information, Daejeon, Korea
[3] The Attached Institute of ETRI, Daejeon, Korea
E-mail(s): farhan@vim.cau.ac.kr; jhkim@vim.cau.ac.kr; inteck@wm.cau.ac.kr; knchoi@cau.ac.kr (*)

* Author to whom correspondence should be addressed; Tel.: +822-820-5316; Fax: +822-825-5304

Received: 10 April 2013 / Accepted: 20 May 2013 / Published online: 6 June 2013

Abstract
Digital mammogram images can have different kinds of artifacts that affect the accuracy of the detection of tumor tissues in the automated computer-aided detection (CAD) system for mammograms. Preprocessing to remove such artifacts is an important step. In this paper, a preprocessing technique for digital mammograms is devised which removes labels, scanning artifacts and the pectoral muscle. First, it removes hurdles like labels, scanning and taping artifacts using an automated algorithm based on thresholding. Then, using the active contours and the proposed stopping algorithm it obtains the contour which contains the boundary of the pectoral muscle. Later, it extracts the pectoral muscle binary image from the contour. Finally, using the pectoral muscle binary image and the original mammogram image it obtains the desired image without any artifacts and the pectoral muscle. We tested the proposed algorithm on the mammograms from the mini-MIAS database and it worked very efficiently. It provided very effective and accurate results for pectoral muscle segmentation. It provided up to 97.84% accuracy, computed from well segmented results.

Keywords: Active contours; Computer-aided detection; Mammogram preprocessing; Pectoral muscle segmentation; Region of interest.

Introduction

Of all cancers in women, breast cancer has the highest incidence rate. It has been reported that about one in eight women will develop breast cancer during their lifetime [1]. Breast cancer takes time to develop usually years. It is divided into 0–4 stages depending on the size of the cancer, whether the cancer is invasive or noninvasive, whether or not it is in the lymph nodes, and whether or not the cancer has spread to other parts of the body beyond the breast. Survival rates are often used by doctors as a standard way of discussing a person's prognosis. The more advanced the stage of cancer, the lower the survival rate. Stage 0 is used to describe non-invasive breast cancer, such as ductal carcinoma in situ (DCIS), in which there is no evidence of cancer cells or non-cancerous abnormal cells having developed into a cancer tumor. The five year survival rate of stage 0 breast cancer is 93%. A stage 1, tumor measures from 0.02 mm to 2cm and will not have spread outside the breast but does involve the lymph nodes. Stage 1 is further divided into two sub-stages 1-A and 1-B, with a five year survival rate of 88%. In stage 2, cancerous cells or tumors no larger than 5cm are formed in the auxiliary lymph nodes, or a tumor larger than 5cm is found outside the lymph

nodes. This stage is also divided into two sub-stages 2-A and 2B, with five year survival rates of 81% and 74%, respectively. In stage 3, either a tumor larger than 5 cm with positive lymph nodes or a tumor with skin and chest wall involvement is present. The surgical intervention performed is relatively dramatic: partial or total breast removal and lymph nodes dissections may be needed. This stage is divided into three sub-stages, i.e., 3-A, 3-B and 3-C with five year survival rates of 67%, 41%, and 49%, respectively. In stage 4, there are obvious metastases to other organs of the body, most often the bones, lungs, liver, or brain, and the five year survival rate drops to 15% [2-3]. It has been shown that early detection and treatment of breast cancer are the most effective methods of reducing the associated mortality rates [4].

X-ray mammography is the most effective image modality for the early stage detection of breast cancer, and thus mammograms are conducted regularly in breast screening [5]. The mammograms results are then examined by expert radiologists. The number of mammograms showing cancer is very small; however, the necessity of thoroughly examining a large number of mammograms in order to detect a small number of cancers can cause high false positives, which can lead to unnecessary biopsies, and some cancers are missed due to radiologist fatigue or distractions [6-7].

A CAD system that prompts suspicious regions and draws the attention of the radiologist to them is highly desirable. Such a system in the diagnosis of breast cancer would potentially provide a consistent and reproducible second opinion to a radiologist, whose primary role is to localize suspicious regions and characterize them as either benign or malignant. It can help radiologists by tapering their search to regions of interest. Use of the proposed CAD system can increase the efficiency of the correct detection of cancer and decrease the number of false positives determined by radiologists [8-9]. Using the CAD system, the mammogram is segmented into several distinct regions including the breast boundary, the nipple, masses, dense glandular regions, and the pectoral muscle. Boundary detection and segmentation of the pectoral muscle is useful in determining mammogram competence and restricting the search space for calcification and lesion detection [10].

Digital mammogram images can have different kind of artifacts, such as high and low intensity labels, taping artifacts, scanning artifacts and the pectoral muscle, as shown in figure 1. The inclusion of the artifacts and the pectoral muscle in the image data being processed could bias the detection procedures [11] and affect the accuracy of the detection of regions of interest in intensity based image processing. Thus, preprocessing plays a vital role in an automated CAD system for mammograms.

In this paper, a preprocessing algorithm for the CAD system, which removes label, scanning, and taping artifacts and segments the pectoral muscle boundary, is proposed. It is divided into three steps: the first step removes the labeling and scanning artifacts; the second step detects the contour to obtain the boundary of the pectoral muscle; and the third step segments the pectoral muscle and removes it from the original mammogram image.

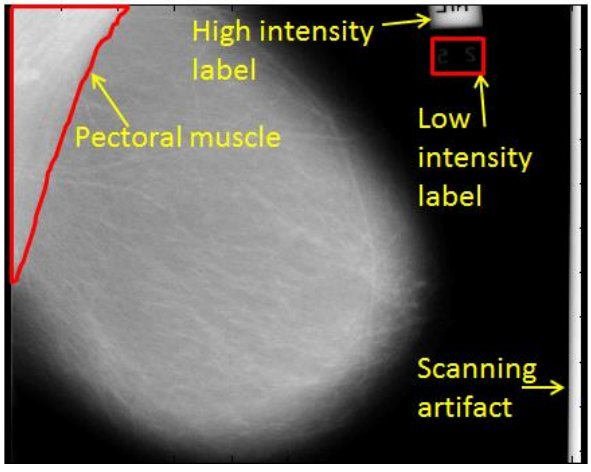

Figure 1. Mammogram image with artifacts and pectoral muscle

In the first part of the proposed algorithm, labels and scanning artifacts are removed using thresholding. First, it obtains a binary image from the original mammogram, and then it keeps the breast boundary only and removes all other objects in the image. Finally using both the binary and original images, it obtains an image without labels or other artifacts. In the second part, a multiphase active contour model is used to extract the desired contour, effectively isolating and removing the pectoral muscle region. It acquires the desired contour using Chan Vese's (CV) [12] active contour model and proposed stopping algorithm. In this model, a new term, M^k is introduced to obtain the contour at each time step. This newly introduced M^k term converts the CV's two phase model into a multiphase model. It uses the 2^n region approach for segmenting an object in an image. It first divides an image into two regions, then discards the outer region and again divides the inner region into two regions and so on, until the stopping condition is fulfilled, or one of the regions is empty. If n is the total number of time steps, then n+1 will be the total number of regions that can be computed with this multiphase model. The stopping algorithm uses the current and old M^k value to stop the contour at a certain point. In the third part, it extracts the pectoral muscle data from the binary image M^k. Then using the pectoral muscle binary image and the original image obtained after the first step, it provides the final image.

We tested the proposed technique on all images in the mini-MIAS database. It provided very effective and accurate results compared to the old techniques. But the problem with the proposed technique is that it takes too much time for high-intensity mammogram images as the contour moves quite slowly on the high-intensity region.

Related Work

Numerous approaches have been proposed to segment the breast boundary profile and the pectoral muscle in mammograms. Some of these have focused on using thresholding [13-14], gradients [15], modeling of the non-breast region of a mammogram using polynomial [16], or active contours [17].

The pectoral muscle segmentation algorithm by Ferari et al [18] which is based on the Hough transform is one of the most frequently used techniques. The main problem with this approach is that the pectoral muscle is approximated by a line. Therefore, these methods give poor results when the pectoral muscle contour is a curve. To solve this problem, another method [19] was introduced based on Gabor wavelets. In [20], the pectoral muscle was once again approximated by a straight line which was further attuned through surface smoothing and edge detection.

Ma et al. [21] described two image segmentation methods: one based on adaptive pyramids and other based on minimum spanning trees. The article [22] chose the longest straight line in the Radon-domain as an approximation to the pectoral muscle localization. There are two problems with this work: the simplification of using a line and the use of a private database. Therefore, the results cannot be compared with other publications. Camilus and co-workers [23] used a graph-cut method, followed by Bezier curve smoothing. Hong et al. [24] proposed an isocontour map based technique which computes the contour map for the given image and extracts the region of interest using a contour tree. A discrete time Markov chain and an active contour model were adopted in [25] for pectoral muscle boundary detection, while in [26] the pectoral muscle was detected using support vector machines (SVMs). Liu et al. [27] proposed a pectoral muscle segmentation algorithm based on the positional characteristics of the pectoral muscle in the breast region to combine the iterative Otsu thresholding scheme and mathematical morphological processing to find a rough border of the pectoral muscle. After finding a rough border, it applies a multiple regression analysis to obtain an accurate segmentation of the pectoral muscle. Finally, Kim et al. [28] proposed a method for adaptive contour map using active contour. They used a region scalable contour map technique in which scale of contour map can be adjusted from coarse scale to fine scale using constants of the force terms inside and outside of the contour. Their method is effective in drawing different scale contour map for the topographic analysis of the structure of objects in an image. Furthermore, that contour map can be used in constructing a contour tree which can be used for

the intensity based depth analysis of different regions of interests in medical imaging. Those regions of interests may refer to tumour tissues, blood rupture, fatty and glandular tissues, etc.

In the second step of this paper we used similar method developed by the Kim et al. in order to obtain the desired contour. Our algorithm is region-based algorithm which does not have the edge indicator function in the length term unlike Kim et al. algorithm which makes the proposed algorithm faster than Kim et al. algorithm.

A Preprocessing Algorithm for the CAD System of Mammograms Using the Active Contour Method

Removing Labels and Scanning Artifacts

In the first step of the proposed algorithm, labels and scanning artifacts are removed using thresholding. First, it takes a mammogram image and applies thresholding to acquire a binary image, shown in figure. 2(b), using the threshold value T=15. Then it divides the image vertically in half, as shown in figure 2(c), to attain two groups. Thereafter, it computes the pixel sum for both groups and compares these to check which one is greater so as to decide the starting point of the pointer. If

$$\sum_{i=0}^{row}\sum_{j=0}^{(col/2)-1} I_{i,j} > \sum_{i=0}^{row}\sum_{j=col/2}^{col} I_{i,j}$$

then the pointer should be on the left corner, otherwise it should be on the right corner.

In the above algorithm, row is the maximum number of rows, col is the maximum number of columns, and $I_{i,j}$ is the binary image. After finding the starting point it traverses the pointer in every row and removes the unnecessary data in the image, i.e., labels and scanning artifacts, as shown in figure 2(e). After removing the unnecessary data, it multiplies the binary mask image with the original mammogram image to obtain the desired result, shown in figure 2(f).

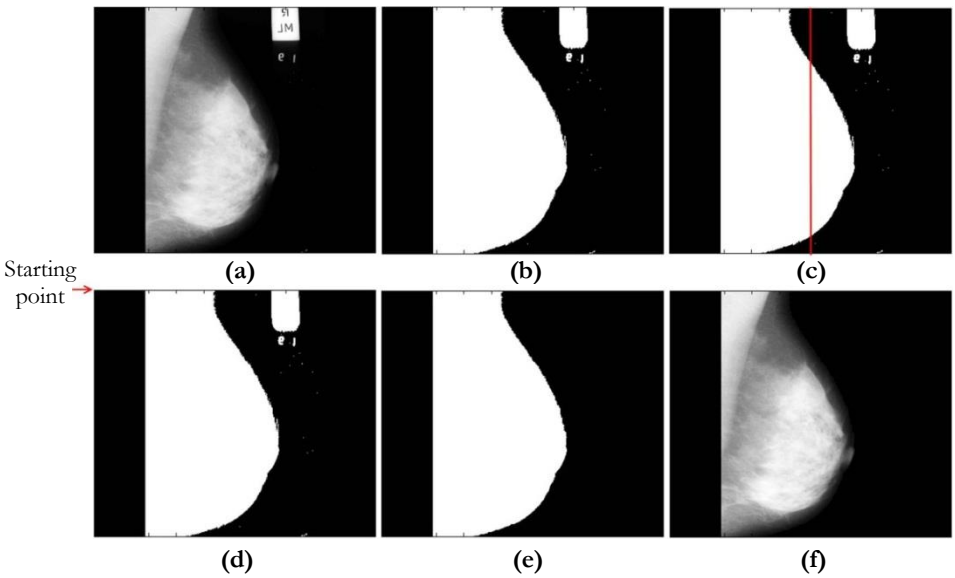

Figure 2. Step 1: (a) Original mammogram; (b) Binary image; (c) Dividing the binary image in half; (d) The binary image with the starting point of the pointer (e) The binary image after removing the labels and artifacts; (f) The mammogram image after removing the labels and artifacts.

Getting Desired Contour

In the second step, it acquires the contour, which contains the pectoral muscle boundary information using the active contour method. Chan and Vese [12] proposed an active contour method based on the Mumford-Shah model [29]. Let $I:\Omega \to R$ be an input image and C be a closed curve. For that, we have defined a new region-based energy functional, E_{CV} by using the CV model as shown below:

$$E_{CV}(c_1,c_2,\phi) = \lambda_1 \int_{\Omega} |I(x)-c_1|^2 H_\varepsilon(\phi(x)) M^k(x)dx + \lambda_2 \int_{\Omega} |I(x)-c_2|^2 (1-H_\varepsilon(\phi(x))) M^k(x)dx + \mu \int_{\Omega} |\nabla H_\varepsilon(\phi(x))|dx + \nu \int_{\Omega} H_\varepsilon(\phi(x))\,dx \qquad \text{where } x \in \Omega \tag{1}$$

where $\mu \geq 0$, $\nu \geq 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$ are fixed parameters. $H_\varepsilon(\phi)$ is the Heaviside function and M^k is the newly introduced mask term, which helps to move the contour inwards shown in (2) and (3) respectively. The M^k term is also used for computing the difference between two consecutive contours in order to determine the stopping point of the contour. It also converts the CV two-phase active contour model into a multiphase active contour model.

$$H_\varepsilon(\phi) = \frac{1}{2}\left(1+\frac{2}{\pi}\arctan\left(\frac{\phi}{\varepsilon}\right)\right) \tag{2}$$

and

$$\begin{cases} M^k(x) = \phi > 0, \\ M^0 : \Omega \to 1. \end{cases} \tag{3}$$

For more accurate computation involving the level set function and its evolution, we need to regularize the level set function by penalizing its deviation from a signed distance function [30]. It eliminates the need for re-initialization of the level set function. The energy functional of the penalizing term and expression for the total energy is shown in (4) and (5) respectively.

$$P(\phi) = \alpha \int_{\Omega} \frac{1}{2}\left(|\nabla\phi|-1\right)^2 dx \tag{4}$$

and

$$E_{Total}(\phi) = E_{CV}(\phi) + P(\phi) \tag{5}$$

c_1 and c_2 shown in (7) and (8) approximate the image intensities inside and outside contour C, respectively. Where ϕ is the level set function initialized by ϕ_0:

$$\phi_0 = \begin{cases} -\rho & x \in \Omega_0 - \partial\Omega_0 \\ 0 & x \in \partial\Omega_0 \\ \rho & x \in \Omega - \Omega_0 \end{cases} \tag{6}$$

Minimizing the total energy functional in (5) by using the steepest descent method [31] and representing the contour C with a zero level set, i.e., $C = \{x \in \Omega \mid \phi(x) = 0\}$, we obtain the variational formulation shown in (7), (8), and (9) respectively.

$$c_1 = \frac{\int_{\Omega} I(x)H(\phi)M^k(x)dx}{\int_{\Omega} H(\phi)M^k(x)dx} \tag{7}$$

$$c_2 = \frac{\int_{\Omega} I(x)(1-H(\phi))M^k(x)dx}{\int_{\Omega}(1-H(\phi))M^k(x)dx} \tag{8}$$

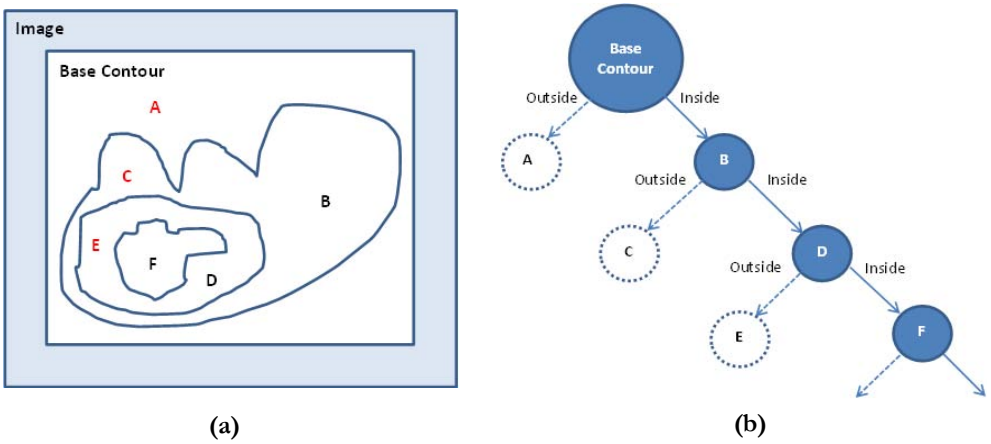

Figure 3. Multiphase segmentation: (a) Calculated isocontour map from a given image; (b) Inclusion tree from (a).

and

$$Q(t) = \delta_{\varepsilon}(\phi)\left[\mu div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \lambda_1\left(I-c_1\right)^2 M^k + \lambda_2\left(I-c_2\right)^2 M^k - v\right] + \alpha\left[\Delta\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right] \tag{9}$$

where Δ is the Laplacian operator. Therefore, the function ϕ that minimizes this functional satisfies the Euler Lagrange equation $\partial E_{Total} / \partial\phi = 0$. The last term in (9) is the regularization term obtained by solving (4) using the gradient descent method. The data fitting term $-\lambda_1\left(I-c_1\right)^2 M^k + \lambda_2\left(I-c_2\right)^2 M^k$, plays a key role in the curve evolution, and λ_1 and λ_2 govern the tradeoff between the first and the second terms. In most cases, we set $\lambda_1 = \lambda_2$ and $v = 0$ and μ as the scaling parameters, and $\delta_{\varepsilon}(\phi) = H'_{\varepsilon}(\phi)$ is the univariate Dirac function. A classical iterative process for minimization of the function $E_{Total}(\phi)$ is the following gradient flow with an artificial time *t*:

$$\begin{cases} \phi_{(t=0)} = \phi_0, \\ \dfrac{\partial\phi}{\partial t} = Q(t) \end{cases} \tag{10}$$

This algorithm uses the 2^n region segmentation approach for segmentation of an object. It divides the image into two sub-regions and then takes the inner of those two regions and divides it further into two sub-regions and so on, until the inner region is empty, or the stopping condition is fulfilled. If n is the number of iterations in order to obtain the final result, then n+1 would be the total number of regions. Figure 3 shows the working of multiphase segmentation using the active contour method. In figure 3(a), the contour map is shown from the base contour to the final contour. The inner regions obtained from the contour map of the segmentation process are displayed by colored nodes followed by the solid lines in figure 3(b). The inner and outer regions' relationship is as follows:

$$\begin{cases} Base\ contour = A + B & \Rightarrow A = Base\ contour - B, \\ B = C + D & \Rightarrow C = B - D, \\ D = E + F & \Rightarrow E = D - F \end{cases}$$

The contour extracting part of the proposed model discards the outer contour and uses the current inner contour for the next step to segment it again into two parts. To stop the contour at a certain point using the stopping algorithm, check the similarity of the pixels in the two consecutive contours. If

$$\sum_{i=0}^{row}\sum_{j=0}^{col} old\ M_{i,j}^{k} - \sum_{i=0}^{row}\sum_{j=0}^{col} M_{i,j}^{k} = stopping\ value$$

then the contour will stop moving any further. Where $0.99 < stopping\ value < 1$, $old\ M^k$ is the mask term of the last computed contour, M^k is the mask term of the current computed contour, row is the maximum number of rows, and col is the maximum number of columns in the original mammogram image. Figure 4 shows the computed contour, which contains the pectoral muscle boundary.

Removing Pectoral Muscle

In the third step, the algorithm removes the pectoral muscle using the M^k value, which was computed in the previous step. M^k provides the binary image of the last computed contour. Using that binary image, it extracts the binary image of the pectoral muscle. First, it finds the starting point for the pointer movement using the same rule used in figure 2(c). After finding the starting point, it keeps the object connected to the starting point and removes the remaining region, considering the information; pectoral muscle is always connected to the starting point. Figure 5 shows how the pectoral muscle is removed from the mammogram.

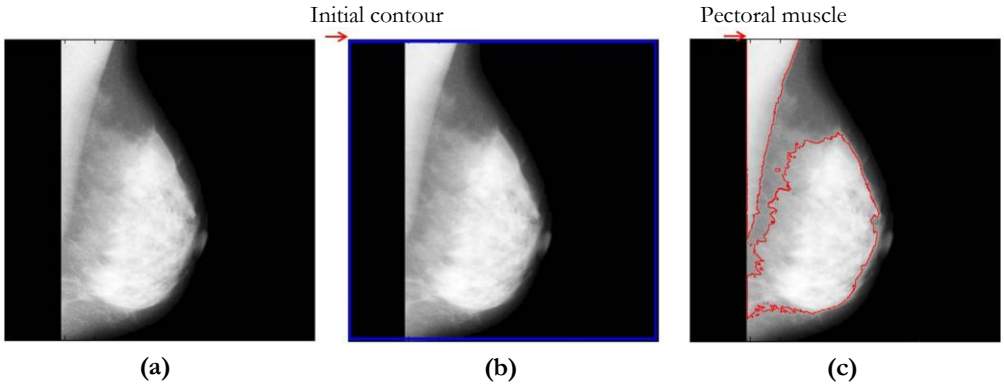

Figure 4. Step 2: (a) Original mammogram image after step 1; (b) Mammogram image with the initial contour; (c) Mammogram image with the contour containing the pectoral muscle boundary, and the contour stopped using the stopping algorithm.

Comparison with other Techniques

We compared our segmentation algorithm with the techniques of Ferrari et al. [19] and Liu et al. [27] and we found our technique to be more accurate in terms of segmentation results. Unlike the other two techniques, the proposed algorithm also removes the unwanted artifacts in the preprocessing stage (in Section removing labels and scanning artifacts).

Figure 6 shows a comparison between the visual results of our technique and methods of Ferrari et al. and Liu et al. Row 1 shows the original mammogram images from the mini-MIAS

database [32] numbering mdb003, mdb005, mdb028, mdb044, mdb095, and mdb123, respectively. Row 2 shows the ground truths of the pectoral muscle regions. Row 3 shows the segmentation results of the pectoral muscle with the method employed by Ferrari et al. Row 4 shows the segmentation results of the pectoral muscle with the algorithm originated by Liu et al. Row 5 shows the segmentation results with our proposed algorithm. The contour stopping value used in our algorithm for mdb003, mdb028, and mdb123 is 0.999, and for mdb005, mdb044, and mdb095, it is 0.9995. The other parameters are the same as given in Section 5.

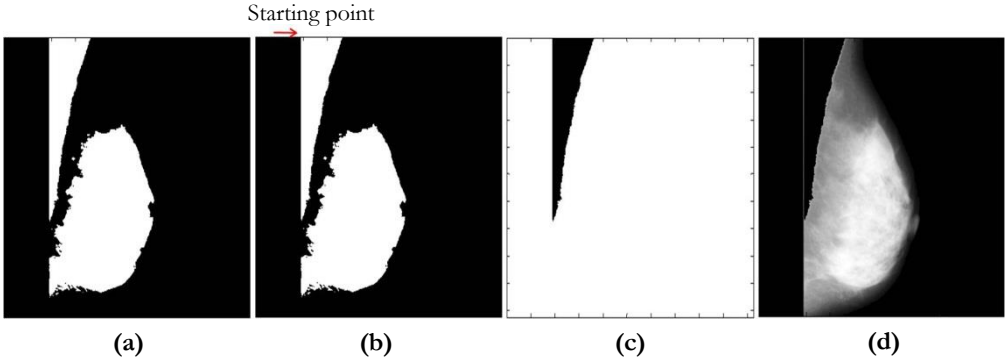

Figure 5. Step 3: (a) The M^k mask image from step 2; (b) The starting point of the pointer for the M^k mask image; (c) The binary image for the pectoral muscle; (d) The mammogram image after removing the pectoral muscle.

If we compare the pectoral muscle segmentation results of mammograms mdb003, mdb005, mdb028, mdb044, and mdb095, we can see that the segmentation results using our algorithm are much closer to the pectoral muscle ground truths compared to the results generated by the techniques of Ferrari et al. and Liu et al. In the Ferrari et al. technique, the pectoral muscle is segmented by a linear straight line approximation; therefore, it always misses the curved pectoral muscle region. The Liu et al. method uses the Otsu thresholding scheme and mathematical morphological processing to find the rough border of the pectoral muscle. But neither approach is as accurate as our technique. By comparing the results, we can see that the Ferrari et al. technique always misses the curved boundary of the pectoral muscle when segmenting it. The results therefore always have an accuracy problem because of the inclusion of the false positive and true negative region with the linear straight line segmentation. In the segmentation results of mdb003, we can see that the Liu et al. algorithm missed some parts of the pectoral muscle, while our technique segmented it well; thereby indicating that it was closer to the ground truth. The segmentation results of mdb005 using our technique and the Liu et al. technique look almost similar. In the results of mdb028, a small hump on the pectoral muscle is missed by the other techniques; however, our algorithm segmented it well. The segmentation results of mdb044 using our technique and the Liu et al. technique look almost similar. In the segmentation results of mdb095, small details of the pectoral muscle boundary are missed by the other techniques; however, our algorithm segmented it well.

The segmentation results of mammogram mdb123 show an accuracy problem for all of the algorithms. The main reason for the accuracy problem with our algorithm is because the region of interest contains a sub-region, which acts like a new region due to the intensity difference between that sub-region and the remaining region. In our algorithm, the contour computing step uses the multiphase active contours method in which the contour is initialized closer to the image boundary. It finds the desired contour by dividing the image into two regions and then discards the outer contour and divides the inner contour further into two regions and so on, until the stopping condition is fulfilled. In mdb123, instead of stopping at the ground truth, the contour moved inward because of the clear intensity difference in the pectoral muscle region. The stopping condition was fulfilled for the sub-region of the pectoral muscle instead of the whole pectoral muscle region. Therefore, it can be said that, if the pectoral muscle has a sub-region with a

prominent intensity difference or a sub-region with a prominent boundary with different intensities, our algorithm may display segmentation problems.

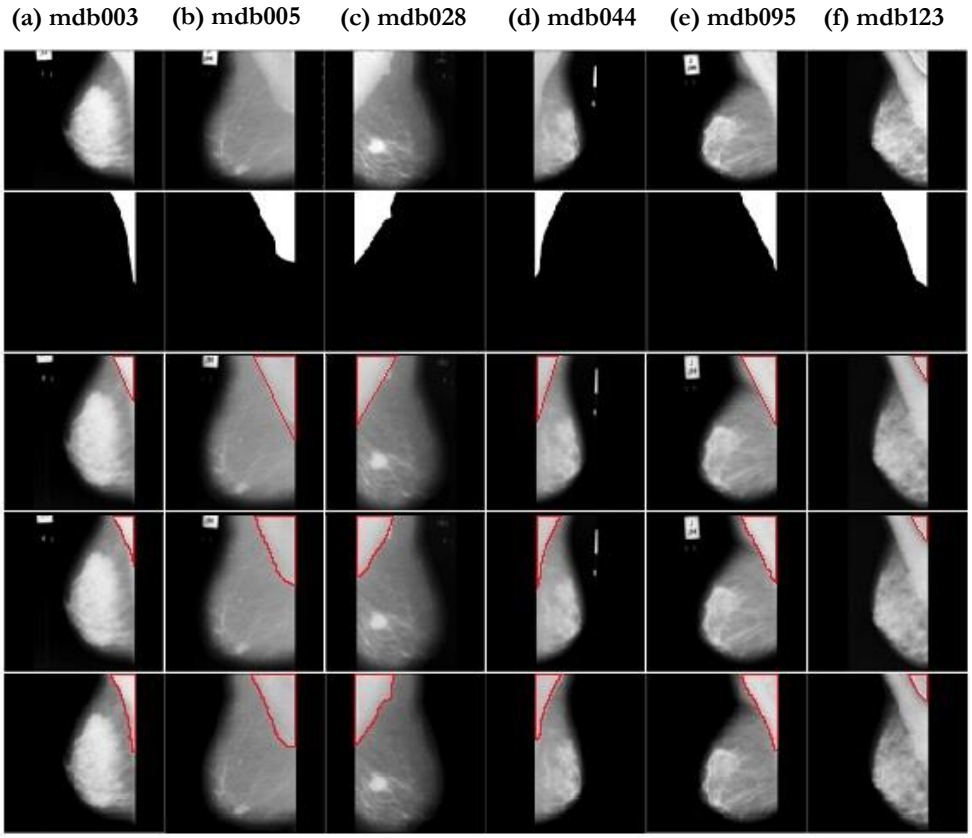

Figure 6. The visual results obtained by using Ferrari et al., Liu et al., and the proposed algorithm on six pectoral muscles from the mini-MIAS database. The first row shows the original mammogram images; the second row shows the ground truth; the third row shows the results obtained by the method employed by Ferrari et al.; the fourth row shows the results obtained from the approach of Liu et al.; and the Last row shows the result of the proposed method.

Results and Discussion

The proposed algorithm was applied to the mammogram images from the mini-MIAS database [32]. The range of intensity in all of the images is represented from 0 to 255, and the images each have a size of 1024x1024 pixels. For the active contour method in step 2 we used $\rho = 1$, $v = 0$, $\lambda_1 = \lambda_2 = 10$, $\mu = 5$, $\varepsilon = 1.5$, $\alpha = 0.1$, $stopping\ value = 0.9995$, and the time step $\Delta t = 0.1$. Figure 7 and figure 8 show the segmentation results using our algorithm. We tested the proposed algorithm on all images from the mini-MIAS database and ascertained that there are two problems affecting the accuracy and efficiency of the algorithm. The accuracy of the preprocessing algorithm is affected if the intensity of the pectoral muscle and the rest of the breast region are almost similar, or the intensity of the pectoral muscle is lower than the remaining breast region. Figure 7 shows some of the results with false positives and false negatives. The efficiency of the algorithm is affected by the mammograms with high intensity smooth regions, as the contour moves very slowly in the high intensity regions. For the rest of the images, which do not have the problems described above, our algorithm gives very accurate segmentation results of the pectoral muscle as shown in figure. 8.

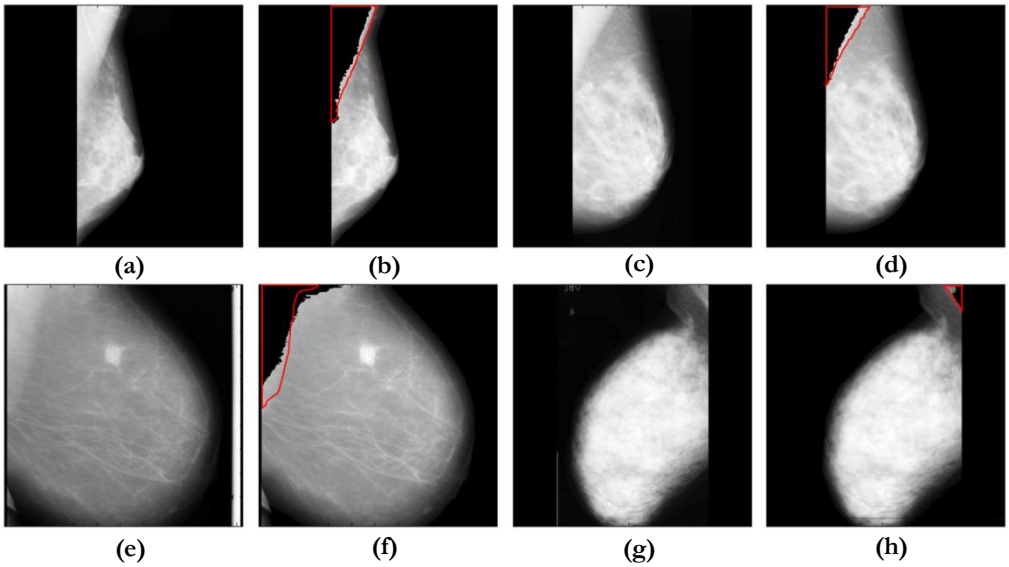

Figure 7. Some of the preprocessing results using the proposed algorithm showing the accuracy problem in the segmentation of the pectoral muscle. The ground truth contour of the actual pectoral muscle boundary is shown in red. **(a)**, **(c)**, **(e)** and **(g)** are images mdb018, mdb030, mdb134, and mdb253, respectively, from the mini-MIAS database.

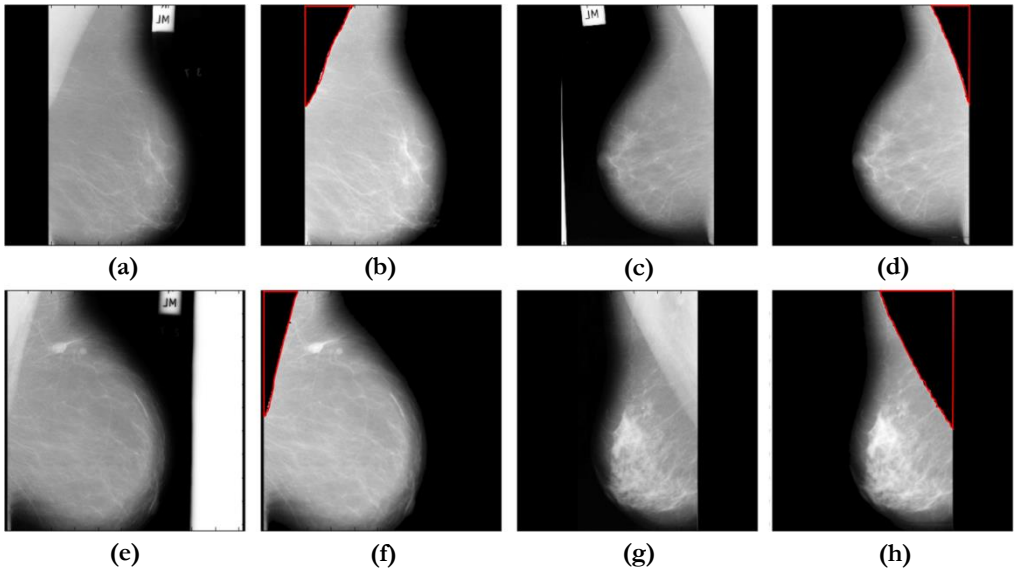

Figure 8. Some of the successful preprocessing results obtained using the proposed algorithm. The ground truth contour of the actual pectoral muscle boundary is shown in red. **(a)**, **(c)**, **(e)** and **(g)** are images mdb006, mdb009, mdb132, and mdb249, respectively, from the mini-MIAS database.

The percentage accuracy of our algorithm in Table 1 and Table 2 is calculated using the following expression:

$$\frac{A \cap B}{A \cup B} \times 100\%$$

where A is the ground truth of the pectoral muscle region, and B is the pectoral muscle region computed by our algorithm. The average percentage accuracy computed from Table 1 and Table 2 is 77.10% and 97.84%, respectively. The accuracy of the proposed algorithm is dependent on the selection of the stopping value of the contour. For the image with high-intensity regions, and the image with the intensity of pectoral muscle closer to the intensity of the remaining breast region,

the stopping value should be closer to 1. However, for the image with a fair intensity difference between the pectoral muscle region and the remaining breast region, the stopping value should be closer to 0.99.

Table 1. Percentage accuracy of our algorithm computed for figure 7

Mammogram No	Accuracy Percentage (%)
018	86.99
030	84.79
134	71.94
253	64.68

Table 2. Percentage accuracy of our algorithm computed for figure 8

Mammogram No	Accuracy Percentage (%)
006	97.70
009	98.25
132	97.35
249	98.07

Conclusion

The edge detection and region segmentation algorithms work more efficiently when the images are preprocessed. Most of such techniques are dependent on the intensity in the image. The unwanted regions with high intensity can affect the segmentation result of the

desired region of interest. Therefore, it is necessary to remove such unwanted regions with high intensity, which can affect the results. In this paper, a preprocessing algorithm which removes undesirable objects, such as labels, scanning artifacts, and the pectoral muscle, from the mammogram image is proposed.

The proposed preprocessing algorithm works in three steps. In the first step, it removes labels and scanning artifacts using thresholding. In the second step, it acquires the contour using the multiphase active contour method, which contains the pectoral muscle boundary information. In the third step, using the previously obtained contour, it extracts the pectoral muscle region and then acquires a final mammogram image without undesired objects.

We tested the proposed preprocessing algorithm on all of the images from the mini-MIAS database. The results appeared to be very effective and accurate compared to the old techniques. The algorithm provided very accurate segmentation results of the pectoral muscle by extracting the pectoral muscle region from the computed contour. The accuracy percentages of our algorithm for bad and accurate preprocessing results are 77.10% and 97.84%, respectively.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgement

This research was supported by the Chung-Ang University Research Grant in 2009.

References

1. Jinn, N B C S P W (2005). NHS Breast Screening Programme. [Online] Available at: www.cancerscreening. nhs.uk/breastscreen.
2. Breastcancer.org (2012) Stages of breast cancer. [Online] Available at: http://www.breastcancer.org/symptoms/diagnosis/staging.
3. American Cancer Society (2012) Breast cancer survival rates by stages. [Online] Available at: http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-survival-by-stage
4. Brady M, Highnam RP. Mammographic Image Analysis. Dodrecht: Kluwer Academic Publishers; 1999.
5. McLelland R. Screening for breast cancer: Opportunity, status and challenges. Recent Results Cancer Res 1990; 119:29-38.
6. Forrest AP, Aitken RJ. Mammography screening for breast cancer. Annu Rev Med 1990; 41:117-132.
7. Astley SM, Boggis CRM, Walker K, Wallace S, Tomkinson S, Hillier V, Morris J. An evaluation of a commercial prompting system in a busy screening center. Proc. 6th Int. Workshop Digital Mammography 2002; 471–475.
8. Astley S.M, Zwiggelaar R, Wolstenholme C, Taylor CJ. Prompting in mammography: How accurate must prompt generators be?. Proc. 4th Int. Workshop Digital Mammography 1998:347-354.
9. Huo Z, Giger ML, Vyborny CJ. Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis. IEEE Transactions on Medical Imaging 2001;20:1285-1292.
10. Chandrasekhar R, Attikiouzel Y. Mammogram-attribute database: A tool for mammogram segmentation and analysis. Proc. IASTED Int. Conf. Signal Processing, Pattern Recognition and Applications 2001:143-148.
11. Homer MJ. Mammographic Interpretation: A Practical Approach. Boston, MA: McGraw-Hill; 1997.
12. Chan TF, Vese LA. Active contours without edges. IEEE Transactions on Image Processing 2001;10(2):266-277.
13. Bick U, Giger ML, Schmidt RA, Nishikawa RM, Wolverton DE, Doi K. Automated Segmentation of Digitized Mammograms. Academic Radiology 1995;2(2):1-9.
14. Yin F-F, Giger ML, Doi K, Metz CE, Vyborny CJ, Schmidt RA. Computerized Detection of Masses in Digital Mammograms: Analysis of Bilateral Subtraction Images. Medical Physics 1991:18(5):955-963.
15. Mendez J, Tahoces PJ, Lado MJ, Souto M, Correa JL, Vidal JJ. Automatic Detection of Breast Border and Nipple in Digital Mammograms. Computer Methods and Programs in Biomedicine 1996;49:253-262.
16. Chandrasekhar R, Attikiouzel Y. Automatic Breast Border Segmentation by Background Modeling and Subtraction. Proc. 5th Int. Workshop Digital Mammography (IWDM), Toronto, Canada: Medical Physics Publishing 2000:560-565.
17. Wirth MA, Stapinski A. Segmentation of the breast region in mammograms using active contours. Visual Communications and Image Processing 2003;5150:1995-2006.
18. Ferrari RJ, Rangayyan RM, Desautels JE, Borges RA, Fr`ere AF. Segmentation of mammograms: identification of the skin-air boundary, pectoral muscle, and fibro-glandular disc. Proc. 5th Int. Workshop Digital Mammography 2000:573-579.
19. Ferrari RJ, Rangayyan RM, Desautels JE, Borges RA, Frere AF. Automatic identification of the pectoral muscle in mammograms. IEEE Transactions on Medical Imaging 2004;23(2):232-245.
20. Kwok SM, Chandrasekhar R, Attikiouzel Y. Automatic pectoral muscle segmentation on mammograms by straight line estimation and cliff detection. IEEE Transactions on Medical Imaging 2004;3(9):1129-1140.
21. Ma F, Bajger M, Slavotinek JP, Bottema MJ. Two graph theory based methods for identifying the pectoral muscle in mammograms. Pattern Recognition 2007;40(9):2592-2602.

22. Kinoshita SK, Azevedo-Marques P, Pereira RR, Rodrigues JA, Rangayyan RM. Radon-domain detection of the nipple and the pectoral muscle in mammograms. Journal Digital Imaging 2008;21(1):37-49.
23. Camilus KS, Govindan VK, Sathidevi PS. Computeraided identification of the pectoral muscle in digitized mammograms. Journal Digital Imaging 2009;23(5):562-580.
24. Hong BW, Sohn BS. Segmentation of regions of interest in mammograms in a topographic approach. IEEE Transactions on Information Technology in Biomedicine 2010;14(1):129-139.
25. Wang L, Zhu M, Deng LP, Yuan X. Automatic pectoral muscle boundary detection in mammograms based on markov chain and active contour model. Journal of Zhejiang University Science C 2010;11(2):111-118.
26. Domingues I, Cardoso JS, Amaral I, Moreira I, Passarinho P, Comba JS, Correia R, Cardoso MJ. Pectoral muscle detection in mammograms based on the shortest path with endpoints learnt by SVMs. Proc. 32nd Annual Int. Conf. IEEE EMBS, Buenos Aires, Argentina 2010:3158-3161.
27. Liu CC, Tsai CY, Liu J, Yu CY, Yu SS. A pectoral muscle segmentation algorithm for digital mammograms using Otsu thresholding and multiple regression analysis. Computer and Mathematics with Applications 2012;64(5):100-1107.
28. Kim JH, Park B-Y, Akram F, Hong B-W, Choi KN. Multipass Active Contours for an Adaptive Contour Map. Sensors 2013;13:3724-3738.
29. Mumford D, Shah J. Optimal approximation by piecewise smooth function and associated variational problems. Communication on Pure and Applied Mathematics 1989;42(5):577-685.
30. Chunming L, Chenyang X, Changfeng G, Fox MD. Level set evolution without re-initialization: A new variational formulation. Proc.s IEEE Conf. Computer Vision and Pattern Recognition, San Diego 2005:430-436.
31. Aubert G, Kornprobst P. Mathematical Problems in Image Processing: Partial Differential Equations and Calculus of Variations. New York: Springer; 2002.
32. Pilot European Image Processing Archive (2003) The mini-MIAS database of mammograms. [Online] Available at: http://peipa.essex.ac.uk/info/mias.html.