

Segmenting Mammography to Detect Breast Cancer with Employing a Multilayer Level Set Approach

Yishuo Huang *

Department of Construction Engineering, Chaoyang University of Technology, Wufeng, Taiwan

Abstract: X-ray mammography provides a non-invasive way of detecting breast cancer early. Detecting breast cancer from mammograms is difficult due to the properties of the image. Mammogram images are images with high-resolution and large size. The low levels of contrast between the cancer cells and normal cells make the automatic and accurate detection of cancer cells difficult. This paper employs two algorithms to solve the above difficulties. Firstly, the given 16-bit mammograms are subjected to a shrinking process to reduce their image sizes. Subsequently, the 16-bit mammograms are converted into 8-bit images employing a pixel-depth conversion algorithm. Without losing generality, the process can significantly reduce the heavy computational burden. Secondly, the multilayer level set approach is applied to the shrunken mammograms to segment a mammogram into several sub-regions, such that each sub-region is homogeneous. The multilayer level set approach has been inspired by the problems proposed by Mumford and Shah. Mumford and Shah proposed the division of an image into a set of homogeneous sub-regions, such that the energy contained in the image can be minimized. Based on this energy minimization, the multilayer level set method implicitly depicts the regional boundaries as several nested level lines. With an increase in the number of iterations and preselected level values, these lines evolve closer to the level boundaries based on the energy minimization criterion. In this paper, the digital database for screening mammography (DDSM), provided by the University of South Florida, is used to evaluate the performance of the proposed algorithms. The experimental results demonstrate that the generated optimal-piecewise-constant approximation can provide the desired segmentation approximation for the given mammogram, and the extracted regional boundaries can effectively assist a physician in locating the breast cancer area.

Keywords: Trapezoidal fuzzy number; bi-criteria fixed charge transportation problem; linear ranking function.

1. Introduction

Breast cancer is the most common cancer found in women, especially in the developed countries [1]. X-ray mammography is a breast cancer screening technique that provides a non-invasive way of detecting breast cancer early. Mammo-

graphy, along with a physical breast examination, is the main choice for screening breast cancer. It is one of the best examination techniques for the early detection of breast cancer. The interpretation of X-ray mammograms requires the skill

* Corresponding author; e-mail: yishuo@cyut.edu.tw
© 2012 Chaoyang University of Technology, ISSN 1727-2394

Received 29 November 2011
Revised 8 January 2012
Accepted 11 January 2012

and experience of a physician. Computer-aided diagnosis (CAD) schemes have been proposed and applied to identify the regions composed of potential microcalcification clusters in mammograms [2]-[4]. Recently, CAD has become an important tool and assists a physician to determine whether or not a woman has breast cancer based on the mammograms.

The detection of microcalcification clusters provides one of the radiographic indications of breast cancer as it has been found that microcalcification clusters are present in 30-50% breast cancer mammographics [5]. Several researchers have proposed different schemes to improve the diagnostic accuracy. For example, Nakayama et al used the filter bank based on the Hessian matrix to decompose the given mammograms, and subsequently, the Bayes discriminant function was employed to detect the abnormal regions of interest within the microcalcification clusters [6]. Selvi et al applied the complex wavelet transform to enable sub-band feature extraction from the given mammograms. The extracted features were used as the training data in the supported vector machine (SVM) to classify the mammograms with microcalcification clusters into abnormal regions of interest [3]. In summary, locating the abnormal regions of interest in the given digital mammogram is a preliminary step in CAD.

Bruce et al employed the Chan-Vese level set approach to enable mammographic mass core segmentation. The algorithm is based on curve evolution and the level set method. The idea of an evolving curve can be traced back to the late 1980s. The concept of an evolving curve is that an initial curve will automatically move to the regional boundaries according to the principle of minimum energy. In 1988, Kass and Witkin proposed the snake theory where the extraction of regional boundaries using the iteration approach is based on the

minimization of the energy contained in the image [7]. There are existing limitations of the snake algorithm: the topology of the evolving curve tends to be affected during evolution and the segmentation procedures depend on parameterization [8].

In this study, the given digital mammograms are segmented into several sub-regions employing active curve evolutions via level sets, such that the pixel values in each sub-region are approximated by a regional constant. The regional constant is defined as the average value of a segmented sub-region. Mumford and Shah have formulated the segmentation problem in computer vision as follows: given an image u_0 , find a set Ω which is composed of the limited sub-regions and an optimal piecewise smooth approximation u , such that the pixel values in the approximation u vary smoothly in the sub-region Ω_i , and discontinuously across the sub-regional boundaries [9]. In order to solve this problem, Chung and Vese proposed a new method (multilayer level set approach), which models the island dynamics for epitaxial growth to obtain an improved curve evolution algorithm for image segmentation [10]. Island dynamics describe the normal velocity of island boundaries to mathematically simulate the epitaxial growth such that the evolution boundaries can form the so-called layer-by-layer growth regime [11]. For examples, let $\phi: \Omega \rightarrow \mathbf{R}$ be a level set function and a first layer of islands can be defined by the region $\{x: \phi(x) \geq 0\}$ and the corresponding boundaries can be defined as $\{x: \phi(x) = 0\}$; then, a second layer of islands which lies on the first one is shown to be $\{x: \phi(x) = 1\}$ and the boundaries are given by $\{x: \phi(x) = 1\}$. The level set approach, initially introduced by Osher and Sethian [12], is applied to implement the island dynamics for epitaxial growth as this approach can automatically

handle topological changes. The aim of this study was to analyze the digitized mammograms using the multilayer level set approach in order to segment the mammograms. In the segmented mammograms, the regional boundaries of the suspicious regions can be extracted to characterize the underlying texture of the breast regions.

The remainder of this paper is organized as follows: In the next section, the modified multilayer level set approach is derived for the given mammograms. Section III illustrates the segmentation results obtained from the application of the multilayer level set approach to the given mammograms. Section IV compares the segmentation results with the circled results provided by physicians to enable a discussion on the proposed approaches. In conclusion, a few comments on the experimental results have been made in Section IV.

2. Multilayer level set model

To find the optimal piecewise-constant approximation of a given image for the minimal partition problems, Mumford and Shah proposed to minimize the energy function defined as follows:

$$E(u, C) = \sum_i \int_{\Omega_i} (u_0 - c_i)^2 dx dy + \nu |C| \quad (1)$$

For a fixed C (let C be a closed subset in Ψ , and made up of a finite set of smooth curves), the energy defined in Equation (1) is minimized for the variables c_i by setting c_i as the mean value for the specified sub-region Ω_i [13]. The vector level set function $\Phi = (\phi_1, \dots, \phi_m)$ and vector Heaviside function $\mathbf{H}(\Phi) = (H(\phi_1), \dots, H(\phi_m))$ were introduced by Vese [9]. The vector level set function forms the edges illustrated

in the segmented image and the vector Heaviside function, whose basic components are 0 and 1, can classify the given image such that the following equation

$$\{I(x, y) | \mathbf{H}(\Phi(I)) = \mathbf{c} \in \mathbf{H}(\Phi(\Omega))\} \quad (2)$$

can be satisfied. In (2), the \mathbf{c} is a constant vector. Furthermore, Vese observed that when employing two level set functions (ϕ_1 and ϕ_2), the given image can be completely segmented into a series of sub-regions, and the boundaries of the segmented sub-regions are formed at $\{\phi_1 = 0\} \cup \{\phi_2 = 0\}$ [9]. The multilayer level set approach proposed by Chung employs the multiphase level set for image segmentation and island dynamics for epitaxial growth [10].

The simulation of epitaxial growth in material science is difficult because of the time and length scales involved. Ratsch et al developed an island dynamics model to describe epitaxial growth, which refers to the deposition of a material on top of a substrate, and the model is implemented using the level set method to simulate thin film growth [14]. In the multilayer approach, let \mathbf{I} be a X-ray mammogram, and the two level set functions partition a given image into sub-regions with distinct levels $\{l_1 < l_2 < \dots < l_m\}$ and $\{k_1 < k_2 < \dots < k_n\}$ for the level set functions, ϕ_1 and ϕ_2 , respectively. The relationships between the sub-regions in the approximation and the functions, ϕ_1 and ϕ_2 , are defined as follows:

$$\Omega_{ij} = \{x \in \Omega : l_i \leq \phi_1(x) \leq l_{i+1}, k_j \leq \phi_2(x) \leq k_{j+1}\} \quad (3)$$

Extending (1), the energy function for a given X-ray mammogram is defined as follows:

$$E(c, \phi) = \sum_{i,j=0}^{m,n} |I - c_{i,j}|^2 \prod_i^{i+1} H((-1)^p(\phi_1 - l_i)) \prod_j^{j+1} H((-1)^q(\phi_2 - k_j)) dx + \mu \left[\sum_{i=1}^m \int_{\Omega} |\nabla H(\phi_1(x) - l_i)| dx + \sum_{j=1}^n \int_{\Omega} |\nabla H(\phi_2(x) - k_j)| dx \right] \tag{4}$$

where $p = 0$ or 1 , $q = 0$ or 1 , H is the Heaviside function, $\mu > 0$ and is a weight parameter, and c_{ij} is the sub-regional constant [10]. Although the Heaviside function has a different form, this paper employs the form proposed by Vese, and is defined as [9]:

$$H(z)_\varepsilon = \frac{1}{2} \left[1 + \frac{2}{\pi} \tan^{-1} \left(\frac{z}{\varepsilon} \right) \right] \tag{5}$$

and

$$H(z)_\varepsilon = \frac{1}{2} \left[1 + \frac{2}{\pi} \tan^{-1} \left(\frac{z}{\varepsilon} \right) \right] \tag{6}$$

where ε is a parameter which determines the value of the function such that the function has a different zero when z is located in the interval, $[-\varepsilon, \varepsilon]$; else, the function is equal to zero. The sub-regional constant, c_{ij} , is defined as follows [10]

$$c_{ij} = \frac{\int_{\Omega} I H(\phi_1 - l_i) H(l_{i+1} - \phi_1) H(\phi_2 - k_j) H(k_{j+1} - \phi_2) dx}{\int_{\Omega} H(\phi_1 - l_i) H(l_{i+1} - \phi_1) H(\phi_2 - k_j) H(k_{j+1} - \phi_2) dx} \tag{7}$$

The optimal piecewise-constant approximation derived from Mumford and Shah can be shown as:

$$\Omega = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} c_{ij} H(\phi_1 - l_i) H(l_{i+1} - \phi_1) H(\phi_2 - k_j) H(k_{j+1} - \phi_2) + \sum_{i=1}^{m-1} c_{i0} H(\phi_1 - l_i) H(l_{i+1} - \phi_1) H(k_1 - \phi_2) + \sum_{i=1}^{m-1} c_{in} H(\phi_1 - l_i) H(l_{i+1} - \phi_1) H(\phi_2 - k_n) + \sum_{j=1}^{n-1} c_{0j} H(l_1 - \phi_1) H(\phi_2 - k_j) H(k_{j+1} - \phi_2) + \sum_{j=1}^{n-1} c_{mj} H(\phi_1 - l_m) H(\phi_2 - k_j) H(k_{j+1} - \phi_2) + c_{00} H(l_1 - \phi_1) H(k_1 - \phi_2) + c_{0n} H(l_1 - \phi_1) H(\phi_2 - k_n) + c_{m0} H(\phi_1 - l_m) H(k_1 - \phi_2) + c_{mn} H(\phi_1 - l_m) H(\phi_2 - k_n) \tag{8}$$

The level set functions, ϕ_1 and ϕ_2 , are used to find the optimum piecewise-constant approximation and are numerically built via the level set method. For minimizing the energy defined in (4), the differential of the energy with respect to the implicit functions can be transformed as the differential of the implicit functions with respect to time, in accordance with the chain rule in calculus. Thus, the differential results can be written as:

$$\begin{aligned} \frac{\partial \phi_1}{\partial t} &= \sum_{i=1, j=1}^{m-1, n-1} |I - c_{ij}|^2 [-\delta(\phi_1 - l_i) H(l_{i+1} - \phi_1) \prod_{p=j, q=1}^{j+1, q=2} H((-1)^q(k_p - \phi_2))] + \\ &\delta(l_{i+1} - \phi_1) H(\phi_1 - l_i) \prod_{p=j, q=1}^{j+1, q=2} H((-1)^q(k_p - \phi_2))] + \\ &\sum_{i=1}^{m-1} |I - c_{i0}|^2 [-\delta(\phi_1 - l_i) H(l_{i+1} - \phi_1) H(k_1 - \phi_2) + \delta(l_{i+1} - \phi_1) H(\phi_1 - l_i) H(k_1 - \phi_2)] + \\ &\sum_{i=1}^{m-1} |I - c_{in}|^2 [-\delta(\phi_1 - l_i) H(l_{i+1} - \phi_1) H(\phi_2 - k_n) + \delta(l_{i+1} - \phi_1) H(\phi_1 - l_i) H(\phi_2 - k_n)] + \\ &\sum_{j=1}^{n-1} |I - c_{0j}|^2 \delta(l_1 - \phi_1) H(\phi_2 - k_j) H(k_{j+1} - \phi_2) - \sum_{j=1}^{n-1} |I - c_{mj}|^2 \delta(\phi_1 - l_m) H(\phi_2 - k_j) H(k_{j+1} - \phi_2) + \\ &|I - c_{00}|^2 \delta(l_1 - \phi_1) H(k_1 - \phi_2) + |I - c_{0n}|^2 \delta(l_1 - \phi_1) H(\phi_2 - k_n) - |I - c_{m0}|^2 \delta(\phi_1 - l_m) H(k_1 - \phi_2) - \\ &|I - c_{mn}|^2 \delta(\phi_1 - l_m) H(\phi_2 - k_n) + \mu \sum_{i=1}^m \left[\delta(\phi_1 - l_i) \operatorname{div} \left(\frac{\nabla \phi_1}{|\nabla \phi_1|} \right) \right] \end{aligned} \tag{9}$$

and

$$\begin{aligned}
 \frac{\partial \phi_2}{\partial t} = & \sum_{i=1, j=1}^{m-1, n-1} |\mathbf{I} - c_{ij}|^2 [-\delta(\phi_2 - k_j) H(k_{j+1} - \phi_2) \prod_{p=i, q=0}^{i+1, 1} H((-1)^q (\phi_1 - l_p)) + \\
 & \delta(k_{j+1} - \phi_2) H(\phi_1 - l_i) \prod_{p=i, q=0}^{i+1, 1} H((-1)^q (\phi_1 - l_p))] + \\
 & \sum_{j=1}^{n-1} |\mathbf{I} - c_{0j}|^2 [-\delta(\phi_2 - k_j) H(k_{j+1} - \phi_2) H(l_1 - \phi_1) + \delta(k_{j+1} - \phi_2) H(\phi_2 - k_j) H(l_1 - \phi_1)] + \\
 & \sum_{j=1}^{n-1} |\mathbf{I} - c_{mj}|^2 [-\delta(\phi_2 - k_j) H(k_{j+1} - \phi_2) H(\phi_1 - l_m) + \delta(k_{j+1} - \phi_2) H(\phi_2 - k_j) H(\phi_1 - l_m)] + \\
 & \sum_{i=1}^{m-1} |\mathbf{I} - c_{in}|^2 \delta(\phi_2 - k_n) H(\phi_1 - l_i) H(l_{i+1} - \phi_1) + \sum_{i=1}^{m-1} |\mathbf{I} - c_{in}|^2 \delta(\phi_2 - k_n) H(\phi_1 - l_i) H(l_{i+1} - \phi_1) + \\
 & |\mathbf{I} - c_{00}|^2 \delta(k_1 - \phi_2) H(l_1 - \phi_1) - |\mathbf{I} - c_{0n}|^2 \delta(\phi_2 - k_n) H(l_1 - \phi_1) + \\
 & |\mathbf{I} - c_{m0}|^2 \delta(k_1 - \phi_2) H(\phi_1 - l_m) - |\mathbf{I} - c_{mn}|^2 \delta(\phi_2 - k_n) H(\phi_1 - l_m) + \mu \sum_{j=1}^n \left[\delta(\phi_2 - k_j) \operatorname{div} \left(\frac{\nabla \phi_2}{|\nabla \phi_2|} \right) \right] \quad (10)
 \end{aligned}$$

While the modified energy function defined in (4) reaches convergence with increasing iterations, the two level set functions are stopped automatically and updated at the end of each iteration. Both partial differential equations (PDEs) are solved using the finite difference method and an iterative approach. In this paper, we extend the numerical algorithm proposed by Vese to ensure that the solutions reach numerical stability [9]. Both PDEs can be solved numerically in a few iterations to obtain the piecewise-constant approximation. Two initial level set functions, ϕ_1^0 and ϕ_2^0 , need to be established in order to implement the numerical algorithm. Eventually, the zero level lines generated using the level set functions lie exactly on the sub-regional boundaries while the defined energy function reaches its convergence value. The regional constants are then calculated to solve the partial differential equations given in (9) and (10) for each iteration. The piecewise-constant approximation, PCA, is iterated to obtain a result similar to the original image for the limited regions.

3. Piecewise-constant approximations for digital mammograms

Experimental data obtained from the Database for Screening Mammography, from the

University of South Florida is employed in this paper [15]. The database contains 2620 cases, and each case has several mammograms. Each mammogram was digitized to a pixel size of 0.0435mm by 0.0435mm and 12-bit gray scale. A mammogram with an image size of 6241 by 4171 pixels was chosen to evaluate the validity of the multilayer level set approach, and it is illustrated in Figure 1. The mammogram analysis requires more computation time due to its large size. Thus, this paper employs an algorithm to shrink the original image to a smaller sized image. The resizing algorithm was implemented by taking every four pixels in the row and column directions to produce the pixels shown in the shrunken mammogram, by employing spline interpretation such that the original mammogram whose size is 6241 by 4171 pixels, shrank to 1606 by 1043 pixels. Similarly, another mammogram was reduced to 803 by 522 pixels, by taking every 8 pixels in the row and column directions. Both mammograms are illustrated in Figure 2.

Both mammograms are processed using the multilayer level set approach to enable a comparison between the resulting differences.

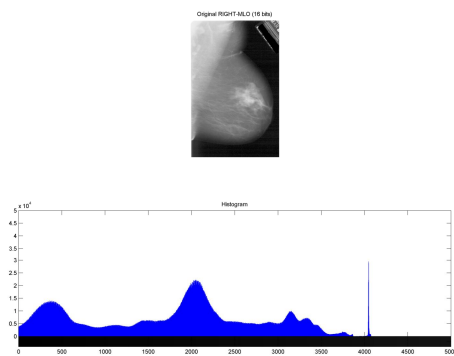


Figure 1. Original mammogram (6241x4171 pixels, 16 bits)

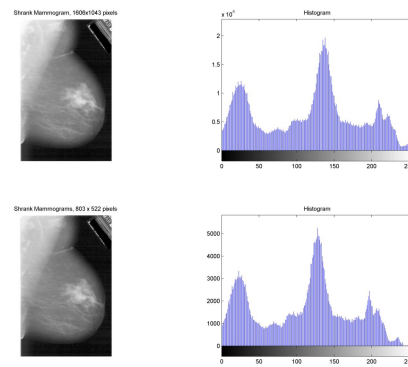


Figure 2. Shrunken mammograms (1606x1043 and 803 x522 pixels, respectively) and their corresponding histograms

Prior to implementing the multilayer level set approach, the predefined level values need to be established. In this paper, the predefined level values are chosen from the respective histograms. In Figure 3, an example illustrates the selection of level values from the histograms. Figure 4 shows the scenario where two level set functions were created and are lying on the shrunken mammogram.

The finite difference scheme was used to solve the partial differential Equations (9) and (10), and a series of piecewise-constant approximations is illustrated in Figure 5. Figure 6 illustrates two final approximations for the shrunken mammograms and the relationships between the iterations and the energies defined in (4).

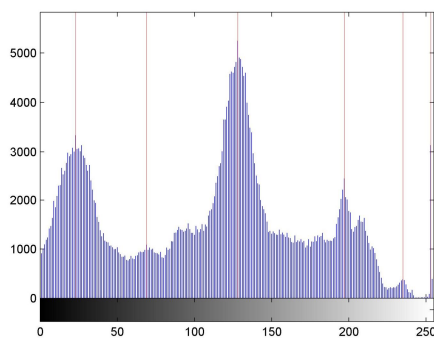


Figure 3. The predefined level values were selected from the histogram of the shrunken mammogram

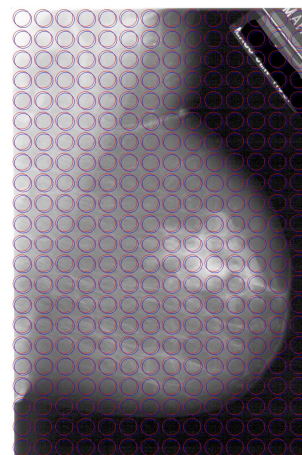


Figure 4. Two level set functions are applied to the shrunken mammogram

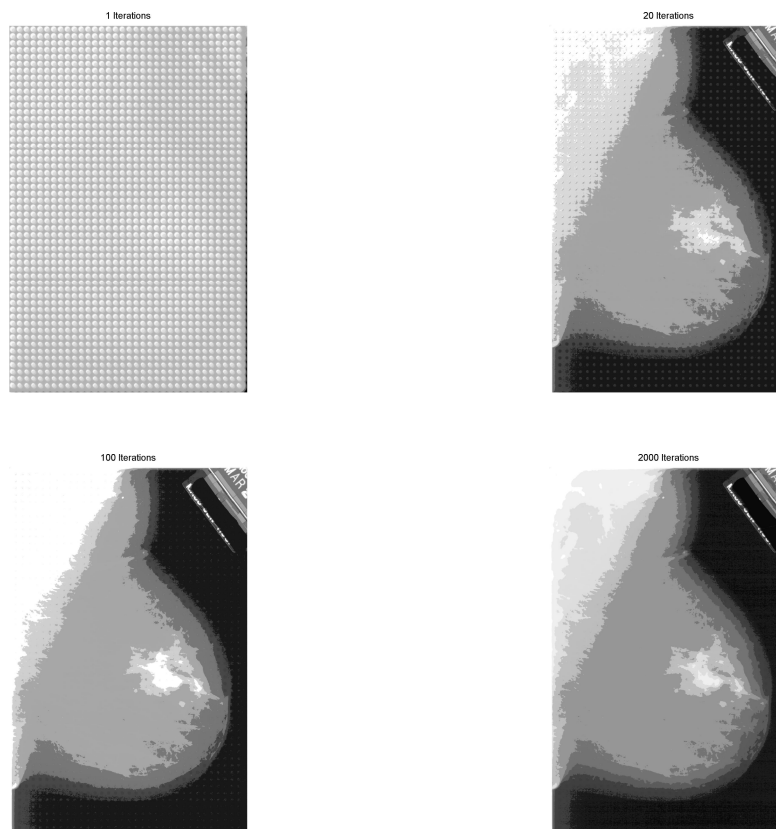


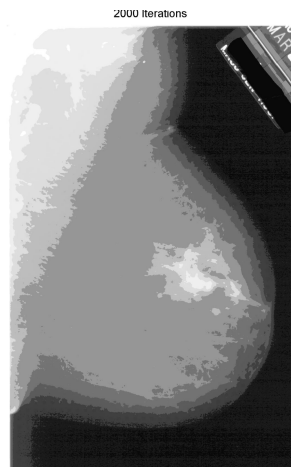
Figure 5. The piecewise-constant approximations were generated using the multilayer level set approach for 1, 20, 100 and 2000 iterations, respectively

4. Discussions and conclusions

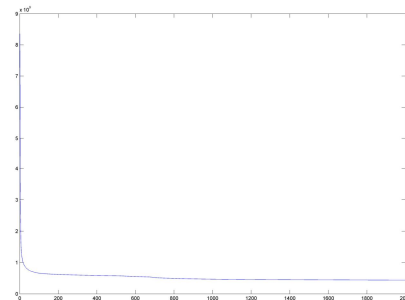
The piecewise-constant approximations generated with the multilayer level set approach provide segmented results such that each segmented region is homogeneous. From Figure 6, it can be seen that there are no large visual differences between the two shrunk

The shrunk algorithm is employed to reduce the mammogram sizes to increase the processing speed. After comparing the processed results illustrated in Figure 7 (b) and (c), the circled region in the central part disappears in the shrunk mammograms with smaller sizes (803 by 522 pixels). The result indicates that the shrinking algorithm can reduce the amount of breast cancer detail in the resulting image. Sometimes, it is

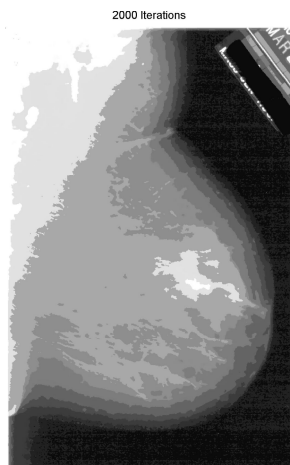
mammograms. In Figure 7, the mammograms containing breast cancer regions identified by physicians are visually close to the approximations generated with the multilayer level set approach. Furthermore, the positions of the central part circled in red, as depicted in the approximations, are very close to the cancer regions circled by the physicians. It is difficult to achieve a balance between the processing speed and amount of information detail required. However, the results demonstrate that the multilayer level set approach helps overcome this difficulty. The layer regions shown in the piecewise-constant approximation highlight their regional boundaries, to indicate the possible locations of the breast cancer.



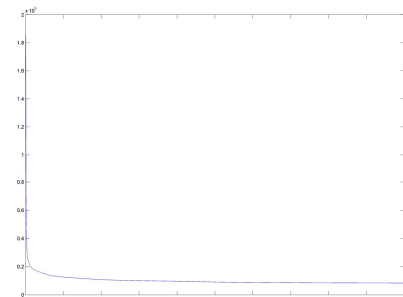
The shrunken mammogram (1606 x 1043 pixels)



The relationship between the iterations and the energy



The shrunken mammogram (803 x 522 pixels)



The relationships between iteration and the energy

Figure 6. The piecewise-constant approximations for 2000 iterations are shown, and the relationship between the iterations and energy defined in equation (4) is also depicted

The multilayer level set approach helps depict the breast cancer regions in CAD. The algorithm offers a stable numerical approach. The iterations and the energy defined in (4) tend to converge quickly. Roughly 100 iterations were required for the shrunken mammograms to achieve convergence when the multilayer level set

approach was applied. From the experimental results, the multi layer level set method helps to extract the features in the breast cancer regions. Furthermore, these extracted features can be incorporated into a training set and employed in a supervised learning algorithm to enable the automatic detection of breast cancer in the figure.

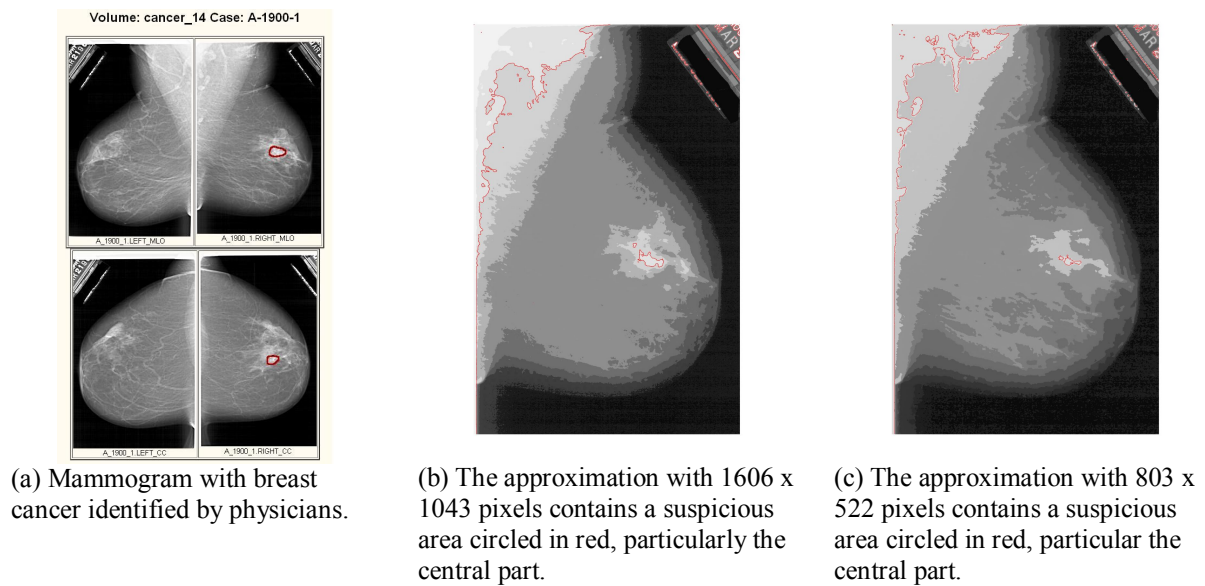


Figure 7. A comparison between the suspicious regions and the results identified by the physicians

References

- [1] World Health Organization, Breast cancer prevention and control, web site:<http://www.who.int/cancer/detection/breastcancer/en/index1.html>.
- [2] Li, H., Liu, K. J., and Lo, S. 1997. Fractal Modeling and Segmentation for the Enhancement of Microcalcifications in Digital Mammograms, *IEEE Trans. Medical Imaging*, 16, 6: 785-798.
- [3] Selvi, S. T. and Malmathanraj, R. 2006. Segmentation and SVM Classification of Mammograms. *IEEE International Conference on Industrial Technology, ICIT*: 905-910.
- [4] Ball, J. E. and Bruce, L. M. 2007. Level Set-Based Core Segmentation of Mammographic Masses Facilitating Three Stage (Core, Periphery, Spiculation) Analysis, *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, Lyon, France: 819-824.
- [5] Sickles, E. A. 1986. Mammographic Features of Early Breast Cancer. *American Journal of Roentgenology*, 413: 461-464.
- [6] Nakayama, R., Uchiyama, Y., Yamamoto, K., Watanabe, R., and Namba, K. 2006. Computer-Aided Diagnosis Scheme Using a Filter Bank for Detection of Microcalcification Clusters in Mammograms. *IEEE Trans. Biomedical Engineering*, 53, 2: 273-283.
- [7] Kass, M., Withkin, A., and Terzopoulos, D. 1988. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1, 4: 321-333.
- [8] Ayed, I. B., Mitiche, A., and Belhadj, A. 2005. Multiregion Level-Set Partitioning of Synthetic Aperture Radar Images. *IEEE. Trans PAMI*, 27, 5: 793-800.
- [9] Vese, L. A. and Chan, T. 2002. A Multiphase Level Set Framework for

- Image Segmentation Using the Mumford and Shah Model. *International Computer Vision*, 50, 3: 271-293.
- [10] Chung, G. and Vese, L. A. 2009. Image Segmentation Using a Multilayer Level-set Approach. *Computing and Visualization in Science*, 12, 6: 267-285.
- [11] Chan, S., Merriman, B., Kang, M., Caflisch, R. E., Cheng, L., Gyure, M., Fedkiw, R. P., Anderson, C., and Osher, S. 2001. A Level Set Method for Thin Film Epitaxial Growth. *Journal of Computational Physics*, 167: 475-500.
- [12] Osher, S. and Sethian, J. A. 1988. Front Propagation with Curvature Dependent Speed: Algorithm Based on Hamilton Jacobi Formulation. *Journal of Computational Physics*, 79: 12-49.
- [13] Mumford, D. and Shah, J. 1989. Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems. *Communication and Pure and Applied Mathematics*, 42: 577-685.
- [14] Ratsch, C., Gyure, M. F., Caflisch, R. E., Petersen, M., Kang, M., Garcai, J., and Vvedensky, D. D. 2002. Level-Set Method for Island Dynamics in Epitaxial Growth. *Physical Review B*, 65: 195403.
- [15] Heath, M., Bowyer, K., Moore, R., and Kegelmeyer, W. P. 2001. The Digital Database for Screening Mammography, in *Proceedings of the Fifth International Workshop on Digital Mammography*: 212-218.