Boundary Segmentation of Microcalcification using Parametric Active Contours

Abdul Kadir Jumaat, Siti Salmah Yasiran, Wan Eny Zarina Wan Abdul Rahman, Aminah Abdul Malek

Abstract—A mammography image is composed of low contrast area where the breast tissues and the breast abnormalities such as microcalcification can hardly be differentiated by the medical practitioner. This paper presents the application of active contour models (Snakes) for the segmentation of microcalcification in mammography images. Comparison on the microcalcification areas segmented by the Balloon Snake, Gradient Vector Flow (GVF) Snake, and Distance Snake is done against the true value of the microcalcification area. The true area value is the average microcalcification area in the original mammography image traced by the expert radiologists. From fifty images tested, the result obtained shows that the accuracy of the Balloon Snake, GVF Snake, and Distance Snake in segmenting boundaries of microcalculation are 96.01%, 95.74%, and 95.70% accuracy respectively. This implies that the Balloon Snake is a better segmentation method to locate the exact boundary of a microcalcification region.

Keywords—Balloon Snake, GVF Snake, Distance Snake, Mammogram, Microcalcifications, Segmentation

I. INTRODUCTION

EARLY detection for breast cancer can save lives. Unfortunately, nearly 40% of new cases of breast cancer identified each year in Malaysia were already in the critical stage [1]. Clinical studies have demonstrated that survival rate is greatly improved if any abnormalities such as masses or microcalcifications are detected at the earliest stage [2]. Currently, a “gold standard” screening modality called mammogram is considered as the first line of defense in early breast cancer detection [2]. A mammography image is composed of low contrast area where the breast tissues and the breast abnormalities can hardly be differentiated by the medical practitioner. Microcalcifications are an early sign of a breast cancer. It is composed of tiny bits of calcium, and may show up in clusters or in patterns. Sometimes tight clusters of microcalcifications can indicate early breast cancer. Scattered microcalcifications are usually a sign of benign breast tissue [3]. However, the existence of microcalcifications is often difficult to detect by radiologists since the structure is almost similar with the other normal breast tissue.

Image segmentation can be defined as an identification of the boundary of object in an image [4]. Image segmentation technique is done in order to delineate the abnormalities region and the background. This means that if the boundary in an object can be identified accurately, all of the basic properties of the object such as size, shape, area, and perimeter can be measured.

Segmentation techniques are classified into low-level and high-level techniques. Low-level techniques such as Canny method, Sobel method, and region growing method simply analyze an image by reducing the amount of data to be processed. However, the reduction of the amount of data to be processed may reduce the image information. Moreover, the low-level segmentation techniques may incorrectly identify region or boundary of an object due to the distraction of noise in an image [5].

High-level techniques such as Active Contours or Snakes are focused to overcome many of the limitations in the low-level image segmentation techniques. Snakes are computer-generated curves that move within an image to find the object boundary. The curve moves under the influences of the internal energy within the curve itself and the external energy which are derived from the image data [6].

Snakes were originally introduced by Kass et al. in 1986 [7]. Research carried out by [4], [7], [8], and [9] agreed that Snakes are proven to be effective methods in line and edge detection, image segmentation, shape modeling, and motion tracking.

There are three types of Active Contours or Snakes namely Parametric Active Contour, Discrete Dynamic Active Contour, and Geometric Active Contour. This paper will only focus on three types of Parametric Snake models to segment microcalcifications in mammography images. They are Balloon Snake, GVF Snake and Distance Snake.

The Balloon Snake, GVF Snake, and Distance Snake are three different examples of Parametric Active Contour. The internal energy apply in the methods consist of two energy terms, namely the elasticity energy and bending energy [6]. Elasticity energy are designed to keep the Snakes contracted or hold together while bending energy keep the curve smooth by avoiding it from bending too much [4], [6].

However, they are using different external energy. The external energy derived from the image data are responsible for driving the snakes to the boundary of the image [4], [6]. GVF Snake uses Gradient Vector Flow (GVF) field as the external energy [6]. The Balloon Snake uses the sum of the image energy and pressure energy as the external energy terms [10], while the Distance Snake uses distance energy plus image energy as its external energy terms. This different external energy used resulted to different image segmentation.

There are lots of studies conducted by other researcher by using the Balloon Snake and GVF Snake in segmentation area. Some of the researchers try to improve and enhance the GVF snake to become a new model [11]-[13]. One of the studies on Balloon and GVF associate with the segmentation has successfully conducted by Jumaat et al [14]-[15]. However, instead of mammogram images, the application of the GVF and Balloon Snakes by Jumaat et al. is only limited on the ultrasound images. The Distance Snake is also originally proposed by Xu & Prince [6]. It has been used as a comparison purpose with the GVF Snake [17].
In 2010, the Distance Snake have been successfully applied to segment the microcalcification [18]. In this paper, the Balloon Snake, GVF Snake, and Distance Snake are used to segment microcalcifications on mammography images. Comparison on the microcalcification areas segmented by the methods is done against the true value of the microcalcification area. The true area value is the average microcalcification area in the original mammography image traced by the expert radiologists. The performances of the methods are measured in terms of percentage area pixel difference.

II. THE ACTIVE CONTOUR (SNAKE)

Snake is a vector valued function in the spatial domain of an image, parametrically expressed as \( v(s) = (x(s), y(s)) \) where \( 0 \leq s \leq 1 \). A curve consists of \( n \) vertices \( v \) connected by straight lines. The parameter \( x \) and \( y \) are the coordinate of the vertices, \( v \) and are functions of the normalized arc length \( s \).

The Snake has a dynamic behavior that deforms from an initial position and converges to the boundary of the object in the image. It moves through the domain of the image by minimizing its energy function, \( E_{\text{snake}} \), which is defined as

\[
E_{\text{snake}} = \frac{1}{2} \int_{0}^{1} \left[ E_{\text{int}}(v(s)) + E_{\text{ext}}(v(s)) \right] ds
\]

(1)

The internal energy function is

\[
E_{\text{int}}(v(s)) = \frac{1}{2} \left\{ \alpha \left| \nu(s) \right|^2 + \beta \left| \nu''(s) \right|^2 \right\}
\]

(2)

The internal energy function, \( E_{\text{int}}(v(s)) \), is computed based on the local shape of the curve \( v(s) \), and is responsible in determining the continuity and the smoothness of the curve. The parameter \( \alpha \) and \( \beta \) are the coefficient of the internal energy function. The parameter \( \alpha \) is the elasticity parameter. For a large value of \( \alpha \), the curve becomes very straight between two points. The parameter \( \beta \) is the rigidity parameter and for a large value of \( \beta \) the curve becomes smooth.

The external energy function \( E_{\text{ext}}(v(s)) \) is derived based on the image information and it drives the curve to the boundary of the object. Different types of Snakes use different type of external energy function.

By calculus of variation, Equation (1) is minimized by solving the associate Euler’s Equation as follows

\[
-\alpha \nu''(s) + \beta \nu^{(4)}(s) + \nabla E_{\text{ext}}(v(s)) = 0
\]

(3)

The Euler’s equation is approximated with finite difference method since it is difficult to be computed analytically. Converted to vector notation with \( v_i = (x_i, y_i) \), now Equation (3) becomes

\[
\begin{bmatrix}
\alpha_i (v_i - v_{i-1}) - \alpha_{ii} (v_{i+1} - v_i) + \beta_{ii} (v_{i-1} - 2v_i + v_{i+1}) + \\
-2v_{i-1} + v_i - 2 \beta_i (v_{i-1} - 2v_i + v_{i+1}) + \\
\beta_{i+1} (v_i - 2v_{i+1} + v_{i+2}) \\
\end{bmatrix} + \\
\left[ \frac{\partial}{\partial x} E_{\text{ext}}(x, y), \frac{\partial}{\partial y} E_{\text{ext}}(x, y) \right] = 0
\]

(4)

or with respect to coordinate components;

\[
\begin{align*}
\alpha_i (x_i - x_{i-1}) - \alpha_{ii} (x_{i+1} - x_i) + \beta_{ii} (x_{i-1} - 2x_i + x_{i+1}) - \\
2\beta_i (x_{i-1} - 2x_i + x_{i+1}) + \beta_{i+1} (x_i - 2x_{i+1} + x_{i+2}) \\
\alpha_j (y_j - y_{j-1}) - \alpha_{jj} (y_{j+1} - y_j) + \beta_{jj} (y_{j-1} - 2y_j + y_{j+1}) - \\
2\beta_j (y_{j-1} - 2y_j + y_{j+1}) + \beta_{j+1} (y_j - 2y_{j+1} + y_{j+2})
\end{align*}
\]

\[
\frac{\partial}{\partial x} E_{\text{ext}}(x, y) = 0 \\
\frac{\partial}{\partial y} E_{\text{ext}}(x, y) = 0
\]

In matrix form

\[
A x + \frac{\partial}{\partial x} E_{\text{ext}}(x, y) = 0 \\
A y + \frac{\partial}{\partial y} E_{\text{ext}}(x, y) = 0
\]

(5)

where \( x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \) and \( y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \). \( A \) is a pentadiagonal banded matrix and the size of the matrix is \( nxn \) where \( n \) being the number of vertices. For example, for a Snake consist of 5 vertices, matrix \( A \) is as follows

\[
A = \begin{bmatrix}
2\alpha + 6\beta & -\alpha - 4\beta & \beta & 0 & 0 \\
-\alpha - 4\beta & 2\alpha + 6\beta & -\alpha - 4\beta & 0 & 0 \\
\beta & -\alpha - 4\beta & 2\alpha + 6\beta & -\alpha - 4\beta & \beta \\
0 & \beta & -\alpha - 4\beta & 2\alpha + 6\beta & -\alpha - 4\beta \\
0 & 0 & \beta & -\alpha - 4\beta & 2\alpha + 6\beta
\end{bmatrix}
\]

In order to solve the matrix equation, the right hand sides of Equation (5) is set equal to the product of a step size \( \gamma \) and the negative time derivatives of the left hand sides [9].
Now, Equation (5) becomes
\[
\begin{align*}
Ax_t + \frac{\partial}{\partial x} E_{ext}(x_{t-1}, y_{t-1}) &= -\gamma (x_t - x_{t-1}) \\
Ay_t + \frac{\partial}{\partial y} E_{ext}(x_{t-1}, y_{t-1}) &= -\gamma (y_t - y_{t-1})
\end{align*}
\]  
where subscript \( t \) is the iteration number.
By matrix inversion \( I \), Equation (6) becomes
\[
\begin{align*}
x_t &= (A + \gamma I)^{-1} \left( \mathbf{x}_{t-1} - k \frac{\partial}{\partial x} E_{ext}(x_{t-1}, y_{t-1}) \right) \\
y_t &= (A + \gamma I)^{-1} \left( \mathbf{y}_{t-1} - k \frac{\partial}{\partial y} E_{ext}(x_{t-1}, y_{t-1}) \right)
\end{align*}
\]  
(7)

The weighting parameter \( k \) is an additional parameter to control the internal and external energy functions. When the right hand side of Equation (7) equals to zero, then the location or position of the vertices in the successive iteration is unchanged.

The Distance snake is originally proposed by Xu & Prince [6]. Distance Snake uses distance energy plus image energy as its external energy terms. The \( E_{ext}(v(s)) \) defines as follows
\[
E_{ext} = d_{(x,y)} = \left[ (x_2 - x_1)^2 + (y_2 - y_1)^2 \right]^{\frac{1}{2}}
\]
where \( d_{(x,y)} \) is pixel distance of the image.

In order to determine the vertex position using Distance Snake, Equation (7) is modified as follows:
\[
\begin{align*}
x_t &= (A + \gamma I)^{-1} \left( \mathbf{x}_{t-1} - k \frac{\partial}{\partial x} \mathbf{d}(x_{t-1}, y_{t-1}) \right) \\
y_t &= (A + \gamma I)^{-1} \left( \mathbf{y}_{t-1} - k \frac{\partial}{\partial y} \mathbf{d}(x_{t-1}, y_{t-1}) \right)
\end{align*}
\]  
(8)

As proposed by Cohen [10], the Balloon Snake uses the sum of pressure energy and image energy as the external energy function, \( E_{ext}(v(s)) \) defines as follows
\[
E_{ext} = -k \frac{\mathbf{F}_{image}}{||\mathbf{F}_{image}||} + k \text{pressure} \mathbf{n}(s)
\]  
(9)

The parameter \( k \text{pressure} \) is the pressure weight or pressure energy and its positive or negative sign causes the Snake to inflate or deflate respectively. The image energy \( \mathbf{F}_{image} \) is the gradient of the image edge map. The parameter \( k \) is the image energy weighting. The symbol \( \mathbf{n}(s) \) represents the unit normal vector to vertices.

In order to determine the vertex position using Balloon Snake, Equation (7) is modified as follows:
\[
\begin{align*}
x_t &= (A + \gamma I)^{-1} \left( \mathbf{x}_{t-1} \pm k \text{pressure} \mathbf{n}(x_{t-1}) \right) - k \frac{\partial}{\partial x} \mathbf{F}_{image}(x_{t-1}, y_{t-1})) \\
y_t &= (A + \gamma I)^{-1} \left( \mathbf{y}_{t-1} \pm k \text{pressure} \mathbf{n}(y_{t-1}) \right) - k \frac{\partial}{\partial y} \mathbf{F}_{image}(x_{t-1}, y_{t-1}))
\end{align*}
\]  
(10)

Xu and Prince introduced GVF Snake in 1997. This new method applies the gradient vector flow field as the external energy function. The gradient vector flow field is defined to be the vector field \( \mathbf{x}(x,y) = [u(x,y), v(x,y)] \) that minimizes the energy functional,
\[
\begin{align*}
\varepsilon &= \iint [\mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + || \nabla f ||^2] \\
&\quad | \mathbf{x} - \nabla f || \nabla f ||^2 \]dx\;dy
\end{align*}
\]  
(11)

where \( f \) is an image edge map, \( \nabla f(x,y) \) is the gradient of image edge map, and \( \mu \) is weighting parameter.

The value of \( || \nabla f || \) is large in a very noisy image and the second term of integrand dominates Equation (11) and is minimized by setting \( \mathbf{x} = \nabla f \). On the other hand, in image with less noise, the value \( || \nabla f || \) is small and the first term of the integrand dominates Equation (11) yielding a slowly-varying field. In order to adjust the trade off between the first and the second integrand, the weighting parameter \( \mu \) is introduced. The more noise in the image, the larger the value of \( \mu \) is needed.

By calculus of variation, Equation (11) is minimized by solving the associate Euler’s Equation below:
\[
\mu \nabla^2 \mathbf{x} = |\nabla f| \nabla f || \nabla f ||^2
\]  
(12)

To solve the equation, \( \mathbf{x} \) is treated as functions of time and solves the system iteratively using central difference method.

After \( \mathbf{x}(x,y) \) is computed, \( \mathbf{x}(x,y) \) replaces the term \( E_{ext} \) in the Equation (7) to yield
\[
\begin{align*}
x_t &= (A + \gamma I)^{-1} \left( \mathbf{x}_{t-1} - k \frac{\partial}{\partial x} u(x_{t-1}, y_{t-1}) \right) \\
y_t &= (A + \gamma I)^{-1} \left( \mathbf{y}_{t-1} - k \frac{\partial}{\partial y} v(x_{t-1}, y_{t-1}) \right)
\end{align*}
\]  
(13)
III. MATERIALS AND METHOD
There are three main stages employed in this study. The first stage is the data collection. Fifty mammography images which contain microcalcifications are obtained from National Cancer Society Malaysia (NCSM). The image is cropped using Adobe Photoshop CS3 software to remove unnecessary data and it is saved in a standard size of 200 by 200 pixels. After that, it will be preprocessed in order to remove the noise and artifacts in the original image using Median Filtering and Histogram Stretching technique. This process is carried out by using the MATLAB R2008a software.

In GVF Snake, the weighting parameter $\mu$ is set to 0.2 with 80 iterations to compute the GVF field. The performance comparison for the segmentation results is carried out in the last stage. Comparison on the microcalcification areas segmented by the methods is done against the true value of the microcalcification area. The true area value is the average microcalcification area in the original mammography image traced by the two expert radiologists. The performances of the methods are measured in terms of percentage area pixel difference. The area is calculated in pixel square unit (pixel$^2$).

IV. EXPERIMENTAL RESULTS
Figure 2 illustrates some of the segmentation results for the Balloon Snake, GVF Snake, and Distance Snake. The original images are shown in part (a), segmented images by Balloon Snake in part (b), Distance Snake in part (c) and the GVF Snake in part (d).

In the first column, the original image 1(a) is successfully segmented by all the three methods. This is due to the properties of the original image which is pretty clear and its boundary also not hard to be differentiated from the background tissue. However, there is a weakest edge for the second image 2(a). It is found that the Balloon Snake overwhelms the boundary of the image. One of the main reasons is because the pressure energy terms used in Balloon Snake is hard to be determined. Balloon Snake can converge to a concave region by increasing its pressure energy. But, increasing the pressure energy may overwhelm the subjective region, i.e. parts of a boundary with weak or zero contrast.
while a small value of pressure energy will make the Snake easily trapped to image noise. On the other hand, the boundary for the third original microcalcification image 3(a) used is very weak and it is not well defined where the foreground and background of the breast tissue are quite hard to be differentiated. On top of that, the actual boundary of the targeted object in the image has a quite similar pixel value with its neighboring area. The result shows that the image is not perfectly segmented by all the three snakes. They are unable to converge to the actual boundary of the microcalcifications due to the blur image. Furthermore, it is shown that the result in part 3(c), the Distance Snake is easily trapped to the neighboring area of the microcalcification compared to the Balloon Snake in part 3(b).

In addition, there are two inaccurate (fail) cases of the GVF Snake in which the final deformations do not converge to the boundary of tumor, whereas no fail cases are noted on the Balloon and Distance Snakes. This is because of the initialization problem. It is a problem where the Snake cannot moves toward an object that is too far away from its initial position to find the object boundary [6]. Figure 3 shows one inaccurate boundary segmentation by the GVF Snake in part (b).

Fig. 3 Original image (a) and GVF Snake fails (b).

In overall, from fifty images tested, the average percentage area difference for the Balloon Snake, GVF Snake, and Distance Snake is 3.99%, 4.26% and 4.3% respectively. These imply that the accuracy of the Balloon Snake, GVF Snake, and Distance Snake in segmenting boundaries of microcalcification are 96.01%, 95.74%, and 95.70% accuracy respectively. Therefore, the Balloon Snake is a better segmentation method to locate the exact boundary of a microcalcification region.

The boundary of the microcalcification identified will assist radiologists in locating potentially cancerous cases for further analysis such as the classification of malignant or benign microcalcification, or in tissue sampling in the biopsy procedure.

ACKNOWLEDGMENT

This research was supported by the Universiti Teknologi MARA (UiTM) under the Research Intensive Faculty (RIF), 600-RMI/DANA 5/3/RIF (58/2012). The author(s) would like to acknowledge UiTM and also the consultant radiologist from University Putra Malaysia, Prof Dr Rozi Mahmud for the support and contributions.

REFERENCES


Abdul Kadir Jumaat is a lecturer in mathematics at centre of mathematical studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA. He received BSc (H) Mathematics in 2008 and Master of Science by Research (Computational Mathematics) in 2011 from Universiti Teknologi MARA. He is a member of IEEE and has research interest is in medical image processing and computational mathematics. 

Siti Salmuah Yasiran is a lecturer in mathematics at centre of mathematical studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA. She received BSc (H) Computational Mathematics in 2008 and Master of Science by Research (Computational Mathematics) in 2010 from Universiti Teknologi MARA. She is a member of IEEE and SIAM. Her research interest is in medical image processing, artificial neural network and algorithm development.

Assoc. Prof. Dr. Wan Emy Zarina Wan Abdul Rahman is a senior mathematics lecturer at the same faculty. She did her master in Newcastle UponTynne, UK, in 1993 and received her PhD at Universiti Teknologi Malaysia in 2006. Her research interest is in mathematical medical imaging.

Anniah Binti Abdul Malek is a lecturer in mathematics at Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan. She received BSc (H) Computational Mathematics in 2007 and Master of Science by Research (Computational Mathematics) in 2012 from Universiti Teknologi MARA. Her research interest is in medical image processing and mathematical image analysis.