Semi-Automated Detection of Breast Mass Spiculation Using Active Contour

Piyatragoon Boonthong[1], Benchaporn Jantarakongkul[2], Suwanna Rasmequan[3],
Annupan Rodtook[4] and Krisana Chinnasarn[5]

[1,2,3,5]Department of Computer Science, Faculty of Informatics, Burapha University, Chonburi, Thailand
[4]Department of Computer Science, Faculty of Science, Ramkhamhaeng University, Bangkok, Thailand

Abstract— The use of computer research for breast cancer diagnosis in digital mammograms has been studied by some researchers for years. The researches based on medical image processing were developed and published continuously. Theirs objective are to create a diagnostic tool that can increase the accuracy of risk analysis for breast cancer. At the early stage, cancers may be identified as spiculated masses revealing architectural distortion. This research proposes a semi-automated method to detect architectural distortion characterized by thin lines radiating from its margins. It will help physicians as second or minor opinion before biopsy operation. The proposed method involves following major steps in sequence. A combination of the object attributes thresholding, hill-climbing and region growing algorithm is applied to digital mammogram for background and breast pectoral muscle removal. The second is a region of interest (ROI) selection based on image splitting and breast ratio estimation. In the third step, the shade corrections of ROI are considered by using the contrast-limited adaptive histogram equalization. Next, we apply the modified hierarchical clustering to detect and enhance the possible cluster of spiculated masses. The other clusters will be a significant reduction. The final step is established to segment spiculated shape by employing the parametric active contour method. The numerical experiments of the proposed method are performed by testing on the digital database for screening mammography (DDSM) made up by the University of South Florida.

I. INTRODUCTION

Breast cancer is the most common cancer in woman worldwide. Risk of breast cancer has increased more than 20% over the past few years. In Thailand, breast cancers are diagnosed with the most frequency. Moreover, it is the leading cause of cancer death.

The precise causes of breast cancer are unclear, however women's risks investigated are ages more than 40, gender, race and family history of breast cancer. The early cancer detection is the most importance to stop cancer developing and spreading by providing an effective treatment. To detect breast cancer in digital mammography, mass screening has been performed with research applications of medical image processing. They were designed for detecting early characteristics of cancer such as calcifications, architectural distortion and spiculated mass. In this paper, we propose multi-stages algorithm for the detection of mass presenting spiculated shapes. It always exists on the periphery of the breast and has a thin, elongated piece of tissue spreading out from its perimeter.

In last decade, medical image processing research have been conducted and improved worldwide for developing effective tools of breast cancer diagnosis. It may help physicians as guidance or minor opinion before biopsy operation. Naga et al. [1] introduced algorithms to detect breast mass in mammographic images. A combination of Gaussian filtering and sub-sampling operation was the first preprocessing step. Angular anisotropy of the mammogram was needed and calculated as flow. Next, textural orientations of flow were analyzed for mass segmentation. Mehul and Alan [2] proposed method of speculated lesions detection in mammography. They applied the combination of enhancement and linear filtering processes for the detection stage. Theirs method was evaluated by testing with the digitized mammography obtained from the digital database for screening mammography (DDSM). Luan et al. [3] present the fully automated detection of breast mass spiculation consisting of principle steps as follows. In the region of interest (ROI), the initial outlines of mass are detected using maximum entropy concept. Next, they apply an active contour model to initial outlines for identifying spiculated lines. Abdel and Mahmoud [4] present a suspicious mass detection in digital mammography. This method based on thresholding scheme used the fuzzy entropy to estimate the optimal threshold for segmentation. Finally, they proposed a block-based performance criterion to measure the result accuracy. F.Zou et al. [5] proposed an algorithm to extract the regions of suspicious mass. They applied an adaptive histogram equalization to enhance mammographic images. Then, the convex hull of edge points subjected to GVF field is used to generate the mass region of mammography.

The main principles of this proposed method involve following conventional image processing steps in sequence. Multiresolution analysis and denoising of digitized mammography of breast are performed using the discrete wavelet transform (DWT). The next process is to investigate mammographic background using the object attribute thresholding (OAT) [6]. The pectoral muscle of breast is discovered by using the combination of hill-climbing [7] and region-growing algorithm [8]. Such locations of background and muscle are marked to ignore processing in the advanced step. The remaining areas of the digitized mammography are split into sub-images. The breast density is evaluated to

achieve the region of interest (ROI). We apply the contrast-limited adaptive histogram equalization [9] and modified hierarchical clustering algorithm [10] to enhance the quality of breast tissue (inside ROI). Final step, active contour model based on generalized gradient vector flow (GGVF) is used to localize spiculated mass. The proposed method offering a simple computational scheme has been tested using the digital mammography dataset obtained by DDSM database of the University of South Florida [11]. It provides a high accuracy of spiculated mass segmentation.

This paper is organized as follows. Section 2 presents background of conventional methods used. In section 3, we describe the proposed methods for detection of breast mass spiculation in this research. Section 4 deals and discuss with the experimental results. Finally, section 5 summarizes this research and suggests directions for future research.

II. THEORY

A. The discrete wavelet transform

The size of image and high noise level may negatively affect processing performance. Multiresolution analyses based on wavelet decomposition are the successive version of resolution and noise reduction levels. The sequence of approximating spaces of image is generated by the so called scaling function φ [12-14], whereas the wavelet function ψ [12-14] is employed to represent the orthogonal complements to the approximating spaces called the detail spaces. We define the approximation and the detail images respectively as following figure.
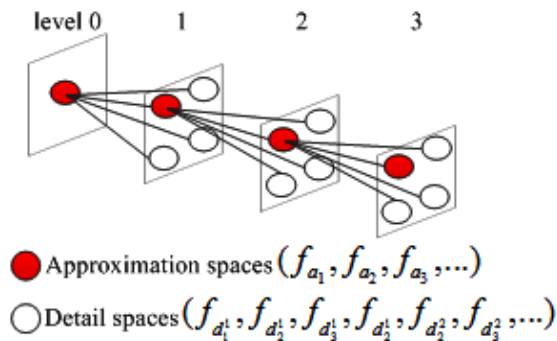

Fig. 1 Wavelet Decomposition

The discrete version of the above decomposition called the filter bank was proposed by Mallat [12], Unser and Aldroubi [13]. It was shown that the discrete wavelet transform can be performed by using the so called finite impulse response filters (FIRF) which produce a tree structured filter bank (see Fig. 2(a)). The characteristics of these filters (h and g filters) associate with wavelet decomposition which h and g are the low pass and the high pass FIR's associated with corresponding scaling and wavelet functions respectively.
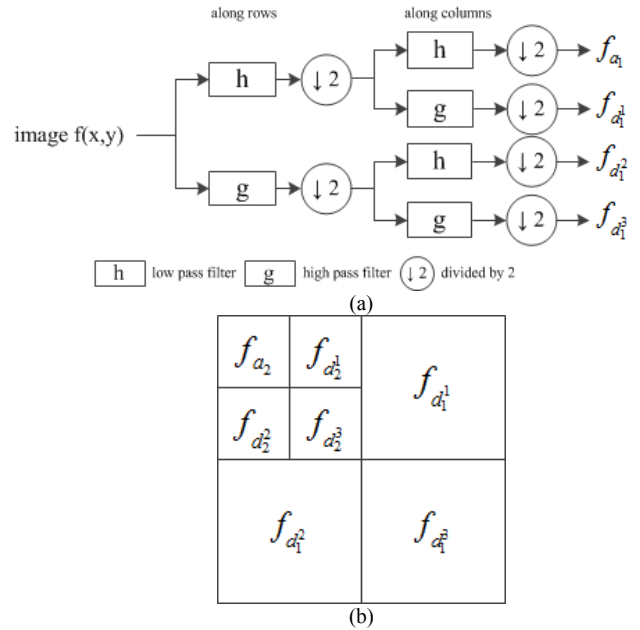

(a)

(b)

Fig. 2 A filter bank, (a) A tree, (b) Wavelet subbands

B. The GGVF Snake

Parametric active contour models or snake defined as $X(s) = (x(s), y(s))$ is a curve evolution method [15-17]. It moves snake inside the image to attach the desired target (spiculated mass boundary). Which obtain the minimum of snake energy. The energy function of snake evolution governed is defined as below.

$$E = \frac{1}{2}\int_0^1 \left(\alpha \left| \frac{dX}{ds} \right|^2 + \beta \left| \frac{d^2 X}{ds} \right|^2 \right) + E_{ext}(X)\,ds. \quad (1)$$

α and β are tension and rigidity parameters of snake deformation whereas E_{ext} is the internal energy pulling the snake towards the boundaries of desired object (spiculated mass). A general definition of E_{ext} is simply defined as a gradient magnitude of the gray level convolved with a Gaussian filtering. The minimum of snake energy must satisfy the following Euler equation.

$$\alpha X''(s) - \beta X'(s) - \nabla E_{ext} = 0. \quad (2)$$

$-\nabla E_{ext}$ is the external force. Regarding to the limitation of conventional snake [15-17], the improved version of the GVF called the generalized gradient vector flow was proposed to replace the external force term $-\nabla E_{ext}$ in Eq. (2). Xu and Prince [15] introduced spatially varying coefficients to decrease the smoothing effect, as a linear elliptic equation given by

$$v_t - g\left(\left|\nabla f\right|\right)\nabla^2 v - h\left(\left|\nabla f\right|\right)\left(\nabla f - v\right) = 0 \qquad (3)$$

where $g\left(\left|\nabla f\right|\right) = e^{-(\nabla f/k)}$, $h\left(\left|\nabla f\right|\right) = 1 - g\left(\left|\nabla f\right|\right)$ and k is a calibration parameter. Note that the weighting functions $g(.)$ and $h(.)$ depend on the gradient of the edge map. In a case of large gradients, $g(.)$ gets smaller as $h(.)$ becomes larger.

III. METHODOLOGY

The semi-automatic detection of spiculated masses in digital mammography proposed in this research consists of different kinds of steps as below diagram (see in Fig. 3).

Digital mammography

↓

1. Multi-resolution analysis using wavelet transform

↓ An appropriate resolution image

2. Region of interest (ROI) definition

2.1 Background detection using OAT algorithm

↓

2.2 Pectoral detection using combination of hill climbing and region growing

↓

2.3 Breast ratio calculation

↓ The ROI of mammographic image

3. Adaptive histogram equalization

↓ The enhanced ROI

4. ROI clustering using hierarchical fuzzy C-mean

↓ The clustered ROI

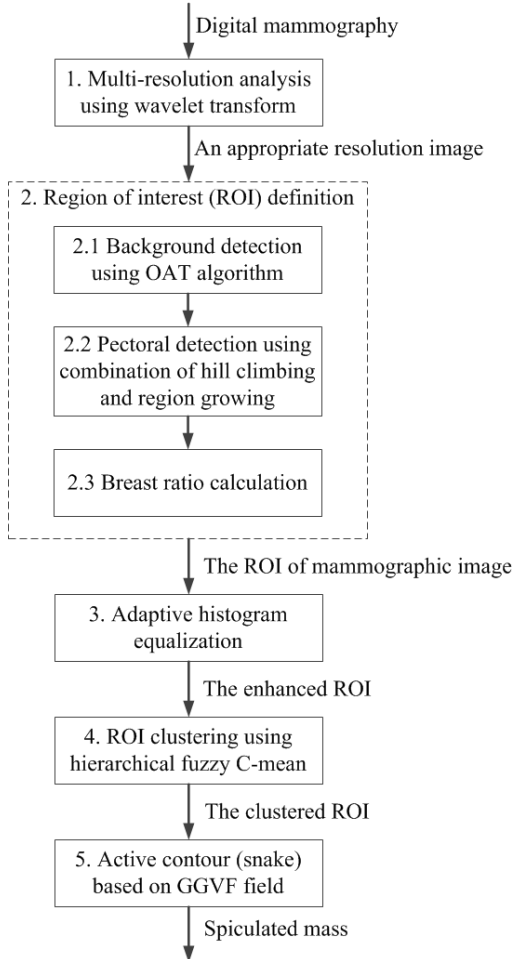5. Active contour (snake) based on GGVF field

↓ Spiculated mass

Fig. 3 Diagram of the proposed method.

1. To obtain a set of different resolution images without affecting significant structures and a noise removal, multiresolution analysis based on the discrete wavelet transform is applied to mammographic images. The next step is to select an appropriate resolution image (approximation sub-band) that unwanted noise is significantly removed.

2. We define the regions of interest (ROI) of breast mammographic image (the results of step 1) by removing background and pectoral muscle sections as a sequence of substeps.

2.1 Backgrounds of digital mammography are defined using the OAT algorithm [6].It recursively employs the Otsu's algorithm to split the image histogram into two classes C_0 (background) and C_1 (object) at threshold value T. The split is performed iteratively; class C_0 is subjected to Otsu's algorithm again. The process continues while $T > A$ and $\sum_{k=0}^{T} H_k > \beta \sum_{k=0}^{G_{MAX}} H_k$, where H is the histogram, G_{MAX} is the maximum gray level, and A is the grey level of background approximation. Finally, β is the calibration parameter used to define some parts of the total number of the image pixels that should belong to the image background. In the resulting image, isolated pixels which theirs gray level are less than the optimal threshold value T_{opt} are marked as non-processing area in the next step.
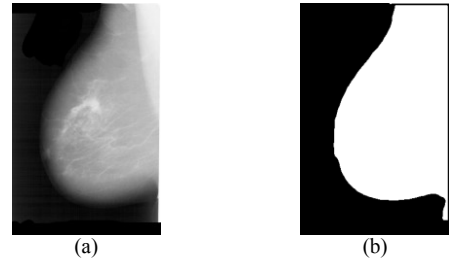

(a)　　　　　　　　　　　　(b)

Fig. 4 Background detection
(a). The digital mammography, (b) Background region

2.2 Pectoral muscles protrude out from each side of breast. Then the combinations of hill-climbing [7] and region-growing [8] algorithms are proposed to detect those areas. A sub-image is extracted from a mammogram image subjected to further k-level wavelet transform and background segmentation. The size of a sub-image is considered as $k*N$ pixels (the first k rows, N : entire columns).A sub-image consists of both of background and non-background areas. As aforesaid, background pixels are not processed anymore. In each row, we apply hill-climbing algorithm including hill-smoothing step with non-background pixels to detect hill area as initial outline of pectoral muscles (see Fig. 5(a), (b)). Different rows hill-climbing are variant in which reaching a local maximum. An initial seed point of region-growing is selected at random from pixels inside the hill. We consider a 4-connected neighborhood surrounding each pixel for growing of region. The region is assumed to be connected pixels with similar intensity.

As aforementioned, regions marking of background and pectoral muscle provide pixels that are not processed in the next step.
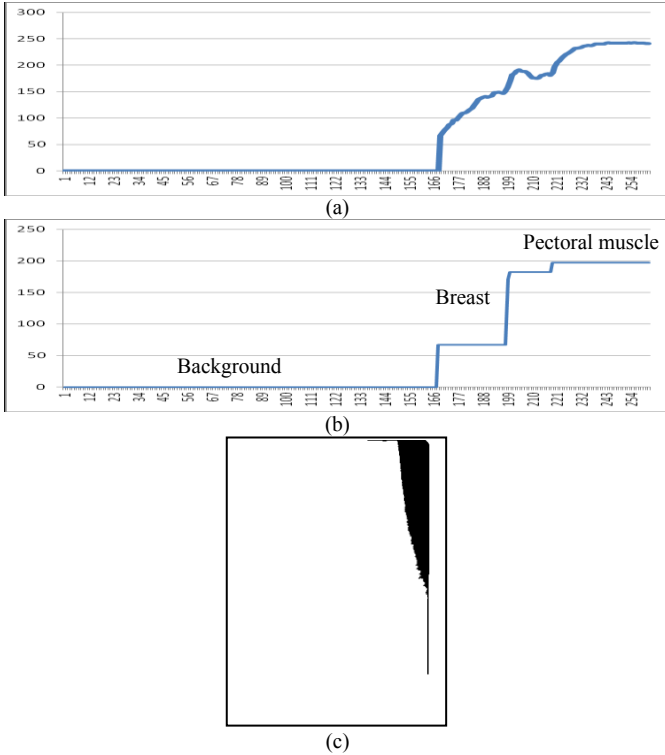
(a)

(b)

(c)

Fig. 5 Pectoral muscle detection
(a) Hill climbing plot original, (b) Hill climbing plot with smoothing,
(c) Pectoral muscle

2.3 The previous step result is divided into overlapped sub-images of size $p*N$ pixels (p : rows, N : entire columns). The rows number overlapped is q . Each sub-image consists of pixels of breast and marked regions (see Fig 6(a)). The volumetric breast density is calculated as an aspect ratio of breast pixels to sub-image size.

$$ B_{Ratio} = \frac{\sum_{x=1}^{p}\sum_{y=1}^{N}\left(I(x,y) \middle/ I(x,y) \right)}{p*N}, \quad where \quad I(x,y) \neq 0. \quad (4) $$

$I(x, j)$ is pixel intensity of ROI image at coordinates *(x, y)* and *I(x, y)* located at marked regions (background and pectoral muscle) equals to 0. Sub images with higher breast ratio that $B_{Ratio} \geq \beta$ indicate the region of interest (ROI) within the mammographic image, otherwise they will be excluded. β is calibration parameter. In case of two sub-images overlapped, the first image is not the ROI but the second one presents characteristics of the ROI. Then, the overlapping rows of two images will be judged as the ROI segment. The result of this step is shown in Fig. 6 (b). The sub-images selection is performed iteratively in columns (Fig. 6(c)). Fig. 6(d) presents the final ROI image
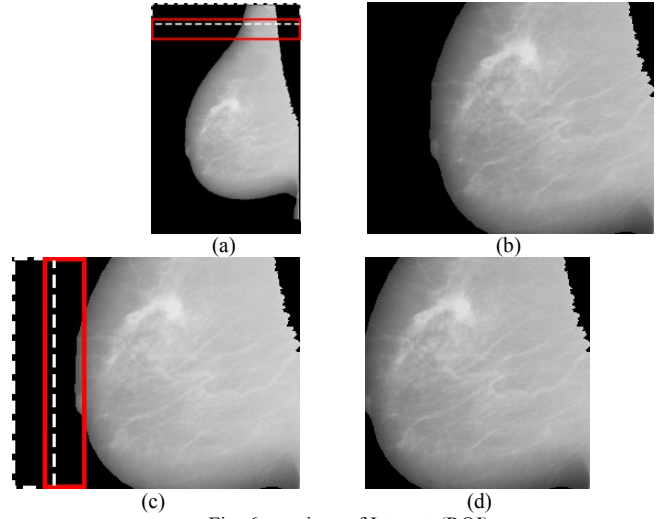

(a)

(b)

(c)

(d)

Fig. 6 a regions of Interest (ROI)
(a) The sub-images selection in rows, (b) The first step ROI
(c) The sub-images selection in columns, (d) The final ROI

3. Preprocessing is a crucial step of the active contour based spiculated mass detection. The contrast-limited adaptive histogram equalization (CLAHE) method [9] is used to improve the ROI's contrast.

4. We partition the ROI using the modified hierarchical fuzzy C-mean clustering [10] based on the top-down dendogram (see Fig. 7(a)). The gray levels of ROI are split into two sub-clusters as lower and higher gray level clusters. The split is performed iteratively; a higher gray level cluster is subjected to fuzzy C-mean clustering again. The process continues while $F_{ratio} > \alpha$. It means that small ratio of between-cluster/within-cluster variance called the F-ratio were kept undivided whereas other conditions are allowed to be split.

In ROI, a particular group of pixels with gray levels in brightest cluster are multiplied by weighting coefficient ω_0 which its value should have more than one. Unwanted components of ROI have an adverse impact on the performance of next active contour step. Then we need to multiply gray levels of pixels in the remaining clusters by ω_1 which $\omega_1 < 1$.
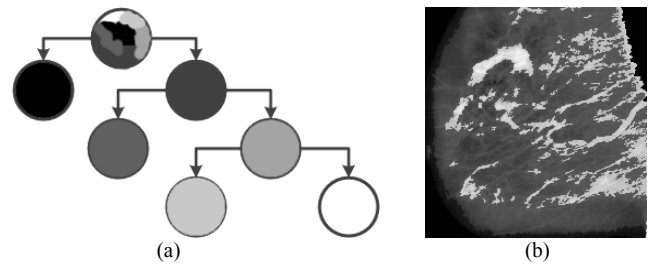

(a)

(b)

Fig. 7 ROI clustering
(a) Clustering dendogram, (b) The clustered ROI

5. The active contour model (snake) based on GGVF field is designed to provide an irregular shape of spiculated mass. First, we apply the GGVF to the raw gradient field ∇f to obtain $V(x,y) = \lfloor u(x,y), v(x,y) \rfloor$ (see Fig. 8(a)) where f is edge map of clustered ROI. An initial snake can be defined specifying it in accordance with the interest (see Fig. 8 (b)). Next, we run an initial snake on the GGVF field. Then, a snake started far from the mass converges to the desired mass. The physical structure (or irregular shape) of spiculated mass can be represented by final contour of snake model (see Fig. 8 (c)). Fig. 8(d) presents a ground truth of spiculated mass supported by the DDSM database.
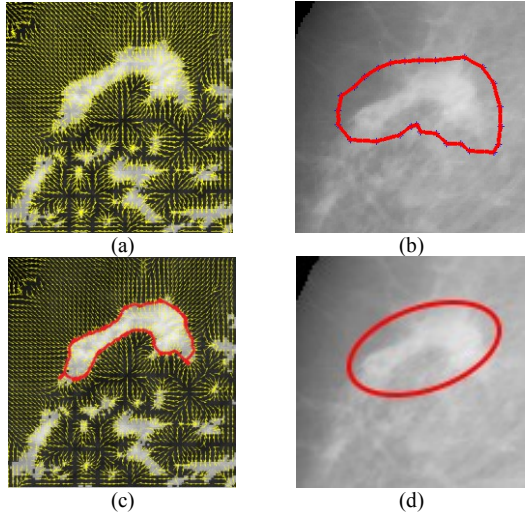

(a) (b)

(c) (d)

Fig. 8 The active contour model (snake)

(a) GGVF field $\left(V(x,y) \right)$, (b) An initial snake,

(c) The final snake, (d) The ground truth

IV. RESULT

Three mammographic images are employed to test our proposed algorithm. The images taken from the DDSM database [11] are shown in Fig. 9(a), Fig. 10(a) and Fig. 11(a). Ground truth information of mammograms is shown in   Fig. 9 -11(g).

Fig. 9(b) and (c) are the marked components and the ROI image respectively. The optimal threshold of OAT algorithm $T = 43$ and an appropriate β of breast ratio is 0.6. The clustered ROI image is shown in Fig. 9(d). Finally, Fig. 9(e) and (f) show GGVF image and the resulting image after performing the active contour model ob GGVF fields.

Fig. 10(b) shows the resulting image after performing a run of OAT algorithm and combination between hill-climbing and growing region algorithms. The threshold T is 45 and breast calibration $\beta = 0.6$. Fig. 10(c) displays the region of interest (ROI). The clustered ROI are shown in Fig. 10(d). The resulting images of active contour model are shown in Fig. 10(e) and (f).
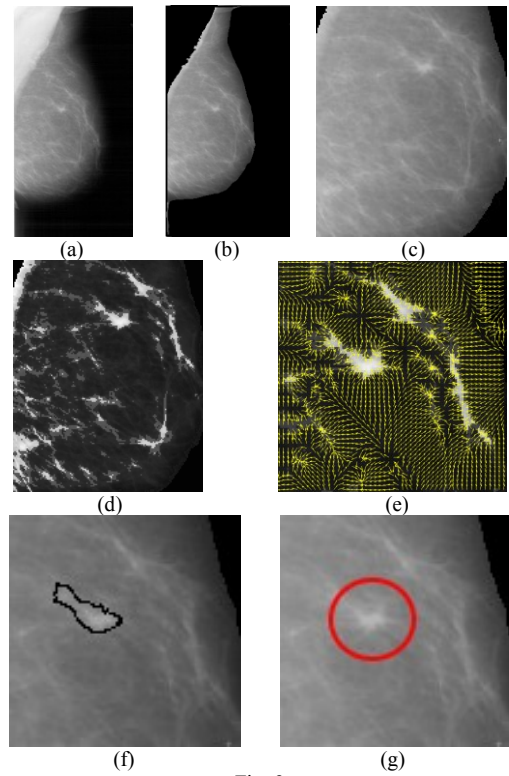

(a) (b) (c)

(d) (e)

(f) (g)

Fig. 9

(a) original mammography, (b) marked regions (background and pectoral muscle regions), (c) the ROI, (d) the clustered ROI, (e). GGVF filed of the clustered ROI, (f) the final snake, (g) The ground truth
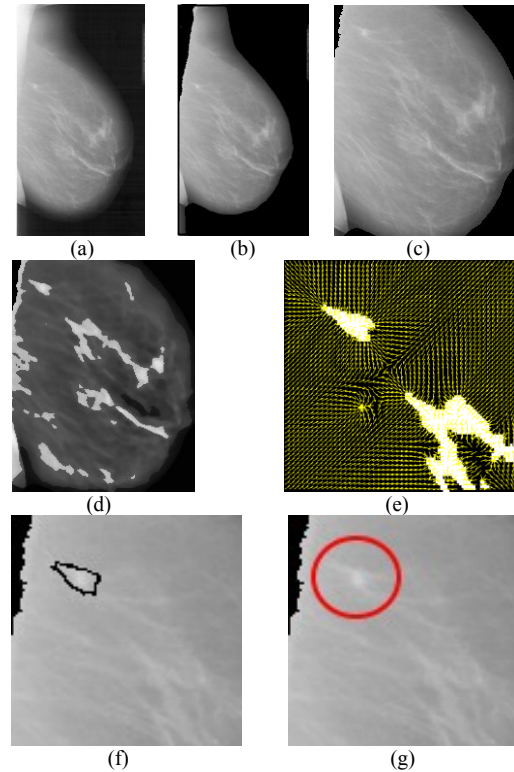

(a) (b) (c)

(d) (e)

(f) (g)

Fig. 10

(a) original mammography, (b) marked regions (background and pectoral muscle regions), (c) the ROI, (d) the clustered ROI, (e). GGVF filed of the clustered ROI, (f) the final snake, (g) The ground truth

Fig.11 shows experimental results on the third mammography performed by the proposed method. T and β equal to 41 and 0.6 respectively. For this experiment (all three images), the diffusion coefficient of GGVF ($K = 0.1$) with 150 iterations. Note that most of irregular shape of spiculation is still present in GGVF images (Fig. 9-11 (e)). Our algorithm produces good results (spiculated mass shape) in dense or low contrast breast images.
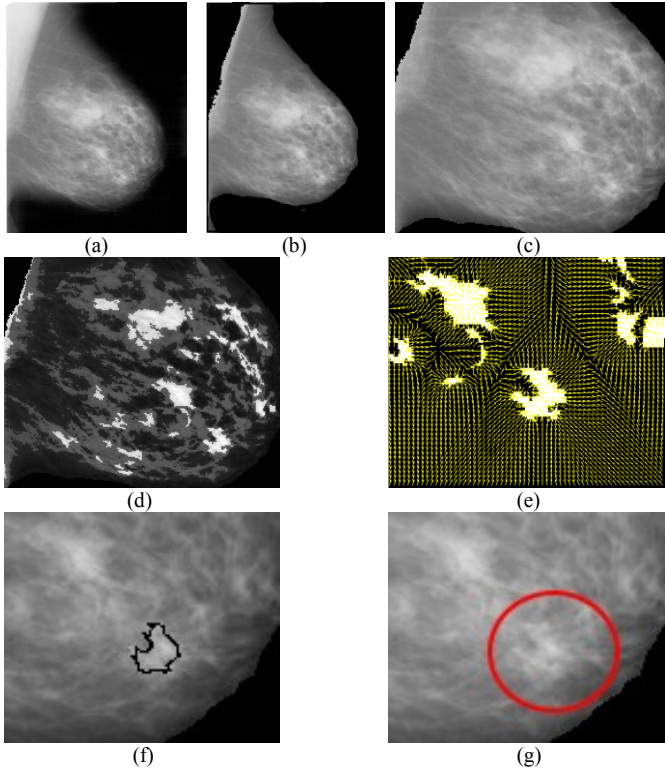

Fig. 11

(a) original mammography, (b) marked regions (background and pectoral muscle regions), (c) the ROI, (d) the clustered ROI, (e). GGVF filed of the clustered ROI, (f) the final snake, (g) The ground truth

V. CONCLUSIONS

In this paper, we propose an algorithm based active contour scheme for spiculated mass detection in digital mammography. To combine efficient techniques, an integrated strategy is presented to extract mass shape with irregularity. Noise reduction is a required step which wavelet transform is proposed. Moreover, multi-resolution concept provides good image localization for next processing. The conventional techniques of object attribute thresholding, hill-climbing and region growing algorithms are combined to define the region of interest (ROI) of mammogram images. Finally, we apply the active contour model to the ROI clustered by a modified hierarchical fuzzy-c-mean to segment the spiculated mass. The experimental results show efficiency and robustness of the proposed algorithm.

REFERENCES

[1] N.R. Mudigonda, R.M. Rangayyan and J.E. Leo Desautels. "Detection of Breast Masses in Mammograms by Density slicing and Texture Flow-Field Analysis", IEEE Transaction on Medical Imaging, Vol.20,No.12 December 2001.

[2] M.P. Sampat and A.C. Bovik, "Detection of Spiculated Lesions in Mammogram", Proceeding of the 25th Annual International Conference of the IEEE. Engineering in Medical and Biology Society (IEMBS), 2003. Vol.1 pp. 810-813.

[3] L. Jiang, E. Song, X. Xu,G. Ma and B.Zheng, "Automated Detection of Breast of Mass Spiculation Levels and Evaluation of Schema Performance", Journal Academic Radiology,2008,Vol.15 No.12, pp. 1534-1544

[4] A Abdel-Dayem and M.R. El-Sakka, "Fuzzy Entropy Based Detection of Suspicious Masses in Digial Mammogram Images", Proceedings of the IEEE Enginerring in Medicine and Biology 27th Annual Conference,2005 pp.4017-4022

[5] F. Zou, IEEE Member, Y. Zheng, Z. Zhou, and K. Agyepong, "Gradient Vector Flow Field and Mass Region Extraction in Digital Mammograms", 21st IEEE International Symposium on Computer-Based Medical Systems,2008 pp.41-43

[6] A. Rodtook, S. Chucherd "Detection of Microcalcifications in Mammograms Using the Object Attribute Thresholding Algorithm", The Stamford journal, Vol 4 (No. 1), June 2012.

[7] S.Kochra and S.Joshi, "Study on Hill-Climbling Algorithm For Image Segmentation", International Journal of Engineering Research and applications, Vol.2(3), 2012,pp. 2171-2174.

[8] R. C. Gonzalez , R. E. Woods, Digital Image Processing (3rd Edition), Prentice Hall; 3 edition (August 31, 2007), ISBN-13: 978-0131687288 ISBN-10: 013168728X

[9] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization", Academic Press Inc., (1994).

[10] Vicenc Torra, "Fuzzy c-means for fuzzy hierarchical clustering", Fuzzy Systems, The 14th IEEE International Conference on, pp. 646-651, May 2005.

[11] DDSM: Digital Database for Screening Mammography. Retrieved July 21,2014 , From the University of South Florida Digital Mammography Home Page website:http://marathon.csee.usf.edu/Mammography/Database.html

[12] S.G Mallat, "Multifrequency channel decompositions of images and wavelet models", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 37, Issue. 12, December 1989, pp. 2091 – 2110.

[13] M. Unser, A. Aldroubi, "Polynomial splines and wavelets – a signal processing perspective", In: Wavelets. A Tutorial in Theory and Application, Academic Press, Boston, pp. 91–122.

[14] A. Rodtook, SS. Makhanov, "A filter bank method to construct rotationally invariant moments for pattern recognitionPattern Recognition Letters", Volume 28, Issue 12, 1 September 2007, Pages 1492-1500

[15] C. Xu, J.L. Prince, "Generalized gradient vector flow external forces for active contours", Signal Processing, vol. 71, issue 2, December 1998, pp. 131–139

[16] A. Rodtook and SS. Makhanov, "Continuous force field analysis for generalized gradient vector flow field", Pattern Recognition, vol.43, Issue 10, October 2010, pp. 3522-3538

[17] A. Rodtook and SS. Makhanov, "Multi-feature gradient vector flow snakes for adaptive segmentation of the ultrasound images of breast cancer", Journal of Visual Communication and Image Representation, vol.24, Issue 8, November 2013, pp. 1414-1430