Automated Segmentation Scheme Based on Probabilistic Method and Active Contour Model for Breast Cancer Detection

Biswajit Biswas, Ritamshirsa Choudhuri and Kashi Nath Dey

Abstract Mammography is one of the renowned techniques for detection of breast cancer in medical domain. The detection rate and accuracy of breast cancer in mammogram depend on the accuracy of image segmentation and the quality of mammogram images. Most of existing mammogram detection techniques suffer from exact continuous boundary detection and estimate amount of affected area. We propose an algorithm for the detection of deformities in mammographic images that using Gaussian probabilistic approach with Maximum likelihood estimation (MLE), statistical measures for the classify of image region, post processing by morphological operations and Freeman Chain Codes for contour detection. For these detected areas of abnormalities, compactness are evaluated on segmented mammographic images. The validation of the proposed method is established by using mammogram images from different databases. From experimental results of the proposed method we can claim the superiority over other usual methods.

Keywords Mammographic images · Segmentation · Breast cancer

1 Introduction and Literature Review

In image analysis, segmentation is the partitioning method for digital image to break into multiple segments. The purpose of segmentation is to simplify or change the demonstration of an image that is more significant for image analysis. Image

B. Biswas (✉) · R. Choudhuri · K.N. Dey
Department of Computer Science and Engineering, University of Calcutta,
Kolkata, India
e-mail: biswajit.cu.08@gmail.com

R. Choudhuri
e-mail: ritam.shirsa@gmail.com

K.N. Dey
e-mail: kndey55@gmail.com

A. Nagar et al. (eds.), *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, Smart Innovation, Systems and Technologies 43, DOI 10.1007/978-81-322-2538-6_57

segmentation is potentially used to locate all the objects and boundaries in images [1]. The performance of an object detection and classification scheme should be consistent and perfect segmentation for the interest of object image [2]. Most important application of image segmentation and classification in mammogram images is for diagnosis of early stage of breast cancer [3]. Mammography is crucial in viewing and diagnosis of breast cancer. So the perfection biopsy using computer-aided methods are advantageous for the detection and classification. We concentrate on image processing for segmentation of breast cancer and estimate the presence of suspicious area based on the image information [4].

There are a number of methods devised towards segmenting human anatomical structures such as variants of active contours [5] and active shape models [6]. The major disadvantages of these methods are they need manual control point initialization and require predefined control [5, 6]. In recent times geometric active contours were established based on theory of curve evolution and geometric flows with level set method [7, 8]. These techniques can detect only the outer boundaries of objects defined by the gradient with well defined structured image. K-means clustering (KM), Fuzzy C means (FCM) clustering are also common for the segmentation of suitable anatomical structures [9]. Markov random field (MRF) based approaches are used for segmentation [10] but uses prerequisite information for segmenting an anatomical structures. Another approach for segmentation is spectral graph clustering method [11], which usually uses for MRI image segmentation [9]. It is a well known robust segmentation method and is commonly used for segmentation applications [2, 4]. A notably different method is proposed in [12]. Commonly a useful image segmentation method is watershed segmentation method that requires exact information about selected pixel set from detected regions [13, 14].

Here we use a probabilistic based segmentation algorithm for the mammogram image to detect breast cancer. This consists of parameter estimation, determine probability distribution of intensity for every pixel in the image, then threshold selection and threshold image creation, morphological processing on segmented region and the active contour model for segmentation of mammogram images. We present a statistical classification method to choose characteristics that satisfies classification condition without from misclassification on observation model [2]. The proposed segmentation approach is applied to various breast cancer images and the results are compared to watershed segmentation method from the related research works. Key steps of proposed method are as follows:

- Find the probability distributions of intensity in image with Gaussian model and estimate Maximum likelihood estimation (MLE) for parameters (Mean and variance).
- Using probabilistic model is prepared a classification method for pixel classification.
- Use morphological operation and Boundary Chain Codes (BBC) for defined active contour map of images.

The organization of the paper is as follows: Sect. 2 deals with background concepts in Probabilistic model, parameter estimation, pixel classification, morphology operations on image, Boundary Chain Codes (BBC) for active contour model and the proposed algorithm has been explained. Section 3 covers the results analysis followed by conclusion on Sect. 4.

2 Proposed Methodology

The lists the various steps of the detection algorithm which are emphasized in detail in the following subsections. The block diagram illustrated in Fig. 1.

2.1 Parameters Estimation and Probability Distribution

The first step of the proposed technique is to estimate the probability density of intensity distribution of each pixel in the input gray scale image $\mathbf{I}(M, N)$ where M, N are number of rows and columns of image $\mathbf{I}$. This task is done by the use of Gaussian probability density function $\mathcal{I}(\mu^{*}, \sigma^{*}2)$ of image $\mathbf{I}$ based on Maximum likelihood estimation of parameters. Here parameters are sample mean μ^{*} and variance $\sigma^{*}2$. Let $x_{i,j}$ denotes the intensity of the pixel at (i, j) in the image $\mathbf{I}(\mathbf{i} \leq \mathbf{M}\,\text{and}\,\mathbf{j} \leq \mathbf{N})$. The value of θ is unknown for the probability density function (PDF) $f_{\mathbf{I}}(x, \theta)$ of image $\mathbf{I}$. We estimate θ based on the sample pixel $\mathbf{x} = x_{i,j}$ where $x_{i,j} \in \mathbf{I}$. We define the following function of θ, with $\mathbf{x}$ for discrete random variable $\mathbf{x} \in \mathbf{I}$, which is called the likelihood function [12]:

$$\mathbf{L}_{\mathbf{x}}(\theta) = p_I(\mathbf{x}, \theta)$$

where any value of θ that maximizes the likelihood function is known as a maximum-likelihood estimate (MLE) and denoted as $\hat{\theta}$:

$$\hat{\theta} = \arg\max_{\theta} L_{\mathbf{x}}(\theta) \tag{1}$$

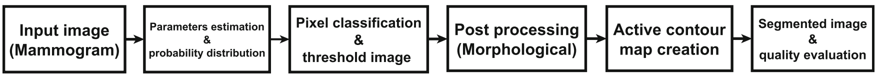

Fig. 1 Main steps of the proposed segmentation scheme for cancer detection

If the likelihood function is differentiable with respect to its parameter, a necessary condition for an MLE to satisfy is

$$\nabla_\theta L_{\mathbf{x}}(\theta) = 0, \text{ i.e., } \frac{dL_x(\theta)}{d_m} = 0, m = 1, 2, \cdots, M \tag{2}$$

where the ∇_θ is the differential operator (1). The logarithmic function is a monotonically increasing and differentiable function, θ that defined the conditions (1) as

$$\frac{\partial \log L_{\mathbf{x}}(\theta)}{\partial \theta_m} = 0, m = 1, 2, \cdots, M \tag{3}$$

We select the solution that gives the largest value of $L_{\mathbf{x}}(\theta)$. Where the function $\log L_{\mathbf{x}}(\theta)$ is called the log-likelihood function, and its partial derivative with respect to θ is called the score function, denoted as $s(\mathbf{x}; \theta)$:

$$s(\mathbf{x}; \theta) \triangleq \left(\frac{\partial \log L_{\mathbf{x}}(\theta)}{\partial \theta_1}, \frac{\partial \log L_{\mathbf{x}}(\theta)}{\partial \theta_2}, \ldots, \frac{\partial \log L_{\mathbf{x}}(\theta)}{\partial \theta_m} \right) \tag{4}$$

The score function specifies the rate at which $\log L_{\mathbf{x}}(\theta)$ changes as θ varies.

Let **n** independent samples are taken from a common Gaussian normal distribution $\mathcal{I}(\mu, \sigma^2)$, where both mean and variance are unknown, and we find an MLE of these parameters based on $\mathbf{x} \in I(i,j)$. By setting $\theta = (\mu, \sigma^2)$, we have the likelihood function

$$L_{\mathbf{x}}(\theta) = \mathrm{I}(\mathbf{x}, \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum_{i=1}^{n} (\mathrm{x}_i - \mu)^2}{2\sigma^2} \right] \tag{5}$$

Then from expression (5) using Eq. (4), yields

$$\frac{1}{2\sigma^2} \sum_{i=1}^{n} (\mathbf{x}_i - \mu) = -n\sigma + \frac{1}{\sigma^3} \sum_{i=1}^{n} (\mathbf{x}_i - \mu)^2 = 0$$

from which we have

$$\mu^* = \frac{\sum_{i=1}^{n} \mathbf{x}_i}{n}, \sigma^2 * = \frac{\sum_{i=1}^{n} (\mathbf{x}_i - \mu^*)^2}{n}, \sigma^2 * \simeq \frac{\sum_{i=1}^{n} (\mathbf{x}_i^2 - \mu^{*2})}{n} \tag{6}$$

where μ^* and $\sigma^2 *$ are estimated parameters.

We used these estimated parameters, viz such as mean μ^* and variance $\sigma^* 2$ in normal Gaussian PDF $\mathcal{I}(\mu, \sigma^2)$ to estimate probability density of every pixel $\mathbf{x}$ of the given image $\mathbf{I}(\mathbf{x} \in \mathbf{I})$. The probability of intensities of image $\mathbf{I}$ are used to pixel classification (background and foreground) and image segmentation for next process. The main necessary steps of parameters estimation is given as follows 1.

Algorithm 1 : Parameters estimation

Input: Gray scale image **I**;
Parameters: Gaussian PDF, unknown parameter μ and σ;
Output: Probabilistic image image **P**;
Step 1: Read gray image image **S**;
Step 2: Estimate meanμ and standard deviation σ using equations (6) on image **I** successively;
Step 3: Evaluate the probability image **P** using equation (5);

2.2 Pixel Classification and Threshold Image

In order to separating a pixel as an object to background and foreground classes, we label the classes π_1 and π_2 where π_1, π_2 are background and foreground class. We assume that if any pixel belong to class π_1 then that pixel assigned by 0 and 1 for class π_2. The pixels are classified on the value of probability $\mathbf{p}$ associated the random variables $\mathbf{x} \in \mathbf{I}(\mathbf{i}, \mathbf{j})$. Thus the background class will be the population of $\mathbf{x}$ values for class π_1 and foreground class as population of $\mathbf{x}$ for class π_2. We represents these two population by probability density function $\mathbf{g_1}(\mathbf{x}) \in \boldsymbol{\pi_1}$ and $\mathbf{g}_2(\mathbf{x}) \in \pi_2$ respectively. Let Ω is collection of all observations of $\mathbf{x}$. We define the sets $\mathbf{R_1} = \mathbf{x}|\mathbf{x} \in \pi_1$ and $\mathbf{R_2} = \mathbf{x}|\mathbf{x} \in \pi_2$ clearly, $\mathbf{R_2} = \Omega - \mathbf{R_1}$.

To classify a pixel object in class π_2 with respect class π_1, we define conditional case, that is

$$\mathbf{P}(\mathbf{X} \in \mathbf{R_2}|\boldsymbol{\pi_1}) = \int_{R_2} f_2(\mathbf{x})\partial x$$

and similarly classify a pixel object in class π_1 with respect class π_2

$$\mathbf{P}(\mathbf{X} \in \mathbf{R_1}|\boldsymbol{\pi_2}) = \int_{R_1} f_1(\mathbf{x})\partial x$$

However an object pixel can be belong to either $\mathbf{R_1}$ or $\mathbf{R_2}$ with basis of a threshold. We estimate threshold parameter η from probability matrix and constructed threshold image based on following rules:

$$\mathbf{R_1} : \sum \frac{\mathbf{g_1}(\mathbf{x})}{\mathbf{g_2}(\mathbf{x})} < \eta|\mathbf{x} \in \pi_1 = 0, \mathbf{R_2} : \sum \frac{\mathbf{g_1}(\mathbf{x})}{\mathbf{g_2}(\mathbf{x})} \geq \eta|\mathbf{x} \in \pi_2 = 1 \tag{7}$$

where η is threshold value defined as $\eta = 1 - \frac{1}{2}\sqrt{\frac{\sigma^*}{\mu^*}}$ with estimated parameters σ^*, μ^*. The overall steps for construction of threshold image represented as follows 2:

Algorithm 2 : Threshold image construction

Input: probability image P;
Parameters: threshold η;
Output: Binary or threshold image $\mathbf{B}$;
Step 1: Read probability image $\mathbf{S}$;
Step 2: Estimating threshold η;
Step 3: Classifying pixel object using (7), (7);
Step 4: Construct threshold image;

2.3 Morphological Image Processing (Post Processing)

The shape based image processing operations known as morphology. The elementary operations of morphology are dilation and erosion. In dilation pixels are added to the boundaries of objects in an image. On the other hand, in erosion pixels are removed from object boundaries [15].

Algorithm 3 : Morphological operation

Input: Binary image B and morphological filter $\mathbf{S}$;
Parameters: Structure elements with size 3×3;
Output: Binary morphological image $\mathbf{M}$;
Step 1: Read binary image $\mathbf{S}$ and morphological filter $\mathbf{T}$;
Step 2: Applying erode and dilate operator (8), (9) on image B in order;
Step 3: Construct binary morphological image $\mathbf{M}$;

The binary dilation of a given binary image $\mathbf{X}$ by structure element $\mathbf{Y}$, denoted $\mathbf{X} \oplus \mathbf{Y}$, is defined as the set operation:

$$\mathbf{X} \oplus \mathbf{Y} = \{p|(\widehat{\mathbf{Y}}_p \cap \mathbf{X}) \neq \phi\} \tag{8}$$

where $\widehat{\mathbf{X}}_p$ is the reflection of the structuring element $\mathbf{X}$ and it is the set of pixel locations $\mathbf{p}$, where the reflected structuring element overlaps with foreground pixels in $\mathbf{Y}$ when translated to $\mathbf{p}$. Similarly, the binary erosion of $\mathbf{X}$ by $\mathbf{Y}$, denoted $\mathbf{X} \ominus \mathbf{Y}$, is defined as the set operation

$$\mathbf{X} \ominus \mathbf{Y} = \{p|(\widehat{\mathbf{Y}_\mathbf{p}} \subseteq \mathbf{X}\} \tag{9}$$

where, it is the set of pixel locations $\mathbf{p}$, where the structuring element translated to location $\mathbf{p}$ overlaps only with foreground pixels in $\mathbf{X}$ [16].

The proposed scheme first applies erode operator on binary image to remove unwanted shapes of the object and dilate is used to operator for filled edge gap in object boundary respectively. Algorithmic steps are as follows 3:

2.4 Active Contour Map Generation (Boundary Chain Codes)

We use Freeman chain code method on binary image to find all contour hierarchy [15, 16]. The contour hierarchy is a polygonal like representation as a sequence of steps in one of eight directions; each step is designated by an integer from 0 to 7 [16]. The closed contour of a region $\mathcal{R}$, is described by the sequence of points $\mathbf{s}_{\mathcal{R}} = [\mathbf{p_0}, \mathbf{p_1}, \ldots, \mathbf{p_{M-1}}]$ with $\mathbf{p_k} = \langle \mathbf{x_k}, \mathbf{y_k} \rangle$. We create the elements of its chain code sequence $\mathbf{s'}_{\mathcal{R}} = [\mathbf{c'_0}, \mathbf{c'_1}, \ldots, \mathbf{c'}_{M-1}]$ by $\mathbf{c'_k} = \mathbf{CODE}(\Delta a_k, \Delta b_k)$ where

$$(\Delta a_k, \Delta b_k) = \begin{cases} (a_{k+1} - a_k, b_{k+1} - b_k), & \text{for } 0 \le k \le M - 1; \\ (a_0 - a_k, b_0 - b_k), & \text{if } k = M - 1; \end{cases}$$

and $\mathbf{CODE}(\Delta a_k, \Delta b_k)$ being defined by Table 1.

For each point on the contour, only the initial point is recorded. The remaining points are encoded by 3 bits. Hence eight directional values can be stored. Algorithmic steps are as follows 4 (Fig. 2):

Algorithm 4 : Chain code operation

Input: Binary image B and a vector of contours $\mathbf{C}$;
Output: Contour image $\mathbf{CI}$;
Step 1: Read binary image $\mathbf{S}$;
Step 2:Get chain coded from image $\mathbf{B}$ using $\mathbf{CODE}$;
Step 3:Draw all retrieved contour on image $\mathbf{B}$;
Step 4: Construct contour image $\mathbf{CI}$;

Table 1 An 8-connected neighborhood

Δa	1	1	0	−1	−1	−1	0	1
Δb	0	1	1	1	0	−1	−1	−1
CODE $(\Delta a_k, \Delta b_k)$	0	1	2	3	4	5	6	7

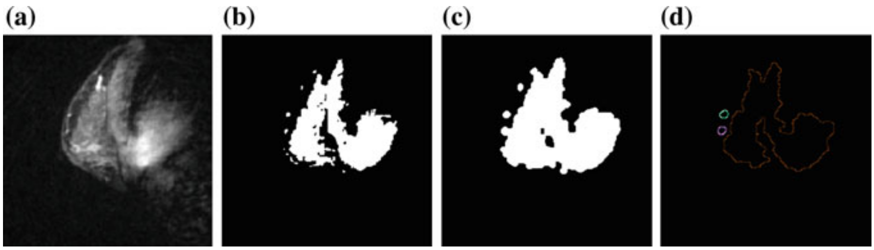

Fig. 2 Results of proposed approach: **a** source image, **b** segmented image, **c** post processing, **d** active contour map

3 Result and Analysis

In this paper all experiments are conducted on images of 30 women diagnosed with breast cancer [17]. The proposed algorithm detects the affected regions in an effective way and the resulting images are shown from Figs. 2, 3, 4, 5, 6 and 7. In each figures the first image is original mammogram, the second image is the threshold image, third is the image after applying morphology (post processing), and last one is the active contour map of cancer detected image. From the detected image, the affected area of breasts could be identified. In order to verify the

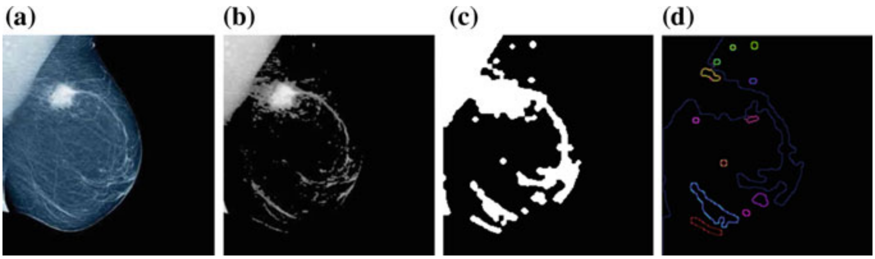

Fig. 3 Results of proposed approach: **a** source image, **b** segmented image, **c** post processing, **d** active contour map

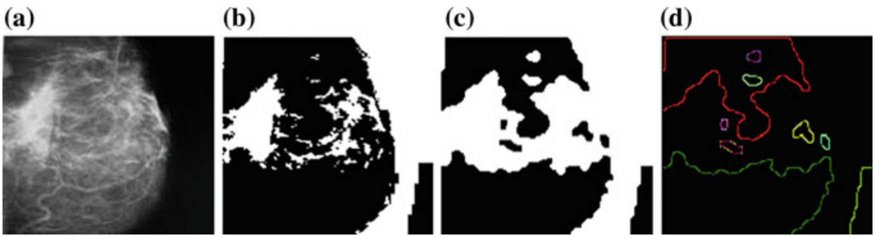

Fig. 4 Results of proposed approach: **a** source image, **b** segmented image, **c** post processing, **d** active contour map

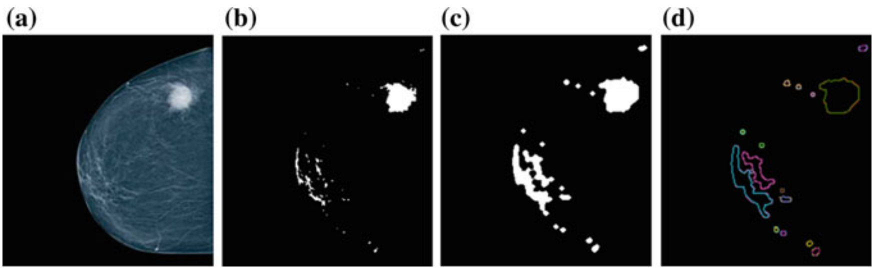

Fig. 5 Results of proposed approach: **a** source image, **b** segmented image, **c** post processing, **d** active contour map

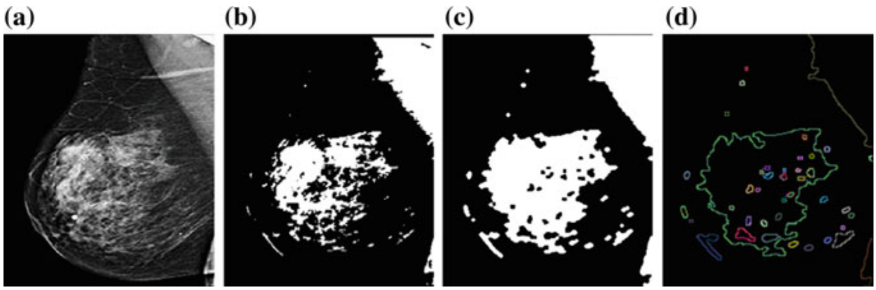

Fig. 6 Results of proposed approach: **a** source image, **b** segmented image, **c** post processing, **d** active contour map

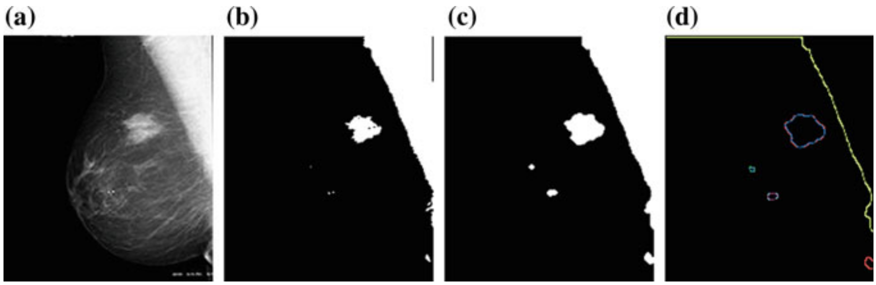

Fig. 7 Results of proposed approach: **a** source image, **b** segmented image, **c** post processing, **d** active contour map

effectiveness of segmentation methods, we used various region descriptors for performance accuracy. These region descriptors are as follows [15, 16]:

- Area of a region in the image plane $\mathbf{A}(\mathbf{R}) = \sum_i \sum_j \mathbf{I}(\mathbf{i}, \mathbf{j})$ where **I** image and **A (R)** the area is measured in pixels.
- The perimeter of an image is $\mathbf{P}(\mathbf{R}) = \sum_i \sqrt{(\mathbf{x}_i - \mathbf{x}_{i-1})^2 + (\mathbf{y}_i - \mathbf{y}_{i-1})^2}$ where x_i and y_i are the co-ordinates of the ith pixel of region.
- Compactness—compactness is an often expressed measure of shape given by the ratio of perimeter to area. That is, $C(R) = \frac{4\pi A(R)}{P^2(R)}$
- Dispersion is the ratio of the maximum to the minimum radius of curve region. That is, $C(R) = \frac{\max\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}}{\min\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}}$ where x_i and y_i are the co-ordinates of the ith pixel of region.

The performance of the proposed segmentation method is tested on benchmark image database [17] and compared with the most popular segmentation algorithm watershed method [13]. From these experiments, we conclude that the proposed scheme in this work presents excellent and accurate segmentation results compared with the watershed algorithm.

Experimentally Tables 2 and 3 has display the performance of proposed segmentation method on different selective benchmark images. Table 2 presents estimation accuracy for individual benchmark image set. For comparative analysis the proposed method and watershed method both are applies on benchmark image sets. The simulation results for proposed scheme w.r.t. the different region descriptor evaluation area (A), perimeter (P), compactness, dispersion are shown in Table 2. In Fig. 8, we have compare the estimated results of proposed method with the existing and popular segmentation technique watershed segment method for the benchmark images (7). From Table 3, it can be observed that proposed segmentation method presents best values (bold faced) w.r.t statistical data of different region descriptor evaluation parameters of segmented output images.

These experiments clearly show that the proposed algorithm excels the limitations of the traditional method by poor structured region and detecting the exact area. The conventional watershed algorithm is suffer from miss-categorization of background and foreground object image in case of poor structured anatomical

Table 2 Assessments of segmentation results of the breast cancer image of proposed scheme

Contour Image	Regional shape descriptors			
	Perimeter	Area	Compactness	Dispersion
Figure 3d	99,978	56,939	0.00073	26.7809
Figure 4d	30,046	45,822	0.00068	67.2463
Figure 5d	174,960	15,515	0.00637	47.9097
Figure 6d	158,690	51,532	0.00026	5.9064
Figure 7d	30,046	45,822	0.00068	67.2463

Table 3 Assessments of regional description parameters of proposed and watershed segmentation scheme

Method	Image	Area	Perimeter	Compactness	Dispersion
Watershed	Figure 8	48,040	38,258	0.00142	42.3725
Proposed	Figure 8	**30,046**	**45,822**	**0.00068**	**67.2463**

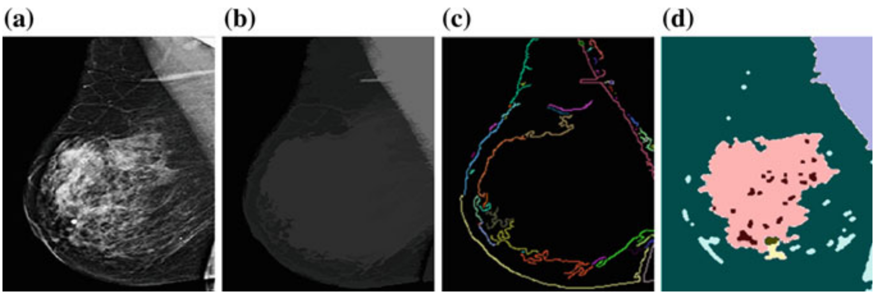

Fig. 8 Results of watershed approach: **a** source image, **b** segmented before watershed, **c** active contour map, **d** watershed image

image. Consequently, the shape and the size of segmented area are fairly accurate. In contrast, the proposed method gives a good partition of image objects without losing any discontinuous boundary. As we can observe in the result sets, the shape and the size of separating boundary of segmented region are detected correctly. In addition, our method gives the precisely continuous and smooth boundaries for image objects.

4 Conclusions

In this paper, we have conceptualized and simulated a probabilistic segmentation algorithm for breast cancer images. To detect abnormalities, the statistical parameters estimation and a classification method is used. All the parameters of the proposed model are estimated automatically from the images. In order to segment, morphological for operation are applied post processing and active contour map for smooth boundary of detected regions. Our method provides a good segmentation of affected cancer regions and smooth borders with continuous active contour map. The experimental results show that aforesaid parameters quantifies the breast abnormality.

References

1. W. Xu, S. Xia, et al., A Model based Algorithm for Mass Segmentation in Mammograms, IEEE, pp. 2543 – 2546, (2006)
2. Cheng, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P., Du, H.N.: Approaches for automated detection and classification of masses in mammograms. Pattern Recogn. **39**(4), 646–668 (2006)
3. Mudigonda, N.R., Rangayyan, R.M., Desautels, J.E.L.: Segmentation and classification of mammographic masses. SPIE Med. Imag. Image Process. **3979**, 55–67 (2000)
4. te Brake, G.M., Karssemeijer, N.: Single and multiscale detection of masses in digital mammograms. IEEE Trans. Med. Imag. **18**, 628–639 (1999)
5. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Int. J. Comput. Vision **1**(4), 321–331 (1988)
6. Xu, C., Yezzi, A., Prince, J.: On the relationship between parametric and geometric active contours. Proc. Asilomar Conf. Sig. Syst. Comput. **1**, 483–489 (2000)
7. Vese, L.A., Chan, T.F.: A multiphase level set framework for image segmentation using the Mumford and Shah model. Int. J. Comput. Vision **50**(3), 271–293 (2002)
8. Brox, T., Weickert, J.: Level set based image segmentation with multiple regions. Pattern Recognit. Springer LNCS **3175**, 415–423 (2004)
9. Chen, W., Giger, M.L., Bick, U.: A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. Acad. Radiol. **13**(1), 63–72 (2006)
10. Li, H.D., Kallergi, M., Clarke, L.P., Jain, V.K., Clark, R.A.: Markov random field for tumor detection in digital mammography. IEEE Trans. Med. Imag. **14**, 565–576 (1995)
11. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000)

12. Fieguth, P.: Statistical image processing and multidimensional modeling. In: Information Science and Statistics. Springer, New York (2011). ISBN:978-1-4419-7293-4
13. Hamarneh, G., Li, X.: Watershed segmentation using prior shape and appearance knowledge. Image Vis. Comput. **27**, 59–68 (2009)
14. Lezoray, O., Cardot, H.: Cooperation of color pixel classification schemes and color watershed: a study for microscopic images. IEEE Trans. Image Process. **11**(7), 783–789 (2002)
15. Nixon, M.S., AguadoFeature, A.S.: Extraction and Image Processing. British Library, UK (2002). ISBN:0-7506-5078-8
16. Burger, W., Burge, M.J.: Principles of Digital Image Processing Core Algorithms. Springer, London (2009). ISSN:1863-7310
17. http://marathon.csee.usf.edu/Mammography/Database.html