

RESEARCH

Open Access

Mass segmentation using a combined method for cancer detection

Jun Liu¹, Jianxun Chen¹, Xiaoming Liu¹, Lei Chun^{2*}, Jinshan Tang^{1*}, Youping Deng³

From BIOCOMP 2010 - The 2010 International Conference on Bioinformatics and Computational Biology Las Vegas, NV, USA. 12-15 July 2010

Abstract

Background: Breast cancer is one of the leading causes of cancer death for women all over the world and mammography is thought of as one of the main tools for early detection of breast cancer. In order to detect the breast cancer, computer aided technology has been introduced. In computer aided cancer detection, the detection and segmentation of mass are very important. The shape of mass can be used as one of the factors to determine whether the mass is malignant or benign. However, many of the current methods are semi-automatic. In this paper, we investigate fully automatic segmentation method.

Results: In this paper, a new mass segmentation algorithm is proposed. In the proposed algorithm, a fully automatic marker-controlled watershed transform is proposed to segment the mass region roughly, and then a level set is used to refine the segmentation. For over-segmentation caused by watershed, we also investigated different noise reduction technologies. Images from DDSM were used in the experiments and the results show that the new algorithm can improve the accuracy of mass segmentation.

Conclusions: The new algorithm combines the advantages of both methods. The combination of the watershed based segmentation and level set method can improve the efficiency of the segmentation. Besides, the introduction of noise reduction technologies can reduce over-segmentation.

Background

Breast cancer is one of the leading causes of cancer death for women all over the world [1] and early detection is one of the main ways to reduce the death rate of the human beings with breast cancer [2-4]. One of the ways to detect the breast cancer is to use mammography. Mammography is thought of as one of the most effective methods to detect early breast cancer. Although mammography is widely used, the rate of correct diagnosis of breast cancer using mammography needs improvement [5]. Thus, in order to improve the diagnosis rate, computer aided diagnosis was proposed

to assist the radiologists in the diagnosis of the breast cancer and used to improve the diagnosis accuracy [6].

In computer aided cancer diagnosis, the detection and segmentation of mass are very important. The shape of mass can be used as one of the factors to determine whether the mass is malignant or benign. In the past, many methods for mass segmentation algorithms have been proposed. These algorithms include manual segmentation [7], semi-automatic segmentation [8], and fully automatic segmentation [9]. Although manual segmentation is considered to be the best mass boundary extraction method [10,11], it is time-consuming. Besides, it subjects to intra-observer and inter-observer variation [11]. In [12], Huo et al. developed a semi-automatic region growing approach based on the choice of the starting point by the radiologist. In [13], Kobatake et al. applied a modified Hough transform to extract lines passing near the centre of the mass and automatically selected candidates based on the number of line-skeletons. In [14], Lou et al.

* Correspondence: Ichuncn@gmail.com; dadatang@yahoo.com

¹College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, Hubei, China

²Key laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China

Full list of author information is available at the end of the article

proposed an algorithm for mass segmentation and the algorithm is based on the assumption that the trace of intensity values from the breast region to the air-background is a monotonic decreasing function. In [15], Zheng et al. proposed an algorithm using the difference image obtained by subtracting the Gaussian filtered image from the original image. In [16], Petrick et al. proposed a method for mass segmentation. The basic idea of the proposed method is to select seeds using local maxima in the original image and generate a gradient image using a frequency-weighted Gaussian filtering. With this image, the thresholds of the regions bounded by the edges are extracted. In [17], Qi and Snyder proposed a method for mass segmentation. They used B'ezier splines to interpolate histograms, from which they extracted the region with threshold values at local maxima. In [18], Guliato et al. proposed a pixel based algorithm. The proposed algorithm aims to preserve the transition between masses and normal tissue to segment the mass boundary. In [19], Mudi-gonda et al. used multilevel thresholding to detect closed edges for mass segmentation. Besides the work mentioned above, there is also other work published in [20-22].

Although many other results on mass segmentation have been published, automatic segmentation of mass is still considered difficult because of the ill-defined boundaries and overlapping with fibro-glandular tissue of many masses [11]. In this paper, we study fully automatic mass segmentation algorithm. Our basic idea is to combine two segmentation algorithms: watershed based segmentation algorithm and level set based segmentation. As is well known, level set based segmentation methods are powerful image segmentation tools and have been used for image segmentation for long time because they have many advantages, for examples, they can handle any of the concavities, splitting, merging and so on. Thus they are still used in many fields including medical image processing [23]. However, there are several disadvantages on level set based segmentation methods. One of the main disadvantages is that the computation is costive. Besides, the level set based algorithms generally need human interaction. In order to reduce the interaction, this paper proposes an algorithm which combines a fully automatic marker-controlled watershed segmentation method with level set based segmentation. In the combined algorithm, the segmentation results from the watershed are used as the input of the level set segmentation and the level set algorithm is used to refine the boundary.

Results

Experimental materials

In the experiments, we selected 200 mammograms randomly from the DDSM database [24] to verify the

proposed algorithm. For reducing computation cost, we resample the original images at a reduced pixel size and 256 gray levels. The mass location was identified by an experienced radiologist and a region of interest (ROI) containing the mass was extracted. The selected samples contain lesions with different breast-tissue density, different degrees of subtlety, and different sizes. The distributions of the size of malignant and benign masses overlapped. 100 of the dataset are benign and 100 of them are malignant.

A program was developed using Matlab to run on all the test images without user intervention. The results show that all cases of segmentation were accurate in comparison with the radiologist-marked on the mammograms. Figure 1 shows some mammograms from DDSM and the segmentation results using watershed transform and level set based segmentation method.

Segmentation evaluation

In the past, there have proposed many segmentation evaluation methods, however, segmentation evaluation is still an open topic [25,26]. There are mainly two evaluation methods. One is subjective evaluation, the other is objective evaluation. In subjective evaluation, visual check is often adopted while the segmentation obtained by the computer is evaluated against the segmentation obtained by a technician in objective evaluation. In this paper, we adopt objective evaluation. The evaluation measures used in the paper are [25]:

$$Hitting = \frac{TP}{TP + FN}$$

$$Missing = \frac{FN}{TP + FN}$$

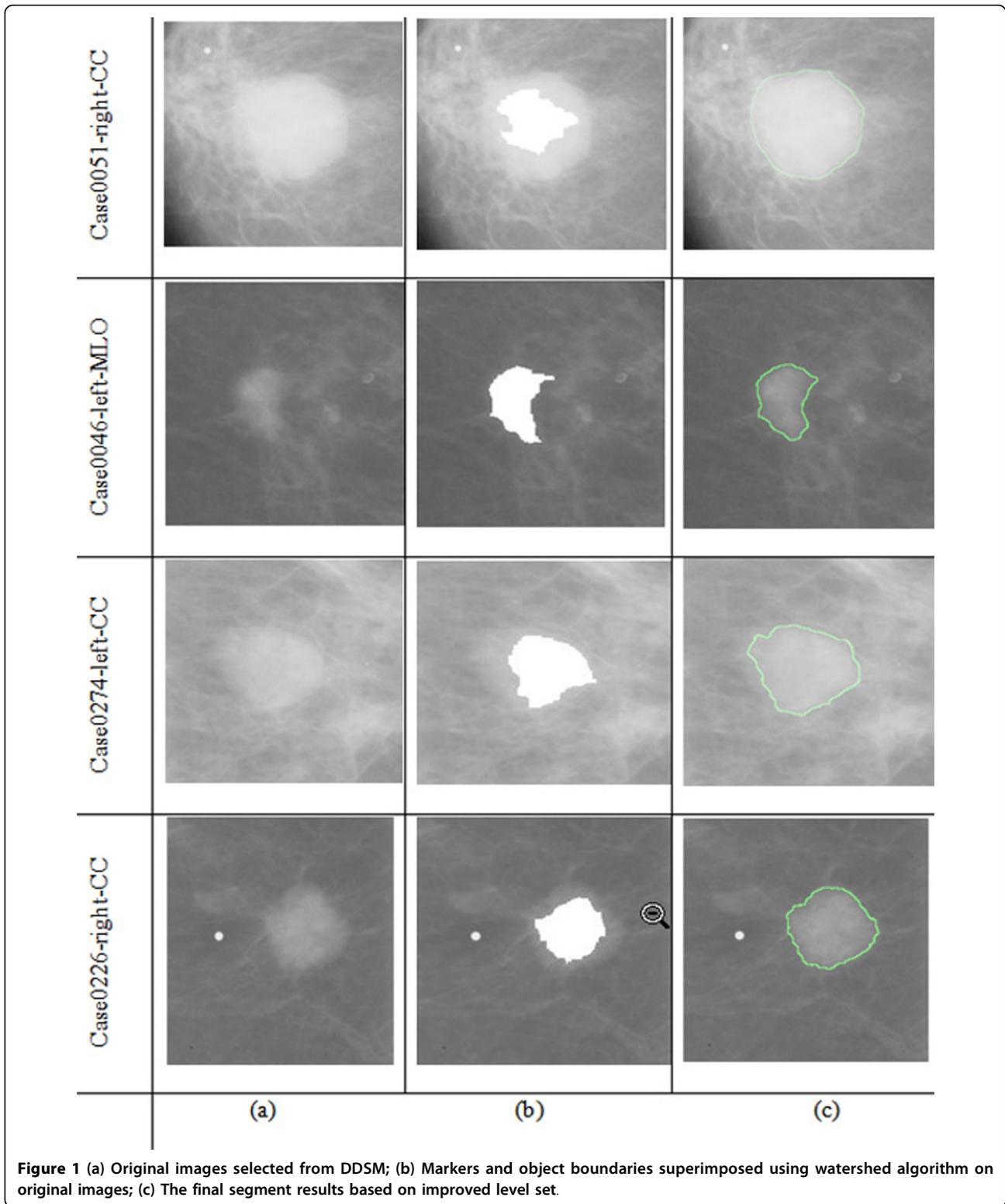
$$OverHitting = \frac{FP}{TP + FN}$$

$$RelativeHitting = \frac{TP}{TP + FP}$$

$$RelativeMissing = \frac{FN}{TP + FP}$$

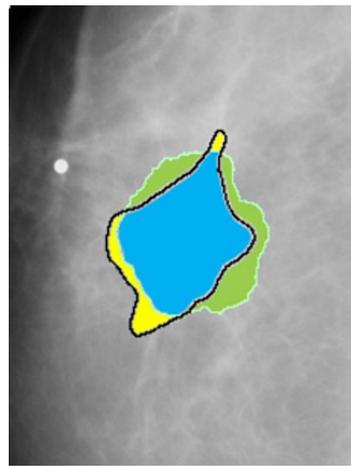
$$Kapps = \frac{2 * Hitting}{2 * Hitting + Missing + OverHitting}$$

where TP, FP and FN are True Positives, False Positives, and False Negatives respectively. Figure 2 shows the basic idea of TP, FP and FN of a mass segmentation. In Figure 2, TP represents the intersection of the radiologist and the algorithm, FP represents the segmentation



results obtained only by the algorithm and the FN represents the segmentation results obtained only by the radiologist [25]. *Hitting* denotes the ratio of correct segmentation, *Missing* denotes the ratio of missing mass,

OverHitting denotes the ratio of false mass segmented, *RelativeHitting* denotes relative correct ratio against segmentation results, and *RelativeMissing* denotes relative missing ratio against segmentation results [25].



-  TP – True Positives
-  FP – False Positives
-  FN – False Negatives
-  Boundary by radiologist
-  Boundary by levelset method

Figure 2 True Positives, False Positives, and False Negatives definition.

Segmentation results

The comparisons of the segmentation results between the proposed method and the manually segmented image by radiologist are shown in Figure 3. In Figure 3, the black contours are the segmentation results using the proposed algorithm and the green contours are the results obtained by a radiologist. From Figure 3, we can find that the proposed method can obtain good results. We can find that the contours obtained by the proposed algorithm are closed to the contours obtained by the radiologist and it proves that the proposed algorithm is effective. Table 1 and Table 2 show the results of quantitative analysis and from the results we can also prove the effectiveness of the proposed algorithm.

Besides the comparison of the proposed algorithm with the human segmentation, we also compared the effectiveness of different noise reduction technologies for over-segmentation reduction. The comparison results are shown in Figure 4. From Figure 4, we can find that effectiveness of average filter is worse than Gaussian filter while Gaussian filter is worse than anisotropic diffusion filter. Anisotropic diffusion filter can reduce the over-segmentation effectively and thus in the proposed algorithm we adopted anisotropic diffusion filter.

Discussion

In this paper, we propose a mass segmentation algorithm which combines watershed method and level set method. The new method is divided into two steps: a marker-controlled watershed transform is first used to segment the mass region roughly, and then a level set is used to refine the segmentation.

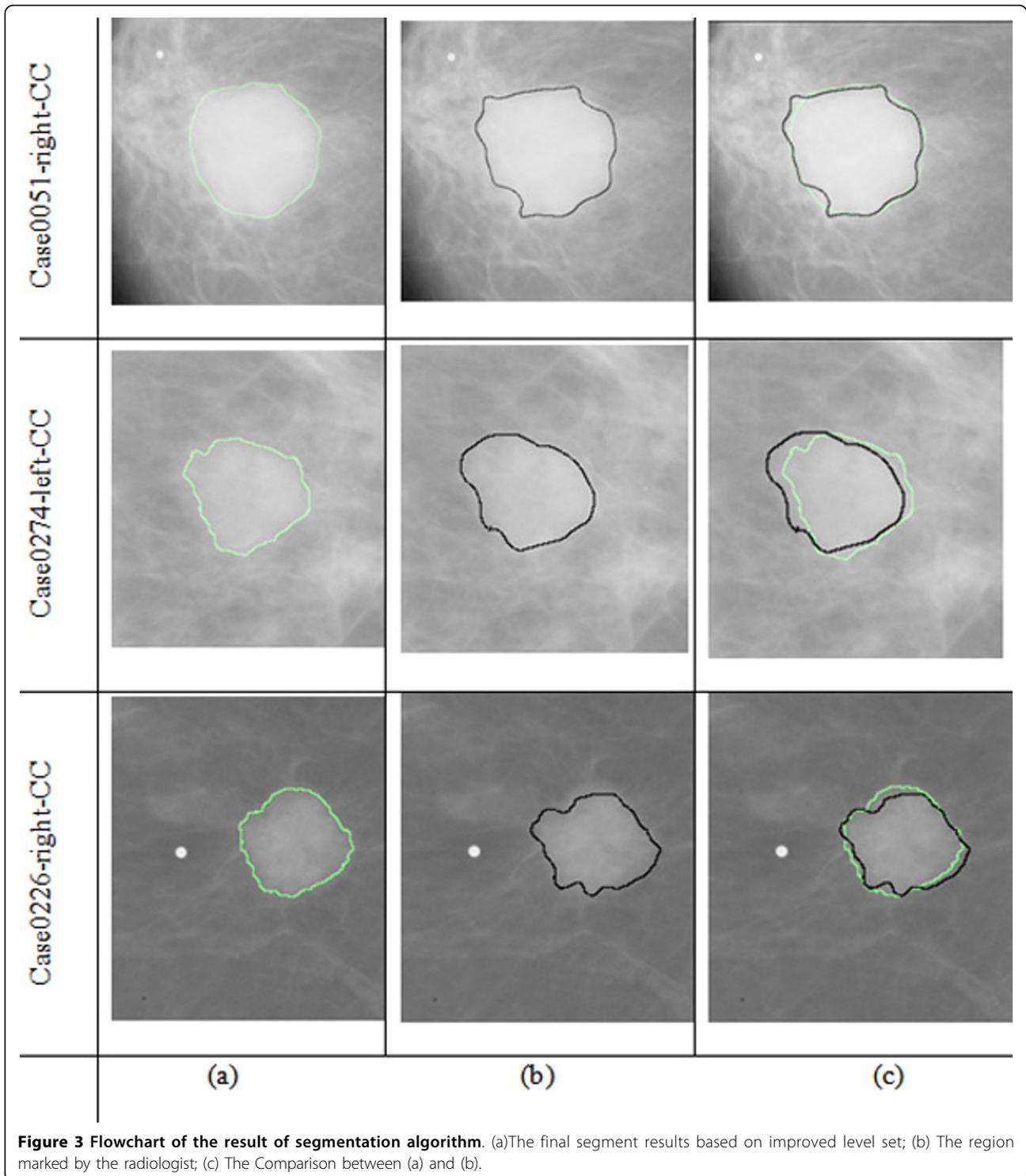
Watershed based segmentation algorithm has many advantages which can overcome the disadvantage in the level set based segmentation. As we know, level set

method usually needs hundreds of iterations to get a good segmentation result. With a good initialization provided by watershed segmentation, the level set method can converge more quickly, thus greatly speed up the whole segmentation procedure. Besides, by using watershed segmentation as the initialization step, we can remove the manual initialization step in general level set segmentation and we can obtain a full automatic segmentation algorithm.

However, the proposed algorithm still has a few limitations. In the proposed algorithm, the object to be segmented is already ROI images which have been preliminarily cut from the whole mammograms. Thus a mass detection step needs to be merged into the algorithm in the future. Although Noise reduction technologies are introduced into the algorithms, over-segmentation still happens on some mammographic images. Over-segmentation affects the efficiency of the algorithm and thus an effective over-segmentation algorithm is needed in the future. Another issue is the time complexity of the level set. By using the result from watershed we can save a lot time but much longer computation time is still needed to achieve the accurate segmentation results.

Conclusions

In this paper, we have developed a hybrid method to segment the mammograms which used watershed algorithm and level set method. We used watershed transform to provide a coarse and fast pre-segmentation, and used the resultant segmentation as the initial contour for the level set segmentation. Automatic selection of the starting point from watershed transform can reduce the user interaction. The combination of the two segmentation methods speeds up the entire segmentation processing and improves the



segmentation efficiency. Besides, the method has good topological adaptability; it can deal with complex and changing shapes of the segmentation of the mammograms well and get high segmentation accuracy. Experimental results show that the proposed segmentation method can obtain good results.

Method

Mass segmentation includes two steps in the proposed algorithm. The first step is to use watershed transform for rough segmentation and the second step is to use level set based method to refine the segmentation obtained by watershed transform. Watershed based

Table 1 The different part Data (pixels) of Fig.3

| CaseNo | TP | FP | FN |
|--------|-------|------|------|
| 0046 | 4517 | 635 | 825 |
| 0051 | 3235 | 370 | 179 |
| 0069 | 2913 | 1475 | 140 |
| 0074 | 12912 | 2611 | 4654 |
| 0123 | 7419 | 1452 | 2566 |
| 0161 | 4339 | 2050 | 858 |
| 0226 | 18834 | 890 | 575 |
| 0274 | 1583 | 704 | 80 |

algorithms are mathematical morphology methods for image segmentation and they have many advantages in comparison with other image segmentation methods. For example, watershed transform based segmentation methods generally have high computation speed and can obtain closed contour lines and accurate position. Besides, watershed based image segmentation algorithms can handle weak edges very well [27].

The basic idea of watershed can be described as follows [27]: let χ be a gray image, $||\nabla\chi||$ is the gradient

Table 2 Validation measure Data (percent) of Fig 3

| CaseNo | Hitting | Missing | OverHitting | RelativeHitting | RelativeMissing | Kappa |
|--------|---------|---------|-------------|-----------------|-----------------|-------|
| 0046 | 0.85 | 0.15 | 0.12 | 0.88 | 0.16 | 0.86 |
| 0051 | 0.95 | 0.05 | 0.11 | 0.90 | 0.05 | 0.92 |
| 0069 | 0.95 | 0.05 | 0.48 | 0.66 | 0.03 | 0.78 |
| 0074 | 0.74 | 0.26 | 0.15 | 0.83 | 0.30 | 0.78 |
| 0123 | 0.74 | 0.26 | 0.15 | 0.84 | 0.29 | 0.79 |
| 0161 | 0.83 | 0.17 | 0.39 | 0.68 | 0.13 | 0.75 |
| 0226 | 0.97 | 0.03 | 0.05 | 0.95 | 0.03 | 0.96 |
| 0274 | 0.95 | 0.05 | 0.42 | 0.69 | 0.03 | 0.80 |

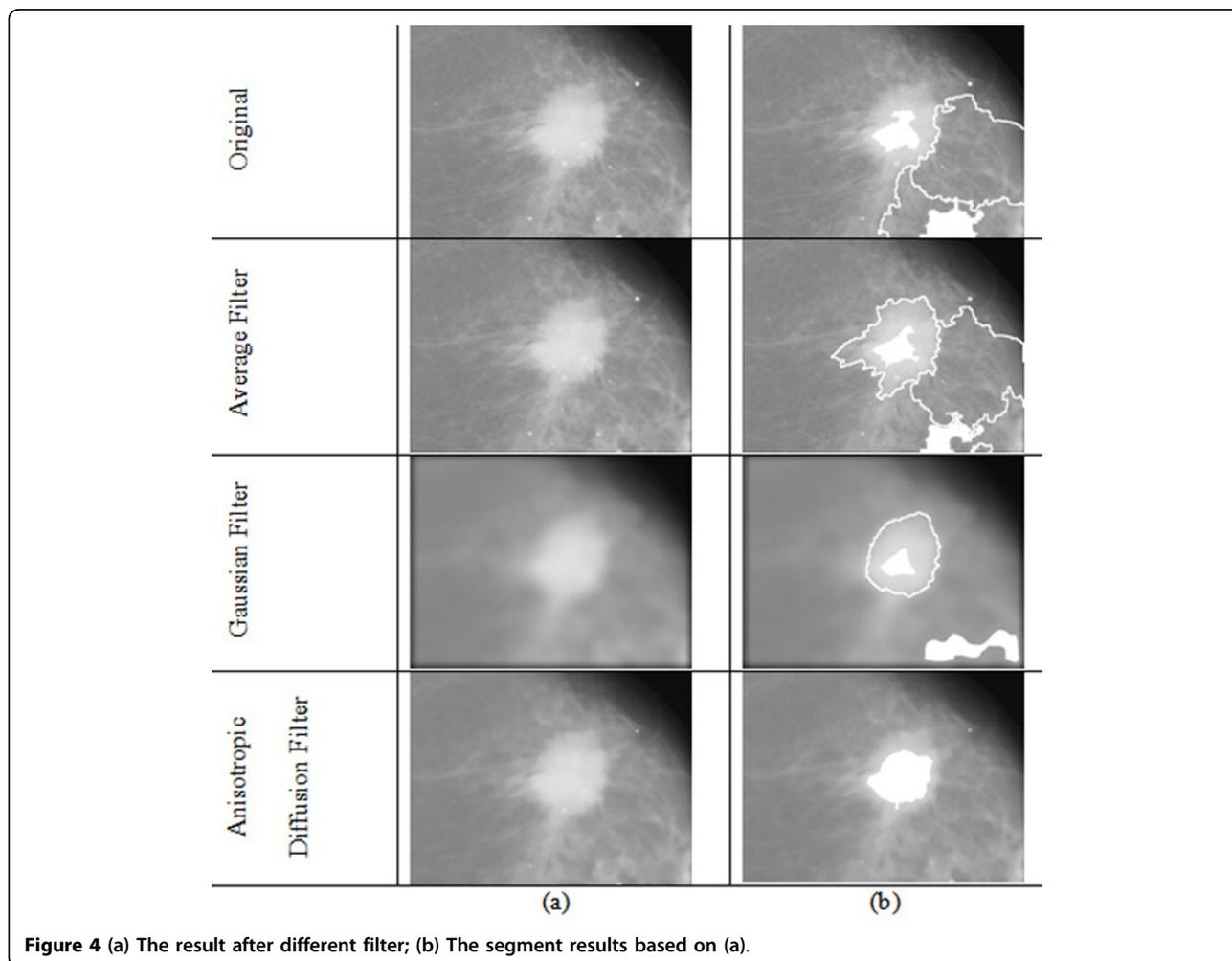


Figure 4 (a) The result after different filter; (b) The segment results based on (a).

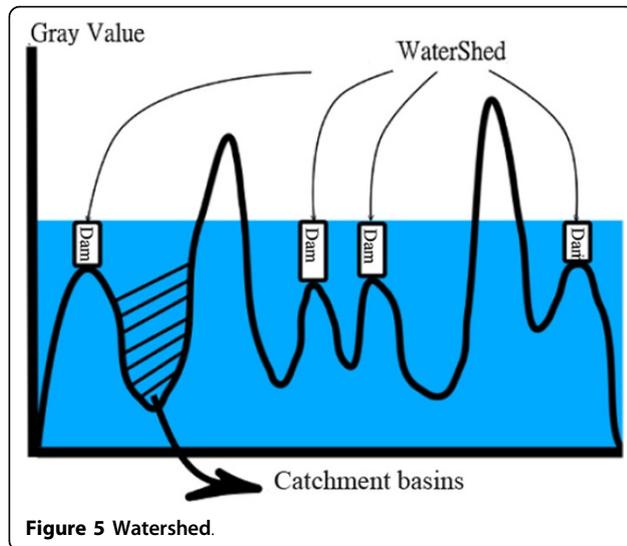


image obtained from χ . In order to segment the objects in the image, the foreground markers will be computed for the objects. After the markers are obtained, the flood waves will propagate from the set of markers to cover the topographic surface $||\nabla\chi||$ [27]. When the water reaches the maximum gray value, the edges of the union of all dams come into being the watershed segmentation. Figure 5 shows the definition of watershed.

In the implementation of the watershed algorithm, if we only use gradient of watershed for segmentation, there are too many ridgelines which will cause over-segmentation (see Figure 6(b)). In order to reduce the over-segmentation, marker-controller watershed is used to reduce over-segmentation. In mark based watershed method, markers are connected through the component. After the marker-based watershed applied, we can get Figure 6(c).

After the image is segmented using watershed transform, we will use the resultant contour as the initial contour for a level set based method to refine the segmentation. The level set algorithm used for the segmentation in the proposed algorithm is from [28]. The level set algorithm proposed in [28] is based on region based

active contour model. This model assumes an image is formed by two homogeneous regions, and can be formulated by the following energy functional [29,30]:

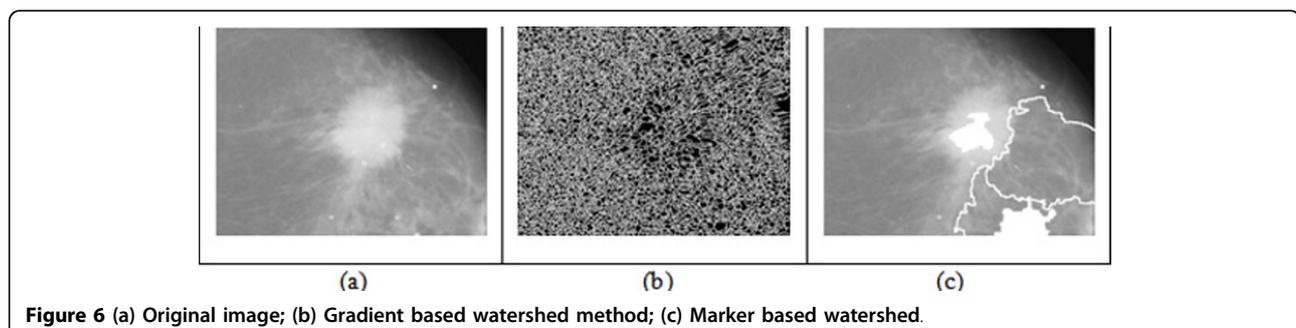
$$E^{CV}(C, c_1, c_2) = \lambda_1 \int_{\text{inside}(C)} |I_0(x, y) - c_1|^2 dx dy + \lambda_2 \int_{\text{outside}(C)} |I_0(x, y) - c_2|^2 dx dy + \mu |C| \quad (\lambda_1, \lambda_2 \geq 0, \mu \geq 0) \quad (1)$$

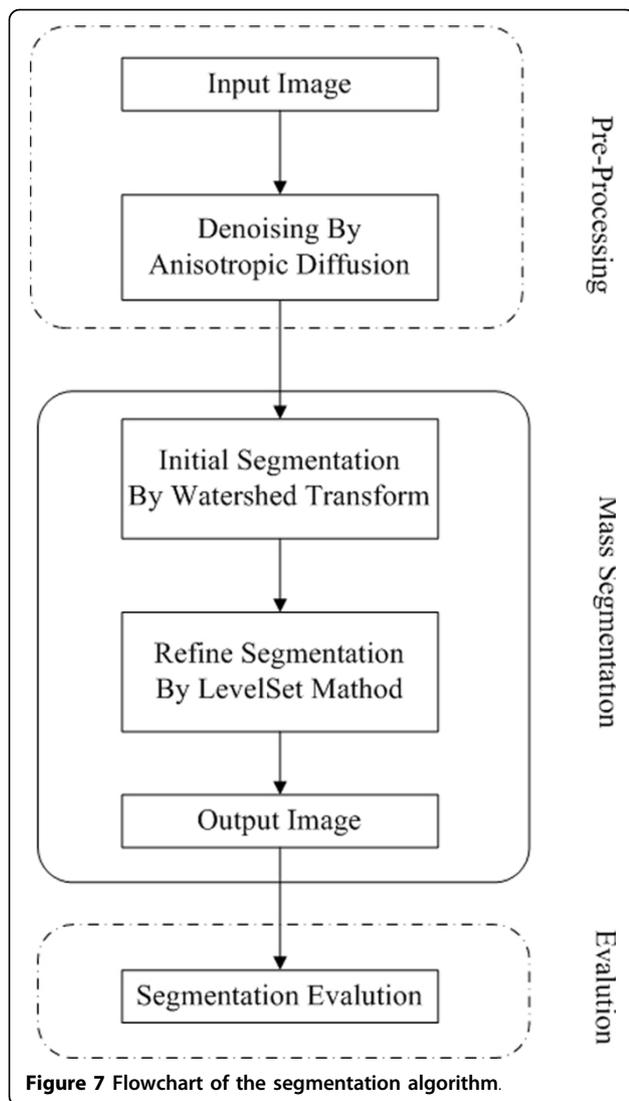
Where $\lambda_1, \lambda_2, \mu, c_1, c_2$ are constants, C is the evolving contour, $|C|$ is the length of contour C , $\text{inside}(C)$ and $\text{outside}(C)$ are the regions inside and outside the contour.

Although the proposed level set method could produce successful segmentation, it needs powerful initialization techniques. In order to solve the problem, in the proposed method, we use the contour obtained from watershed segmentation step as the initial contour of the level set. We resolve the drawbacks of the two methods mentioned above by combining them.

Besides the initialization issue, there is also noise issue. In general, the mammograms have a lot of noise. If the watershed algorithm was applied on the image directly, over-segmentation will happen because the watershed algorithm is very sensitive to noise. To avoid over-segmentation, we need to remove the noise. When the noise is removed, we can get the coarse segmentation using watersheds. The noise reduction methods investigated in the proposed paper include average filter, Gaussian filter and anisotropic diffusion [31]. Anisotropic diffusion was introduced by Perona and Malik [31] and it uses the gradient between the image area to control diffusion degree. Anisotropic diffusion can eliminate the noise effectively while preserve the edge of the image. The anisotropic diffusion used in the proposed algorithm is the method developed in the [32].

The proposed algorithm is shown in Figure 7. It is composed of several steps, the original image will be pre-processed and then used as the input of the watershed segmentation and the rough segmentation is obtained. The rough segmentation will be used as the start contour for the level set segmentation. This approach combines the advantages of the two methods and overcome the disadvantages of each single method: marker-based watershed is rough but fast and the level set segmentation





needs a certain number of iterations, which produces the final, highly accurate, smooth results.

Acknowledgements

The paper is supported by NSFC 61100055, NSF of Hubei Province (NO. 2008CDB345), Educational Commission of Hubei Province (NO.Q20101101), Department of Science and Technology of Hubei Province (NO. D20091102), and Science Foundation of Wuhan University of Science and Technology Project 2011xz019. This article has been published as part of *BMC Systems Biology* Volume 5 Supplement 3, 2011: BIOCOMP 2010 - The 2010 International Conference on Bioinformatics & Computational Biology: Systems Biology. The full contents of the supplement are available online at <http://www.biomedcentral.com/1752-0509/5?issue=53>.

Author details

¹College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, Hubei, China. ²Key laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China. ³Rush University Cancer Center, Rush University Medical Center, Chicago, Illinois, USA.

Authors' contributions

JL, XL, LC and JC developed the algorithm using watershed and level set and wrote the original paper. JT proposed the investigation of over-segmentation issue and revised the paper. YD did data analysis. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Published: 23 December 2011

References

1. American Cancer S: **Breast cancer facts & figures 2007-2008**. American Cancer Society Atlanta, GA; 2007.
2. Tang J, Rangayyan RM, Xu J, El Naqa I, Yang Y: **Computer-aided detection and diagnosis of breast cancer with mammography: recent advances**. *IEEE Trans Inf Technol Biomed* 2009, **13**(2):236-251.
3. Elter M, Horsch A: **CADx of mammographic masses and clustered microcalcifications: a review**. *Med Phys* 2009, **36**(6):2052-2068.
4. Liu X, Tang J, Zhang X: **A multiscale image enhancement method for calcification detection in screening mammograms**. 2009: *IEEE* 2009, 677-680.
5. Chan HP, Sahiner B, Helvie MA, Petrick N, Roubidoux MA, Wilson TE, Adler DD, Paramagul C, Newman JS, Sanjay-Gopal S: **Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study**. *Radiology* 1999, **212**(3):817-827.
6. Sahiner B, Petrick N, Chan HP, Hadjiiski LM, Paramagul C, Helvie MA, Gurcan MN: **Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization**. *IEEE Trans Med Imaging* 2001, **20**(12):1275-1284.
7. Mudigonda NR, Rangayyan RM, Desautels JE: **Gradient and texture analysis for the classification of mammographic masses**. *IEEE Trans Med Imaging* 2000, **19**(10):1032-1043.
8. Kilday J, Palmieri F, Fox MD: **Classifying mammographic lesions using computerized image analysis**. *IEEE Trans Med Imaging* 1993, **12**(4):664-669.
9. Shi J, Sahiner B, Chan HP, Ge J, Hadjiiski L, Helvie MA, Nees A, Wu YT, Wei J, Zhou C, et al: **Characterization of mammographic masses based on level set segmentation with new image features and patient information**. *Med Phys* 2008, **35**(1):280-290.
10. Rangayyan RM, Mudigonda NR, Desautels JE: **Boundary modelling and shape analysis methods for classification of mammographic masses**. *Med Biol Eng Comput* 2000, **38**(5):487-496.
11. Guliato D, de Carvalho JD, Rangayyan RM, Santiago SA: **Feature extraction from a signature based on the turning angle function for the classification of breast tumors**. *J Digit Imaging* 2008, **21**(2):129-144.
12. Huo Z, Giger ML, Vyborny CJ, Bick U, Lu P, Wolverton DE, Schmidt RA: **Analysis of spiculation in the computerized classification of mammographic masses**. *Med Phys* 1995, **22**(10):1569-1579.
13. Kobatake H, Yoshinaga Y: **Detection of spicules on mammogram based on skeleton analysis**. *IEEE Trans Med Imaging* 1996, **15**(3):235-245.
14. Lou SL, Lin HD, Lin KP, Hoogstrate D: **Automatic breast region extraction from digital mammograms for PACS and telemammography applications**. *Comput Med Imaging Graph* 2000, **24**(4):205-220.
15. Zheng B, Good WF, Armfield DR, Cohen C, Hertzberg T, Sumkin JH, Gur D: **Performance change of mammographic CAD schemes optimized with most-recent and prior image databases**. *Acad Radiol* 2003, **10**(3):283-288.
16. Petrick N, Chan HP, Sahiner B, Helvie MA: **Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms**. *Med Phys* 1999, **26**(8):1642-1654.
17. Qi H, Snyder WE: **Lesion detection and characterization in digital mammography by Bezier histograms**. 1999: *IEEE* 1999, 1022:1021-1024.
18. Guliato D, Rangayyan RM, Carnielli WA, Zuffo JA, Desautels JEL: **Segmentation of breast tumors in mammograms by fuzzy region growing**. 1998: *IEEE* 1998, 1002:1002-1005.
19. Mudigonda NR, Rangayyan RM, Desautels JE: **Detection of breast masses in mammograms by density slicing and texture flow-field analysis**. *IEEE Trans Med Imaging* 2001, **20**(12):1215-1227.
20. Dominguez RA, Nandi A: **Toward breast cancer diagnosis based on automated segmentation of masses in mammograms**. *Pattern Recognition* 2009, **42**(6):1138-1148.

21. Song E, Jiang L, Jin R, Zhang L, Yuan Y, Li Q: **Breast mass segmentation in mammography using plane fitting and dynamic programming.** *Acad Radiol* 2008, **16**(7):826-835.
22. Chu Y, Li L, Clark R: **Graph-based region growing for mass-segmentation in digital mammography.** *Proceedings of SPIE* 2002, **4684**:1690-1697.
23. Malladi R, Sethian JA, Vemuri BC: **Shape modeling with front propagation: a level set approach.** *IEEE Trans Patt Anal Mach Intell* 1995, **17**(2):158-175.
24. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer P: **The digital database for screening mammography.** Medical Physics Publishing; 2001, 212-218.
25. Li X: **Automatic image segmentation based on level set approach: application to brain tumor segmentation in MR images.** Université de Reims Champagne-Ardenne; 2009.
26. Zhang H, Fritts JE, Goldman SA: **Image segmentation evaluation: a survey of unsupervised methods.** *Computer Vision and Image Understanding* 2008, **110**(2):260-280.
27. Vincent L, Soille P: **Watersheds in digital spaces: an efficient algorithm based on immersion simulations.** *IEEE Trans Patt Anal Mach Intell* 1991, **13**(6):583-598.
28. Tang J, Liu X: **Classification of mass in mammography with an improved level set segmentation by combining morphological features and texture features.** In *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies. Volume 2.* Springer Verlag;
29. Chan T, Vese L: **An Active Contour Model without Edges.** *Scale-Space Theories in Computer Vision* . In *Lecture Notes in Computer Science. Volume 1682.* Springer; 1999:(1999):141-151.
30. Chan TF, Vese LA: **Active contours without edges.** *IEEE Trans Image Process* 2001, **10**(2):266-277.
31. Perona P, Malik J: **Scale-space and edge detection using anisotropic diffusion.** *IEEE Trans Patt Anal Mach Intell* 1990, **12**(7):629-639.
32. Tang J: **A Multi-direction GVF snake for the segmentation of skin cancer images.** *Pattern Recognition* 2009, **42**(6):1172-1179.

doi:10.1186/1752-0509-5-S3-S6

Cite this article as: Liu et al.: Mass segmentation using a combined method for cancer detection. *BMC Systems Biology* 2011 **5**(Suppl 3):S6.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

