

# An Improved Approach for Breast Cancer Detection in Mammogram based on Watershed Segmentation

Alaa A. Hefnawy  
Electronics Research Institute  
Dokki, Cairo, Egypt

## ABSTRACT

An accurate technique for breast tumor segmentation is a critical step for monitoring and quantifying breast cancer. The fully automated tumor segmentation in mammograms presents many challenges related to characteristics of an image. In this paper, a hybrid segmentation algorithm, which combines the watershed transform and level set techniques, is proposed. Since watershed segmentation is based on pixel density variation that is present in all mass tumors, it was fairly successful in locating tumors under all conditions. However it is very sensitive to small variations of the image magnitude and consequently the number of generated regions is undesirably large and the segmented boundaries are not smooth enough. Meanwhile Level set methods offer a powerful approach for the medical image segmentation since it can handle any of the cavities, concavities, convolution, splitting or merging. However, this method requires specifying initial curves and can only provide good results if these curves are placed near symmetrically with respect to the object boundary. In our proposed technique a watershed segmentation algorithm was developed to initially locate breast mass tumor candidates. In order to facilitate and improve the detection step, the segmentation results is treated as the initial localization of the desired contour, and used in the following level set method, which provides closed, smoothed and accurately localized contours or surfaces. Experimental results show the significant improvement of the final segmentation accuracy.

## General Terms

Image Processing.

## Keywords

Watershed Segmentation; Breast Cancer Mammogram Detection.

## 1. INTRODUCTION

Breast cancer is one of the leading cancers in the female population. About 25% of all cancers diagnosed in women are breast cancers and about 20% of all lethal cancers are breast cancers. It is the leading cause of death due to cancer in women [1]. Because the means to prevent breast cancer have not yet been found, early detection is important. However the appearance of breast cancer is very subtle and unstable in their early stages, therefore doctors and radiologists can miss the abnormality easily if they only diagnose by experiences. In the majority of cases, however, the abnormalities are either micro-calcifications or masses [2]. Micro-calcifications usually form clusters and individual micro-calcifications can range from 20 to several hundred microns in diameter. On the

other hand, a breast mass is a generic term to indicate a localized swelling, protuberance, or lump in the breast. Mammography is a low dose x-ray used for visualizing the internal structure of the breast. Mammography has been proven to be the most reliable method and it is the key screening tool for the early detection of breast cancer [3]. When a mammogram image is viewed, breast tissue appears white and opaque and fatty tissue appears darker and translucent.

The computer aided detection technology can help doctors and radiologists for getting a more reliable and effective diagnoses. There are numerous tumor detection techniques have been analyzed [4][5], including region-based techniques, thresholding-based techniques, edge-based techniques, fuzzy techniques, bilateral image subtraction and multi-scale techniques.

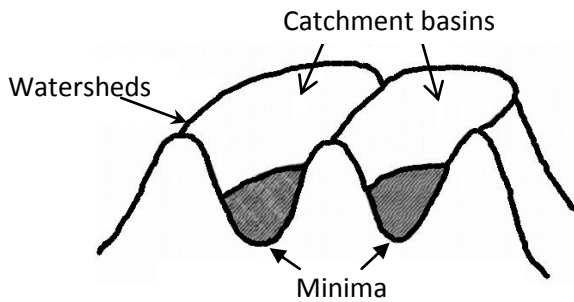
Image segmentation is a process that partitions an image into its constituent regions or objects. Effective segmentation of complex images is one of the most difficult tasks in image processing. Various image segmentation algorithms have been proposed to achieve efficient and accurate results. Among these algorithms, watershed segmentation is a particularly attractive method. The major idea of watershed segmentation is based on the concept of topographic representation of image intensity. Meanwhile, watershed segmentation also embodies other principal image segmentation methods including discontinuity detection, thresholding and region processing. Because of these factors, watershed segmentation displays more effectiveness and stableness than other segmentation algorithms [6].

In this paper, an algorithm belonging to the category of hybrid techniques is proposed, since it results from the integration of morphological watershed transform and level set method [7], and develops robust and flexible object segmentation approach. The output of the watershed detection function is used as rough approximation of the desired regions in the image, and guide for the initial location of the seed points used in the following level set method. Experimental results show that this hybrid method can solve the weakness of each method, and provide accurate, smoothed segmentation results.

## 2. THE WATERSHED TRANSFORM

The watershed transform proposed by Vicent and Soille [8] is a well-known segmentation technique, which is based on immersion simulation, and allows the generation of an initial image partition into regions. It is based on visualizing a gray-level image as topographic surface with three dimensions: two special coordinates versus intensity. In such a topographic surface view, the gray level of a pixel is interpreted as its

altitude. Suppose a water source is placed in each regional minimum (also called 'catchment basins') and the entire topography structure is flooded from below. When water from two sources (regional minima) are about to meet, a dam is constructed to prevent the merging as in Fig. 1. The flooding and dam construction process continues until only the dams are visible from above. These dams (watershed lines or 'watersheds') effectively segment the image into regions.



**Fig. 1: Watershed immersion simulation**

Watershed segmentation has been used in various image processing and computer vision tasks. Sheshadri and Kandaswamy [9] developed a breast cancer detection technique based on watershed segmentation. The study first used a preprocessing step to remove or attenuate the curvilinear structures present and then applied segmentation on gradient images. Ramani et al. [10] compared two different semi-automated segmentation methods, i.e., level set bias field corrected image segmentation method and watershed method. The study found that the level set segmentation gives better image quality than watershed approach. Lewis and Dong [11] satisfied a high detection rate of the mass tumors by selecting foreground and background markers, and then applied watershed segmentation algorithm to isolate a tumor region from its surrounding tissue. Xu et al. [12] applied watershed transformation to the smoothed gradient image to obtain the lesion boundary between the internal and external markers. The watershed approach showed better performance than dynamic programming boundary tracing method and the plane fitting and dynamic programming method. Ponraj et al. [13] proved that, for mass detection, watershed segmentation yields an accurate result compared to morphological component analysis method. They used extracted features from tumors to compare both techniques.

The strength of watershed segmentation is that it produces a unique solution for a particular image, and it can be easily adapted to any kind of digital grid and extended to  $n$ -dimensional images and graphs. Furthermore, in mammograms, a mass tumor region is brighter and has more uniform intensity than its surroundings, which makes a good candidate for watershed segmentation. However, direct watershed segmentation on mammograms often generates unreliable results; the noise in the image results in over segmentation. Another disadvantage of watershed segmentation, again related to the image noise and the image's discrete nature, is that the final boundaries of the segmented region are lack of smoothness. So it is not an efficient idea to treat the watershed segmentation as the final segmentation.

### 3. THE LEVEL SET METHOD

Since its inception [7], the level set method, originally used as numerical technique for tracking interfaces and shapes, has

been used in well wide fields including the medical image processing [14]. It offers a powerful approach for the image segmentation since it can handles any of the cavities, concavities, splitting/merging, and convolution. However, despite of all the advantages which this method can provide, it requires the prior choice of the most important parameters such as the initial location of seed point, the appropriate propagation speed function and the degree of smoothness.

This method, contours or surfaces are represented as the zero level set of a higher dimensional function, usually called a level set function. With the level set representation, the image segmentation problem can be formulated and solved in a principled way based on well-established mathematical theories, including calculus of variations and partial differential equations (PDE) [15]. An advantage of the level set method is that numerical computations involving curves and surfaces can be performed on a fixed Cartesian grid without having to parameterize these objects. Moreover, the level set method is able to represent contours/surfaces with complex topology and change their topology in a natural way.

$X \equiv (x, y)$  The initial position of the  $C_0(X)$  moving interface (where  $\emptyset$ ) is embedded as the zero level  $C_0$  set

of a higher dimensional function, the signed distance to  $\emptyset$ , and link the evolution of this new  $\emptyset$  function to  $C(t)$  the evolution of the interface  $\emptyset$  itself through a time-dependent initial value problem. At each time, the contour is given by the zero level set of  $C(t)$ .

$$(C(t), t) = 0 \Rightarrow \tau + \nabla \cdot (C(t), t) \cdot \frac{\partial C}{\partial t} = 0$$

$$(1) \quad \frac{\partial C}{\partial t} = F_n \quad \text{condition states that}$$

$$\frac{\nabla_{\emptyset}}{|\nabla_{\emptyset}|} \quad \text{Since } \emptyset \text{ and outward normal vector is given by}$$

$$\tau + F|\nabla| = 0$$

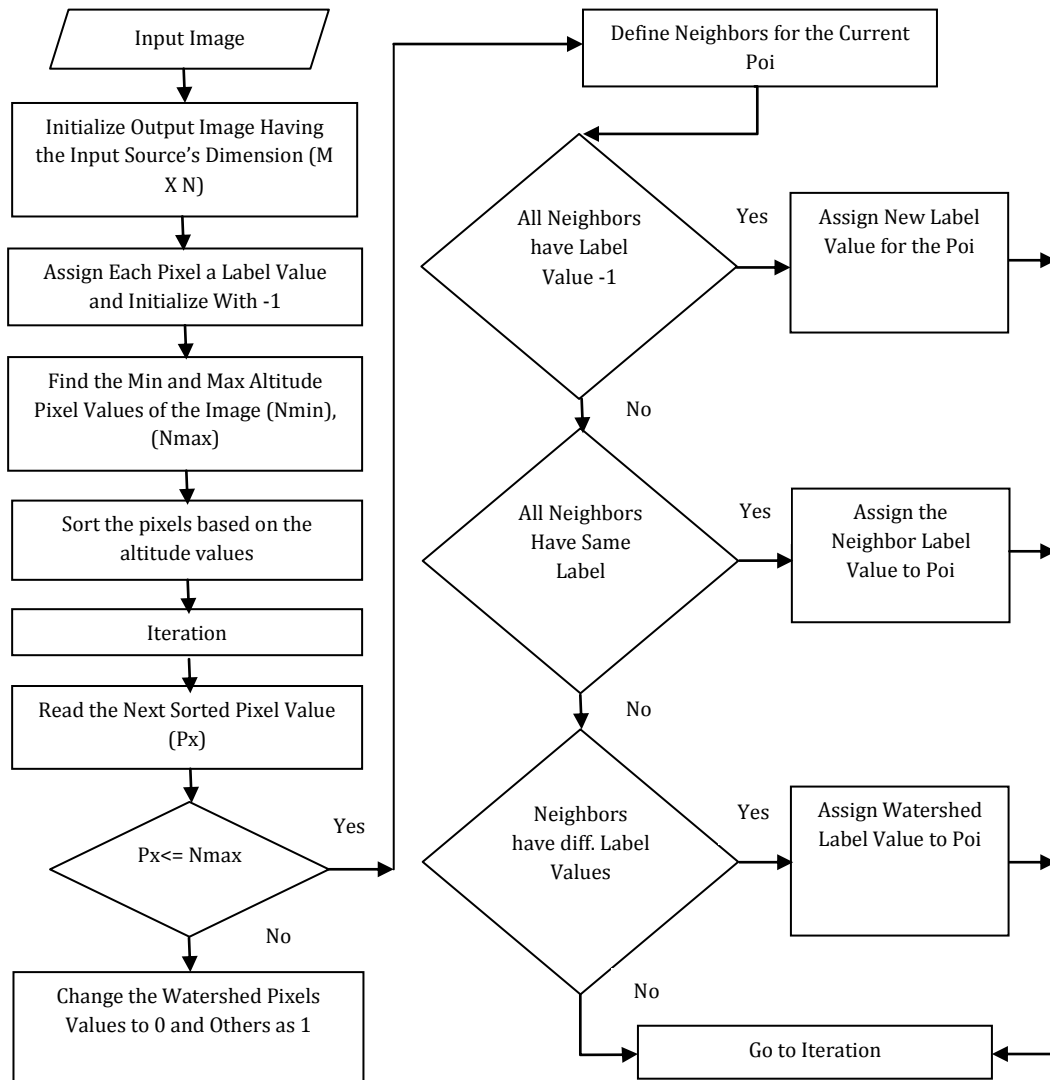
, this yields the following evolution equation for :

$$(X, 0) = C_0(X)$$

$$(2)$$

For reasons of causality, it is possible to restrain the computation domain to a band of cells around the zero-level set of  $\emptyset$ , for the decrease of the computational cost. Classical approaches are referred to as narrow-band methods.

According to the above discussion, the level set method requires specifying initial curves and can only provide good results if these curves are placed near symmetrically with respect to the object boundary. When using the level set method in image segmentation, an initial front should be chose appropriately, and let it propagate with a speed function that stops the motion when the boundary is reached. So the initial  $\emptyset(X; t)$  front and the speed function are important ones to decide the accuracy of the final segmentation. Thus, level set segmentation is not sufficient for the segmentation of complex medical images; they must be combined with powerful initialization techniques in order to produce successful segmentation. We proposed a rough approximation to the desired region in the image based on the watershed algorithm which then smooth the resulted region boundaries based on level set method. The next section gives the detailed algorithm.



**Fig. 2: Flow chart of Watershed segmentation**

#### 4. THE PROPOSED TECHNIQUE

Except for the preprocessing stage, our segmentation strategy will consist in using watershed transform as a pre-segmentation tool, and then refine the segmentation result with the level set method. This approach combines the advantages of both methods: the watershed transform pre-segmentation is rough but quick, and the level set needs only a few iterations to produce the final, fast, highly accurate, and smooth segmentation.

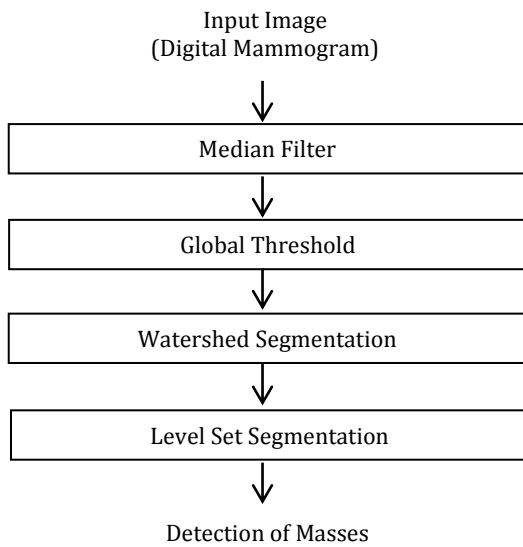
The goal of image preprocessing is to remove noise; black background and unwanted artifacts (such as medical labels) from mammograms as much as possible. Thus, to find breast regions and cropped out the unwanted image portions, a

global thresholding is applied to segment the mammogram. Then watershed algorithm is applied to the image data set. Among the existing watershed algorithms, the immersion-based approach of Vincent and Soille's [8] watershed algorithm was used because of its accuracy and speed of computation. Figure 2 shows the flowchart of the used watershed segmentation algorithm. The output of the watershed algorithm is a partitioning of the input data in volume regions of which the interior does not contain any sharp gray value transitions. As we know, the algorithm leads inevitably to an over segmentation of the data because all the crest lines of the data set are detected. Therefore, the noise filtering preprocessing (using median filter) need to be applied to the image data first (see Fig. 3).

After the initial segmentation that based on watershed transform, the final segmentation is accomplished based on level set method. By combining watershed transform and level sets, this method is able to produce highly accurate segmentations of topologically and geometrically complex structures in much less time than where level sets alone.

The detailed algorithm is as the following:

1. Smooth the original mammogram image by median filters to reduce the noise.
2. Global thresholding; to crop out the unwanted image portions.
3. Watershed segmentation; for initial region extraction.
4. Smooth the boundary of the regions of interest using the level set method



**Fig. 3: Proposed Mass Detection Process**

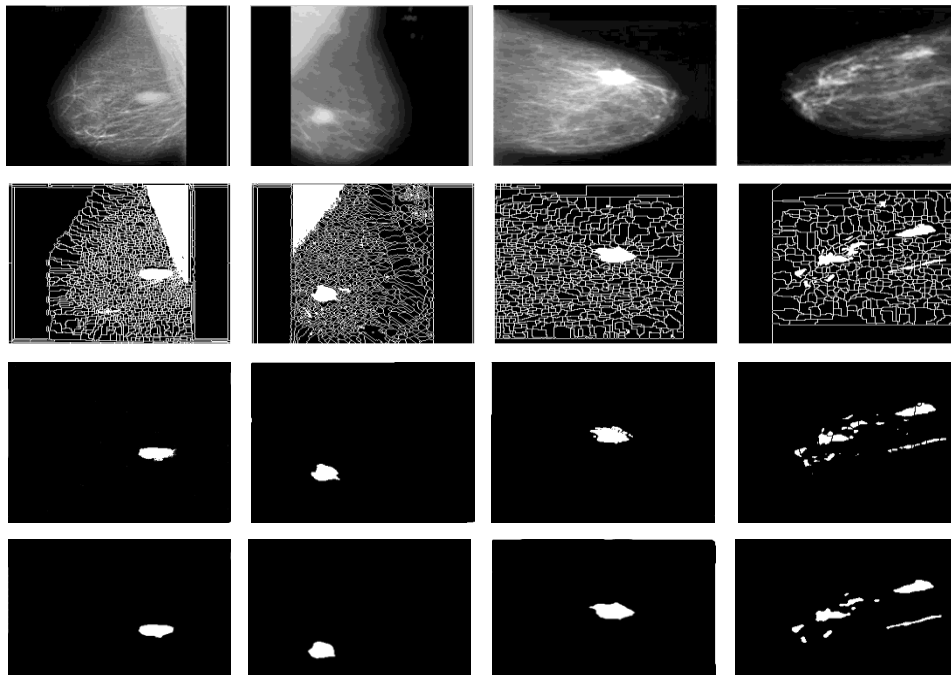
## 5. EXPERIMENTAL RESULTS

The image collection used in this study comes from the Mammographic Image Analysis Society (MIAS) database [16]. Cancerous mammograms images were taken for processing and the results are discussed below. Original images with 1024 by 1024 matrix are taken into consideration. Figure 4 shows the original mammograms, used for mass detection, with tumor locations correctly identified. Four original mammograms of different cases are shown at the top row. Based on immersion algorithm, watershed Image is obtained which results in efficient segmentation of regions of interest as shown in the second row. Watershed image is further processed to extract the region with mass. The third row shows the resultant segmented images. The final enhanced segmentation results, after applying the level set approach, are shown at the bottom row.

In some cases, the segmented region matched the mass tumor very well, e.g. the first column. In other cases, the segmented regions were either bigger, e.g. the second column, or smaller, e.g. the third column, than actual mass tumor sizes. These were generally because watersheds were either not restrictive enough or too restrictive. A very obvious improvement could be notice after applying the level set approach. Another issue shown was that there were multiple false positive regions located in the segmented image, e.g. the fourth column. As the primary goal for this work was to locate all real tumors, false positives will be removed later. That is part of our ongoing work.

## 6. CONCLUSIONS

This paper has highlighted a semi-Automated method for detection of cancer tumors in breast by digital mammogram image analysis. A developed fast hybrid segmentation algorithm has been proposed, which integrates watershed transform and level set method. Both methods have a common advantage; they have no constraints or hypothesis on topology, which may change during convergence.



**Fig. 4: Examples with tumor correctly located; top row: the original mammogram images, second row: Watershed images, third row: segmented images (initial region extraction), bottom row: final segmentation (enhanced region boundaries using level set method).**

of the pre-segmentation (for the level-sets forces): looking at the local distribution of the different regions extracted with the competitive front, we are able to give the initial values of the region descriptors.

In the domain of medical image analysis, time-consuming algorithms are synonymous with non-interactive methods, and are therefore limited to a very small number of specific applications. In others words, the computation times of level sets are an obstacle to a wider use of these methods. The combination of the two minimizes user interaction and speeds up the entire segmentation process. The proposed algorithm produced very satisfactory results both with respect to segmentation performance and execution times.

In our future research, we would also like to remove false positives from the results. As the goal of the current segmentation step is to find all mass candidates even with some false positives, a natural step after segmentation will be to remove false positives as much as possible.

## 7. REFERENCES

- [1] <http://www.cancer.gov/>, The American College of Radiology (ACR), <http://www.acr.org/>.
- [2] H. D. Cheng, X. D. Cai, X. W. Chen, L. M. Hu, and X. L. Lou, "Computer-aided Detection and Classification of Micro Calcifications in Mammograms: a Survey", *pattern recognition*, Vol. 36, pp. 2967–2991, 2003.
- [3] American Cancer Society, "Breast Cancer Facts and Figures", 2007–2008.
- [4] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai and H. N. Du, "Approaches for Automated Detection and Classification of Masses in Mammogram", *Pattern Recognition*, Vol. 39, No. 4, pp. 646-668, 2006.
- [5] M. P. Sampat, M. K. Markey, and A. C. Bovik, "Computer-aided Detection and Diagnosis in Mammography", *Handbook of image and video*, pp. 1195-1217, 2005.
- [6] S. Gupta, M. Sadiq, M. Gupta, and N. Rao, "Semi-automatic Segmentation of Breast Cancer for Mammograms Based on Watershed Segmentation", *Proceedings of the 5th National Conference; Computing For Nation Development*, 2011.
- [7] S. Osher, J. A. Sethian, "Fronts Propagating with Curvature Dependent Speed: Algorithms based on the Hamilton-Jacobi Formulation", *Journal of Computational Physics*, Vol. 79, pp. 12-49, 1988.
- [8] L. Vincent, P. Soille, "Watersheds in Digital Spaces: An Efficient Algorithm based on Immersion Simulation", *IEEE Trans. PAMI*, Vol. 13, No. 6, pp. 583-598, 1991.
- [9] H. S. Sheshadri, A. Kandaswamy, "Detection of Breast Cancer Tumor based on Morphological Watershed Algorithm", *ICGST-GVIP Journal*, Vol. 5, Issue 5, 2005.
- [10] R. Ramani, N. S. Vanitha, and S. Valarmathy, "A Comparative Study of Algorithms for Breast Cancer Detection in Mammogram", *European Journal of Scientific Research*, Vol. 91, No. 1, pp. 100-111, 2012.
- [11] S. H. Lewis, and A. Dong, "Detection of Breast Tumor Candidates Using Marker-controlled Watershed Segmentation and Morphological Analysis", *Proceedings of the IEEE southwest symposium on Image Analysis and Interpretation*, April 2012.
- [12] S. Xu, H. Liu, and E. Song, "Marker-Controlled Watershed for Lesion Segmentation in Mammograms", *Journal of Digital Imaging*, Vol. 24, No.5, pp. 754-763, 2011.
- [13] D. N. Ponraj, M. E. Jenifer, P. Poongodi, and J. S. Manoharan, "Analysis of Tumor Characteristics Based on MCA Decomposition and Watershed Segmentation", *International Journal of Computer Applications*, Vol. 42, No. 4, pp. 1-6, 2012.
- [14] J. A. Sethian, "Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics", *Computer Vision and Material Science*, Cambridge University, UK, 1996.
- [15] T. R. Raviv, N. Kiryati, and N. Sochen, "Segmentation by Level Sets and Symmetry", *Proceedings of the IEEE Conf. On Computer Vision and Pattern Recognition*, pp. 1012-1022, June, 2006.
- [16] J Suckling et al., "The Mammographic Image Analysis Society Digital Mammogram Database", *Exerpta Medica, International Congress Series 1069*, pp.375-378, 1994