Semi Automatic Segmentation of Breast Cancer for Mammograms Based on Watershed Segmentation

Shavi Gupta1, Mohd. Sadiq2, Mona Gupta3 and Naseem Rao4

1Department of Electronics and Communication Engineering, Apeejay College of Engineering, Sohna Affiliated to Maharshi Dayanand University, Rohtak, Haryana, India
2Section of Computer Engineering, University Polytechnic, Faculty of Engineering and Technology, Jamia Millia Islamia (A Central University), New Delhi-110025, India
3Software Engineer, Steria India Ltd, Noida, UP, India
4Department of Electronics and Communication Engineering, Priyadarshini College of Computer Science, Greater Noida, U.P. India

ABSTRACT
Mammograms are the soft X-rays meant for the detection of any lesions or cysts in breasts. Nowadays it is playing a major role in the detection of breast cancer at an early stage. In this paper we have developed a computer aided diagnosis tool using MATLAB for breast cancer in Mammograms. This segmentation technique is based on the extraction of catchments basins through a topographic representation of the mammography image. Applying the watershed algorithm on a rough image leads to over-segmentation problems. To avoid this, we have first carried out a preprocessing step. Then, we have found the distance transform of the preprocessed image. Finally the watershed segmentation is applied to this latter image.

KEYWORDS
Mammogram, Cancer detection, Morphological operation, segmentation, Watershed algorithm, filtering

1. INTRODUCTION
According to the national cancer institute breast cancer is the leading cause of cancer deaths in women today and it is the most common type of cancer in women. Each year about 180,000 women in United States develop breast cancer and about 48,000 lose their lives to the disease. It is also reported that a women’s lifetime risk of developing breast cancer is one in eight. Currently digital Mammography is one of the most promising cancer control strategies in earliest stages. But Mammography has low positive predictive value PPV (Only 35% have malignancies). Several research works have tried to develop computer aided diagnosis tools. They could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis. In order to perform a semi-automated tracking of the breast cancer, it is necessary to detect the presence or absence of lesions from the mammograms. The rest of this paper is organized as follows. Section 2; give a brief description about the watershed morphological segmentation algorithm. In section 3, we have given description about distance transform. The proposed system discussion and their respective results are described in section 4. Conclusion and future scopes are given in section 5.

2. MORPHOLOGICAL WATERSHED ALGORITHM

2.1 INTRODUCTION
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.

2.2 BASIC CONCEPTS
As pointed already, the basic concept of watershed is based on visualizing a gray level image into its topographic representation, which includes three basic notions: minima, catchment’s basins and watershed lines. Fig. 1 illustrates the meanings of these definitions. In the image of Fig. 1. (A), if we imagine the bright areas have "high" altitudes and dark areas have "low" altitudes, then it might look like the topographic surface illustrated by Fig. 1. (b). In this surface, it is natural to consider three types of points:

(1) Points belonging to the different minima ;(2) point at which water would fall with certainty to a single minimum; and
(3) points at which water would be equally likely to fall to more than one minimum. The first type of points forms different minima of the topographic surface. The second type points, which construct a gradient interior region, is called catchment’s basin. The third type of points form crest lines dividing different catchment’s basins, which is termed by...
The objective of watershed segmentation is to find all of the watershed lines (the highest gray level). The most intuitive way to explain watershed segmentation is the Immersion Approach: imagine that a hole is drilled in each minimum of the surface, and we flood water into different catchment’s basins from the holes. If the water of different catchment’s basins is likely to merge due to further immersion, a dam is built to prevent the merging. This flooding process will eventually reach a stage when only the top of dam (the watershed lines) is visible above the water line. An efficient algorithm to implement this approach proposed by Luc Vincent and Pierre Soille [3] involves two steps:

1. The sorting step: In this step, the algorithm first determines the frequency of each gray level of the image, which allows allocating the needed memory for each list of pixels. Then, algorithm scans the image a second time and inserts the pixels in the list corresponding to their gray level value in increasing order.

2. The flooding step: Suppose we begin the flooding from the minimum of gray level \( h \), and \( X_h \) is made of the pixels belonging to the minimum. Now, we consider the gray level \( h+1 \). \( Y_h \) is one component made of the pixels of gray level \( \leq h+1 \).

We can find three possible relations between and \( Y_h \), illustrated by Fig. 2:
- \( Y_h \cap X_h = \emptyset \): in this case \( Y_h \) is obviously a new minimum at gray level \( h+1 \);
- \( Y_h \cap X_h \neq \emptyset \) and is connected: in this case, \( Y_h \) belongs to the catchment’s basin corresponding to the minimum of \( X_h \); \n- \( Y_h \cap X_h \neq \emptyset \) and is not connected: in this case, we can infer that \( Y_h \) must contain different minima at gray level \( h \).

3. DISTANCE TRANSFORMS
Watershed segmentation is applied to gradient image or distance transformation image for binary images, however, there are only two gray levels 0 and 1 standing for black and white. If two black blobs are connected together in a binary image like Fig. 3, only one minimum and catchment’s basin will be formed in the topographic surface. To use watershed to segment the connected blobs, we need to use distance transforms (DTs) to preprocess the image to make it suitable for watershed segmentation.

We define the DT of a binary image as the distance from every pixel of the object component (black pixels) to the nearest white pixel. There are many different ways to define the distance between two pixels \([i_1, j_1]\) and \([i_2, j_2]\) in a digital image. Several commonly used DT functions [4] for image processing are:

- **Euclidean:**
  \[
  d_{\text{Euclidean}}([i_1, j_1],[i_2, j_2]) = \sqrt{(i_1 - i_2)^2 + (j_1 - j_2)^2}
  \]

- **City block:**
  \[
  d_{\text{City block}}([i_1, j_1],[i_2, j_2]) = |i_1 - i_2| + |j_1 - j_2|
  \]

- **Chessboard:**
  \[
  d_{\text{Chessboard}}([i_1, j_1],[i_2, j_2]) = \max(|i_1 - i_2|,|j_1 - j_2|)
  \]

4. PROPOSED SYSTEM DISCUSSION AND RESULTS
The method we have developed consists of different modules. Before carrying out segmentation; the mammogram must undergo a pre-processing step to avoid the over segmentation. The pre-processing consists in removing or attenuating the curvilinear structures (CLS) by applying an adapted filtering mask, which correspond to the blood vessels, the veins, the milk ducts, speculations and fibrous tissue. The second module consists of finding out distance transform and finally the third module is the application of watershed algorithm as mentioned above.

4.1 PREPROCESSING STEPS
We have tested this proposed method on digital mammograms taken from the mini-MIAS database.

**Step 1** of the proposed software is to Read the Images by Mat lab software and convert them to matrix of values each value represent two-dimensional light intensity function \( f(x, y) \), where \( x \) and \( y \) denote spatial coordinates and the value of \( f \) at any point \((x, y)\). The image contains many objects of different sizes that are touching each other.

**Fig 4:** The original image

**Step 2** aims to create the structuring element, because we want to apply dilation and erosion operations and it is an essential part of doing this is to create the structuring element that is used to probe the input image. The structure element is a matrix consisting of only 0’s and 1’s that can have any arbitrary shape and size, we used Mat lab strel function to create a diamond shape-structuring element to do this as shown below

**Fig 5:** Diamond structure element

**Step 3** of the proposed software is to Enhance the Image Contrast and to do this we used imtophat and important functions in Mat lab image processing toolbox that returned the top-hat and bottom-hat transformations, respectively, of the original image the result of this step is shown in Fig. 6 and 7 and the equation used to do this is shown in (2) and (3).

\[
I_{\text{top}} = \text{imtophat}(a\text{fin}, \text{se}) \quad (2)
\]

\[
I_{\text{bot}} = \text{imbothat}(a\text{fiui}, \text{se}) \quad (3)
\]

After that we convert objects of Interest to do that we do that by watershed transform that detects intensity valleys in the image and also to do that we enhanced the image By highlighting the intensity valleys using the imcomplement function in the Mat lab image processing toolbox as shown in Fig. 9.

**Step 4** exaggerate the gaps between objects, the top-hat image contains the peaks of objects that fit the structuring element. To do this we used the important function, which shows the gaps between the objects. To maximize the contrast between the objects and the gaps that separate them from each other in our proposed software we added the top-hat image to the original image and then subtract the bottom-hat image from the result (Fig. 8). Our software uses the toolbox image arithmetic functions, imadd and imsubtract, to perform these operations as shown in (3).

\[
\text{lenhance} = \text{imsubtract}(\text{imadd}(I_{\text{top}}, \text{afm}), I_{\text{bot}}) \quad (3)
\]

**Fig 6:** Bottom-hat image

**Fig 8:** Original + top-hat image-bottom-hat image
Fig 9: Enhanced image
After that the proposed software that detects all the intensity valleys deeper than a particular threshold with the imextendedmin functions in the Mat lab image processing toolbox. The output of this function is a binary image. The location rather than the size of the regions in the imextendedmin image is important.
Then the imimposemin function modifies the image to contain only those valleys found by the imextendedmin function. The imimposemin function also changes a valley’s pixel values to zero, all regions containing an imposed minima are detected by the watershed transform as shown in Fig. 10(a) and 10(b).

Fig 10(a): Extended minima image
Fig 10(b): Limposed minima image

Step 7 is applying watershed segmentation, to accomplish watershed segmentation of the imposed minima image as shown in (5)

War = watershed (limpose)  

Fig 11: Watershed Segmented Image
The watershed function in (5) returns a label matrix containing nonnegative numbers that correspond to watershed regions. Pixels that do not fall into any watershed region are given a pixel value of 0; a good way to visualize a label matrix is to convert it to a color image, using the label2rgb function. In the color version of the image, each labeled region displays in a different color and the pixels that separate the regions display white as shown in Fig. 11.

5. CONCLUSION AND FUTURE WORK
Watershed segmentation is an effective method for gray level image segmentation. To apply watershed segmentation to binary images we need to preprocess the binary images with distance transform to convert it to gray level images, which are Suitable for watershed segmentation. This paper has highlighted a semi-Automated method for detection of cancer tumor/lesions in breasts at an early stage by digital mammogram image analysis.
As compared to manual segmentation process (operator time approx. 1 hr.) the semi-automated method here presented has quite a reduced operator time (approx. 15 to 20 min.). So using this semi-automated method a large number of images can be process in a shorter time as compared to manual segmentation.
The presented method also has the reproducibility and accuracy almost comparable to that of manual segmentation.
In the future work we are planning to extract certain feature (Mean, variance, uniformity, shape, texture) of the segmented image. Based on these features these tumors can be classified as benign or malignant

REFERENCES
Semi Automatic Segmentation of Breast Cancer for Mammograms Based on Watershed Segmentation

International Workshop on Digital Mammography


[7]. Mohammad Sameti, Rabab E. Ward, Branko Palcic, Jacqueline Morgan-Parkes “Texture Feature Extraction for Tumor Detection in Mammographic Images” 2007


[10]. Qing Chen 1, Xiaoli Yang 2, Emil M. Petriu1 “Watershed Segmentation for Binary Images with Different Distance Transforms” 2006


[16]. Xiaoheng Yang , Nakajima Lab, Titech “Image segmentation” 2007