Detection of Breast Tumor Candidates Using Marker-controlled Watershed Segmentation and Morphological Analysis

Samual H. Lewis and Aijuan Dong
Department of Computer Science, Hood College, Frederick, MD, 21701, USA

Abstract

Computer Aided Diagnosis (CAD) was approved to automate breast cancer detection with mammograms in 1998. But due to the great variability in tumor sizes and shapes, and underlying breast tissue structures, pattern recognition algorithms have a difficult time adapting to different situations. In this paper, a marker-controlled watershed segmentation algorithm was developed to locate breast mass tumor candidates. The approach first selected foreground and background markers, and then applied watershed segmentation algorithm to isolate a tumor region from its surrounding tissue. Since watershed segmentation is based on pixel density variation that is present in all mass tumors, the proposed approach was fairly successful in locating tumors under all conditions. Experiment results with Mammographic Image Analysis Society (MIAS) data set showed the overall detection rate for mass tumors is 90%.

1. Introduction

Breast cancer continues to be a major public health problem around the world [1]. The best way to improve the prognosis for breast cancer is early detection when treatment is more effective, and a cure is more likely.

A variety of abnormalities may appear in mammograms, including masses, microcalcifications, and architectural distortions. In this study, we were interested in mass detection. Masses are groups of cells that are clustered together. They are often denser than the surrounding tissue. Based on shape, masses can be classified as circumscribed, spiculated, or ill-defined. Researchers have found out that the main obstacle of mass detection is the great variability of mass appearance [2].

In mass tumor detection, image segmentation is often used to locate suspicious regions. Various segmentation techniques have been proposed for mass detection in the literature[3][4], including region-based techniques, thresholding-based techniques, edge-based techniques, fuzzy techniques, bilateral image subtraction and multi-scale techniques. Since the study reported here was based on watershed segmentation, the review that follows were focused on this aspect.

Watershed segmentation has been used in various image processing and computer vision tasks. Sheshadri and Kandaswamy [5] 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. Dubey, Hanmandlu and Gupta [6] compared two different semi-automated segmentation methods, i.e., modified gradient magnitude region growing technique (MGMRGT) and watershed method. The study found that the MGMRGT segmentation is better than watershed approach. Huang and Chen [7] integrated neural network (NN) classification and watershed segmentation to extract precise contours of breast tumors from ultrasound images. Computer simulation results showed similar contours and regions-of-interest (ROIs) to those obtained by manual contouring. Xu et al. [8] 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.

Mammograms have low image contrast and a lot of noise from natural breast structures such as milk ducts. Direct application of watershed segmentation often leads to over-segmentation. This is due to the noise and other local irregularities that lead to a large number of potential but trivial regional minima. Previous research of applying watershed segmentation on mammograms also focused on locating the precise contour of mass lesions. In our study, we were interested in investigating the overall detection rate of using marker-controlled watershed segmentation on mass tumor candidate detection. This quantitative study is important because it determines the overall sensitivity performance of a tumor detection algorithm.

For the proposed marker-controlled watershed segmentation algorithm, we first used morphological operations, including opening-by-construction and closing-by-construction, to smooth the images, and then identified foreground markers (also called internal

978-1-4673-1830-3/12/$31.00

marker) via regional minimum and background markers (also called external markers) via adaptive thresholding segmentation on mammographic images. Both markers worked together and provided a good control of number of segments generated.

The rest of the paper is organized as follows. Section 2 first gave an overview of the proposed approach, and then detailed on the major steps of the proposed algorithm. Experiment results were presented and discussed in Section 3. Section 4 concluded the paper and discussed future work.

2. Proposed method

2.1. Overview

Watershed segmentation 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 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. The flooding and dam construction process continues until only the dams are visible from above. These dams (connected boundaries) effectively segment the image into regions.

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, as indicated before, direct watershed segmentation on mammograms often generates unreliable results. In our study, we constrained the watershed segmentation with foreground and background markers that were selected via multiple morphological operations.

The proposed marker-controlled watershed segmentation method consists of a few phases as shown in Figure 1. Image processing enhances image quality and removes noise as much as possible. Instead of segmenting the original image, the second step generates a gradient image for segmentation. After that, foreground and background markers are computed. At last, water segmentation, constrained by markers, carries out on the gradient image.

Image preprocessing

A typical mammogram has two kinds of background noises: black background and artifacts such as medical labels (Figure 2(a)). The goal of image preprocessing is to remove black background and unwanted artifacts from mammograms as much as possible. Thus, the first step in image preprocessing is to find breast regions and cropped out the unwanted image portions.

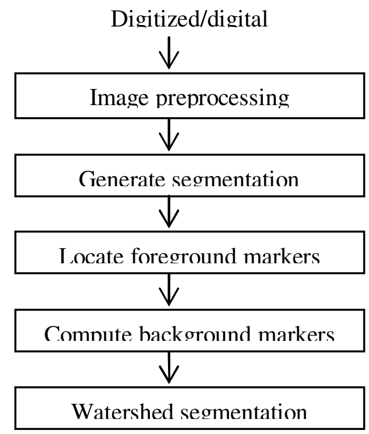
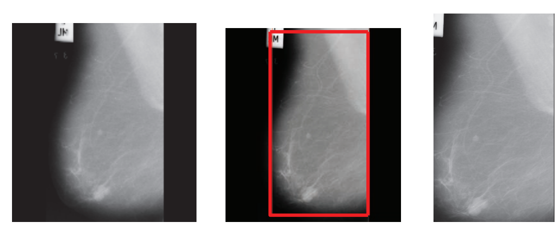


Figure 1. The proposed method

To locate breast region in a mammogram, global thresholding is applied to segment the mammogram. The global threshold of a mammogram is determined using the algorithm proposed in [9]. Since the breast is generally the largest region in a mammogram, the issue of locating breast region is converted to the task of finding the largest segment. The smallest rectangle containing the largest segment, i.e. the breast region, is then used to remove the background and unwanted medical labels (Figure 2 (c)). Since the intensity difference between black background and breast regions and size difference between artifacts and breast regions are large, the global thresholding segmentation can locate breast regions with satisfying accuracy.



(a) The original (b) The segmented (c)The cropped

Figure 2. Image preprocessing

Generate Segmentation Function

One common application of watershed segmentation is to extract regions or blob-like objects of uniform or near-uniform intensity. Since a region with low intensity variation has small gradient, therefore in practice watershed segmentation often applies on gradient images instead of original images. In such cases, high gradient magnitudes are at object or region boundaries, while low gradient magnitudes occur inside regions or objects.

The goal of this step is to generate gradient magnitude images that are to be used for later segmentation step. To this end, two Sobel filters, one is horizontal (Formula (1)) and the other is vertical (formula (2)), are applied on the cleaned image I from preprocessing step, respectively. The process generates two edge images, Gx and Gy. These images are then used to calculate gradient image G.

$$Gx = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * \quad I \tag{1}$$

$$Gy = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * \quad I \tag{2}$$

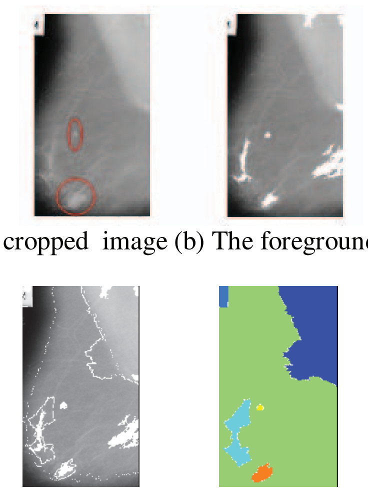
$$G = \sqrt{Gx^2 + Gy^2} \tag{3}$$

Locate foreground markers

One reason that leads to over-segmentation results from a large number of potential but trivial regional minima. The goal of this step is to locate regional minima that are more likely containing mass tumors. A regional minima has to satisfy three conditions 1) surrounded by higher value; 2) region areas are bigger than a predefined threshold value; and 3) points inside a region are connected and are of the same intensity value.

In our approach, since markers are picked from original images and mass tumors have higher intensity values than its surroundings, the foreground markers are located at regional maxima. The locations of these regional maxima are mapped back for segmentation on gradient images on last step.

To pick foreground markers, a variety of morphological functions are applied: opening by reconstruction, closing by reconstruction. Opening by reconstruction is erosion followed by image morphological reconstruction. This operation is used to remove some of the bright pixels from the edges of regions. Depending on the size of the structure element, the step will also effectively remove some regional maxima. Closing-by-reconstruction is a dilation followed by image reconstruction. This operation is applied to shrink background color holes inside regions. This will actually lead to the merging of regions, which also remove trivial regional maxima. Compared with normal opening and closing, opening by reconstruction and closing by reconstruction are less destructive and can maintain an object's shape better. At last, regional maxima are selected using 8-neighborhood. Pixels inside are connected with same intensity value, t, whose external boundary pixels all have value less than t. Continue with the example, Figure 3(b) shows the cropped image with foreground markers overlaid on it. The bright regions are foreground markers. In Figure 3(a), the two red ovals indicate the locations of the tumors.



(a) The cropped image (b) The foreground markers

(c) The background markers (d) The segmentation result

Figure 3. Marker-controlled watershed segmentation

Compute Background markers

While the foreground markers determines allowable regions to start the flooding process, the background markers constrain the flooding process so that it will stop at the edges of the objects we are trying to segment. To this end, an adaptive thresholding segmentation method is applied on the original image. After that, with holes inside regions filled and very small regions removed, the segment boundaries are used as background markers in our approach. Figure 3(c) shows the image with both foreground and background markers overlaid on it.

Watershed Segmentation

In this step, first the locations of foreground makers and background markers are mapped to the gradient magnitude image. And then modify the gradient magnitude image so that its only regional minima occur at those locations where foreground and background markers are. Finally, watershed segmentation is performed on the modified gradient magnitude image. Continue with the example, Figure 3(d) shows the final segmentation result. As can be seen, the two mass tumors in this image are correctly located. There are also a few false positives, which can be removed by investigating region properties, such as area, average intensity, circularity, etc. In this study, the goal is not missing real tumors.

3. EXPERIMENT

The image collection used in this study comes from the Mammographic Image Analysis Society (MIAS)[10]. The corpus has a total of 161 paired images, with each image falling into one of seven categories: calcification, circumscribed masses, spiculated masses, architectural distortion, asymmetry, other ill-defined masses and normal. The first six categories are considered abnormal. The mammograms were originally digitized with 50 micron pixel edge but then reduced to 200 micron pixel edge. The size of each image is 1024x1024 and the bit depth per pixel is 8. Other information, such as breast tissue type, severity of abnormality, and location of abnormalities, is also provided.

In this study, we selected circumscribed, spiculated and miscellaneous masses as types of breast tumors that were to be identified. A total of 48 images were used with 22 being circumscribed (CIRC), 19 being speculated (SPIC), and 7 being miscellaneous (MISC). A couple images have more than one tumor; the total number of tumors in these images was 50. Table 1 showed the results of the study. 87% of the circumscribed masses, 94% of the spiculated masses, and 85% of the miscellaneous masses were correctly identified. The overall detection rate was 90%.

Table 1. Experiment results on mass lesion detection

	CIRC	SPIC	MISC	Total
# of image	22	19	7	48
# of mass lesions	24	19	7	50
# of hit	21	18	6	45
# of miss	3	1	1	5
Detection rate	87%	94%	85%	90%

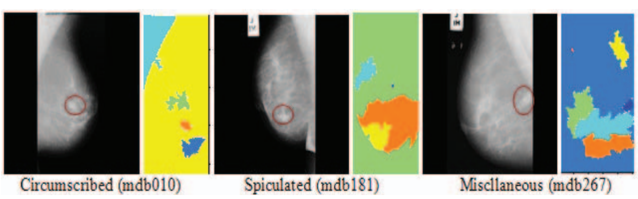


Figure 4. Examples with tumor correctly located

Figure 4 showed three pairs of mammograms with tumor locations correctly identified. For each pair of images, the one on the left was the original mammogram with tumor location marked in red circle; the one on the right showed the watershed segmentation result. Regions of different colors represented different segments. As can be seen, the proposed watershed segmentation is promising in locating tumor candidates. In some cases, the segmented region matched the mass tumor very well, e.g. mdb010. In other cases, the segmented regions were either bigger, e.g. mdb181, or smaller, e.g. mdb267, than actual mass tumor sizes. These were generally because background markers were either not restrictive enough or too restrictive. An improvement could be made in locating background markers more accurately. Another issue shown was that there were multiple false positive regions located in each pair. As we indicated in Section 2, 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.

4. Conclusion and future work

In this paper, a marker-controlled watershed segmentation for breast tumor candidates detection was investigated. Instead of applying watershed segmentation directly on mammograms, we studied a morphological approach to clean up images and then determined foreground and background markers, which addressed the issue of over-segmentation and made the watershed segmentation result more reliable. The experiment with MIAS showed a 90% detection rate for mass tumors.

Future work will be directed toward finding ways to better locate foreground and background markers in order to improve the detection rate. Another focus is to remove false positives from the results. The goal of 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. Finally, we would like to test our algorithm on a larger data set to generate stronger results.

References

[1]. American Cancer Society, "Cancer facts and figures", 2011.

[2]. I. Christoyianni, E. Dermatas, G. Kokkinakis, "Fast detection of masses in computer-aided mammography," IEEE Signal Process. Mag., 17(1), pp. 54-64, 2000.

[3]. 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, 39(4), pp. 646-668, 2006.

[4]. M. P. Sampat M. K. Markey, A. C. Bovik, "Computer-Aided Detection and Diagnosis in Mammography," Handbook of image and video, pp. 1195-1217, 2005.

[5]. H. S Sheshadri A. Kandaswamy, "Detection of Breast Cancer Tumor based on Morphological Watershed Algorithm," ICGST-GVIP Journal, Volume 5, Issue 5, 2005.

[6]. R. Dubey, M. Hanmandlu, S. Gupta S., "Segmentation and Characterization of Masses in the Digital Mammograms," WebmedCentral BREAST 2010, 1(9):WMC00726.

[7]. Y. L. Huang and D. R. Chen, "Watershed segmentation for breast tumor in 2-D sonography," Ultrasound in Medicine & Biology, vol. 30, pp. 625-32, 2004

[8]. S. Xu, H. Liu and E. Song, "Marker-Controlled Watershed for Lesion Segmentation in Mammograms," Journal of Digital Imaging, 24(5), pp. 754-763, 2011.

[9]. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.

[10]. J Suckling et al., "The Mammographic Image Analysis Society Digital Mammogram Database", Exerpta Medica, International Congress Series 1069, pp.375-378, 1994