

AUTOMATIC SEGMENTATION IN BREAST CANCER USING WATERSHED ALGORITHM

Hanan Alshanbari¹, Saad Amain¹, James Shuttelworth¹, Khaled Slman² and Shadi Muslam²

¹Faculty of Engineering and Computing, Coventry University

²King Abdullah Medical City (KAMC), Saudi Arabia

ABSTRACT

Accurate and reproducible delineation of breast lesions can be challenging, as the lesions may have complicated topological structures and heterogeneous intensity distributions. Diagnosis using magnetic resonance imaging (MRI) with an appropriate automatic segmentation algorithm can be a better imaging technique for the early detection of malignant breast tumours. The main objective of this system is to develop a method for automatic segmentation and the early detection of breast cancer based on the application of the watershed transform to MRI images. The algorithm was separated into three major sections: pre-processing, watershed and post-processing. After computing different segments, the final image was cleared of all noise and superimposed on the original MRI image to generate the final modified image. The algorithm successfully resulted in the automatic segmentation of the MRI images, and this can be a beneficial tool for the early detection of breast cancer. This study showed that the automatic results correctly agree with manual detection.

KEYWORDS

Image Processing, Automatic Segmentation, Watershed Segmentation, Breast Cancer & MRI

1. INTRODUCTION

Breast cancer is one of the most common cancers in the female population in the world. About 25% of all cancers diagnosed in women are breast cancers and about 20% of all lethal cancers are breast cancers. It is the most common cause of mortality in women. The appearance of breast cancer is subtle and unstable in the early stages; therefore, radiologists and physicians can take much time and still miss the abnormality easily if they only diagnose by experience [1].

Early detection of a tumour at an early stage is the key to improved chances of successful treatment. It also improves patient survival rates significantly. Today, in most hospitals, radiologists perform the diagnosis of a breast tumour manually on mammographic or other radiological images, which is a time consuming and error prone process due to the small sizes and various shapes of lesions, and these manual images are affected by the presence of low contrast and unclear boundaries between surrounding normal tissues. It has been observed that even experienced radiologists miss 10–30% of breast cancers during routine screening in manual diagnosis [2].

Today, a promising technology that can enhance breast cancer detection rates is the use of magnetic resonance imaging (MRI) as a complement or even alternative to mammography for the diagnosis of breast cancer. New techniques while using MRI have led to the better characterisation of lesions and can help diagnose malignancy at an earlier stage. An accurate and standard technique for breast tumour segmentation will play a pivotal role in detecting and quantifying breast cancers [3, 4].

The segmentation of images plays a key role in practical applications, such as in medical science. Medical images are important to object recognition of the human organs. 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 processing. The purpose of segmentation is to partition images that have different characteristic tissues into semantically interpretable regions, such that the characteristics of each region and extract interest objects [6].

Various image segmentation algorithms have been proposed to achieve efficient and accurate results. Among these algorithms, watershed segmentation is a particularly useful method. Watersheds are a classic technique in the field of topography and have long been considered a useful tool in image segmentation. Watershed segmentation is based on morphological concepts and it was originally proposed by Digabel and Lantuejoul. However, many modifications and improvements have been carried out in the past on the original algorithm. The idea of watershed segmentation is primarily based on the concept of the topographic representation of image intensity. Watershed segmentation also incorporates other principal image segmentation methods, including discontinuity detection, thresholding and region processing. Ultimately, watershed segmentation displays more effectiveness and stableness than other segmentation algorithms. Automatic processing using this algorithm is a promising approach. The objective of watershed transformation is to find the watershed lines on the topographic surface [6].

It has been found that the interpretations of MRI images of breast cancer by radiologists many times provide high rates of false positive diagnoses. Therefore, efforts have been made in the past, as well, by many researchers to develop effective diagnostic methods based on the use of image processing algorithms. The main advantage of MRI is that it offers a clearer image of the breast on which potential tumours are more easily detected, especially when contrast agents are used. The purpose of the automatic segmentation process is to detect the relevant points of the MRI. The watershed transform is a powerful tool for image segmentation [7]. This study aimed to develop an automatic watershed segmentation algorithm for MRI images, which can help lead to a better appearance and detection of early breast cancer lesions.

2. METHODOLOGY

Images of 100 patients were retrieved from the medical imaging department via the picture archives of King Abdullah Medical City (KAMC). From these, 20 images were tested and others will be used in future tests. The same machine with the same sequence parameters examined all patients. The images were obtained under well-controlled conditions by an experienced technician to ensure geometrically aligned orientations. The diagnoses of the patients were kept confidential.

3. PROCESSING APPROACH

The algorithm was separated into three major sections: pre-processing, watershed and post-processing. During pre-processing, the MRI was subject to median filtering for meaningful segmentation. This was followed by image rotation for the orientation and identification of the pectoral muscles. In the watershed section, a Sobel filter was applied to approximate a gradient map of the image. Post-processing consisted of masking the original image and recombining the image. Figure 1 shows the steps while using the watershed algorithm.

3.1. Pre-processing

The only pre-processing done on the image is the application of a median filter to the image. This filter makes the intensity of each pixel the average intensity of it and its direct neighbours. First,

watershed segmentation divides the image into meaningful regions, which describe structures in the MRI, such as large continuous bright and dark patches of similar intensity. Smoothing the image in pre-processing reduces the number of meaningless regions.

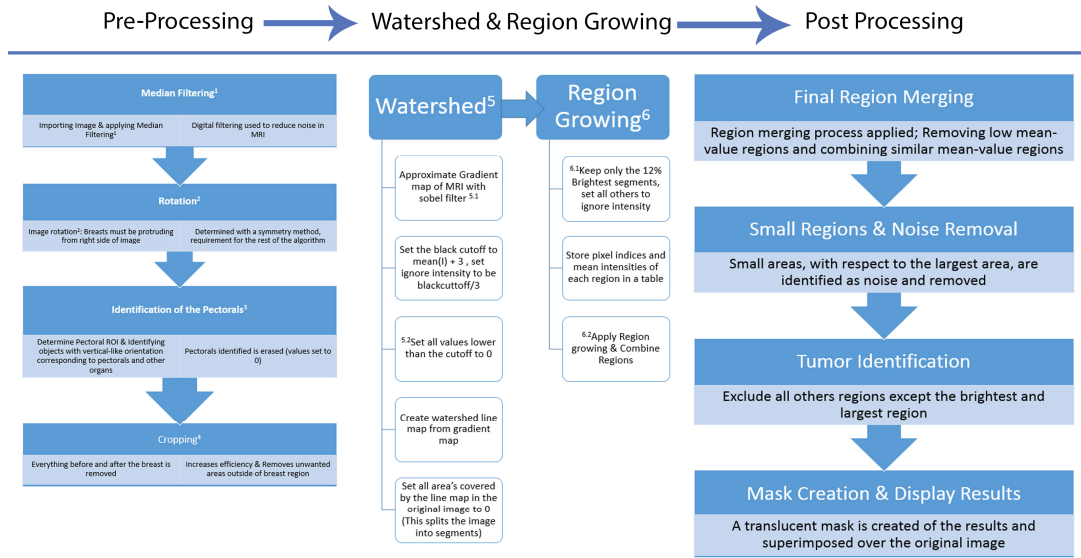


Figure 1. Steps in using the watershed algorithm

3.1.1. Rotation and Alignment

The algorithm defines the orientation of the image by first computing a summation of all intensity values in a given row. It locates the rows where the summations are above the threshold and it proceeds to rotate the image if the row is located below the top 20% of the image. Afterwards, a histogram equalisation is applied, which increases contrast to define better the areas within the breast that may be of interest.

3.1.2. Identification of the Pectorals

The identification of the pectoral muscles is done by first determining the region of interest in which the muscles will be located, such as finding the torso/breast border and then ignoring everything to the left of the border. This is effective, as pectoral muscles in almost all circumstances lie behind the breasts. The remaining portion is then converted to a black and white image with a modified histogram-based threshold. The resulting black and white image is scanned for an area of the minimum size and vertical orientation of the pectoral muscle.

3.2. Watershed Algorithm

The first step is to apply the Sobel filter to the MRI[8]. This is used to approximate a gradient map of the image. In this gradient image, sharp changes in contrast, such as the edge between a dark and light region, appear as high intensity points, while regions with monotone intensity appear as dark patches. Instead of applying watershed segmentation to the original MRI, watershed segmentation is applied to this gradient map. The watershed algorithm can handle monotone plateaus poorly and it over-segments them. In the gradient, however, these areas are monotone, so they appear as a single dark patch surrounded by a high intensity edge at the boundaries, which is segmented better. A list of the regions created by segmentation is stored and

through a few different means, and the list is reduced to just the regions that could contain tumours.

The algorithm starts by marking any pixel below a certain threshold in the original image as a possible background pixel. The pixels inside the largest connected region, where all pixels are below that value, are labelled background pixels. The background regions are removed from a prepared list of breast regions.

Next, the skin segments were removed, and they need to be handled specially because of their tendency to appear extremely brightly on MRIs. A distance map of the breast is generated, which defines the distance between each pixel in the breast and the nearest background pixel to remove the noise. Skin segments are removed from the list of breast regions.

The breast MRI is now composed of only inner breast regions. Most of these are dark regions, and cancer appears brightly on MRIs, so these need not be kept on the list of breast regions to scan. The algorithm removes the darkest 90% of regions from the list of breast regions. The remaining regions in the list of breast regions are the regions that could have a tumour in them. Having reduced the list to just those segments, the algorithm moves on to the combining step.

3.2.1. Combining Regions

The resulting segments are first sorted by average intensity before being placed through a process that combines segments by highest to lowest intensity. First, the index of the region is converted into a position for the seed point. Using this seed, region growing is applied on the segment and segments that are contained within the grown regions are added to the seed region, creating a new region. If both the grown region and the new region are contained within each other, the regions are combined. This process is applied multiple times for ideal region growing to define region growing more formally, such as:

$$(a) \bigcup_{i=1}^n R_i = R.$$

Segmentation must be complete; that is, every pixel must be in a region.

$$(b) R_i \text{ is a connected region, } i = 1, 2, \dots, n$$

Points in a region must be connected in some predefined sense.

$$(c) R_i \cap R_j = \emptyset \text{ for all } i = 1, 2, \dots, n.$$

Regions must be disjoint.

$$(d) P(R_i) = TRUE \text{ for } i = 1, 2, \dots, n.$$

Properties must be satisfied by the pixels in a segmented region. In this case, $P(R_i)$ is true if all pixels in R_i have the same intensity.

$$(e) P(R_i \cup R_j) = FALSE \text{ for any adjacent region } R_i \text{ and } R_j.$$

Region R_i and R_j are different in the sense of predicate P .

$$R_i P(R_i)$$

is a logical predicate defined over the points in set and \emptyset is the null set.

3.3. Post-Processing

After computing different segments, very small areas (0.1% of the image size) are identified as noise and removed. A translucent mask is created from the resulting segments and superposed over the MRI. Similarly, a different translucent colour mask is used to highlight the previously identified pectoral muscle in a separate image. At this point, the image is recombined with the regions previously cropped from the pre-processing stage and is rotated back to its original orientation. The resulting image is then displayed, including both the tumour regions of interest ROI and the identified pectoral muscle.

4. RESULTS AND DISCUSSION

The automatic algorithm successfully segmented, and one example of each step is shown in Figure 2. The proposed method of segmentation has been applied to a series of 20 patients. In this paper, a method for the automated segmentation of cancerous areas in an MRI image has been developed. The suspicious areas in the image were highlighted so they could be analysed further to determine whether the masses detected are cancerous.

As medical images are complex, pre-processing aids are required when using the Sobel operation. Then, the watershed transform is applied. The areas in the image that are highlighted could be analysed to determine whether they are cancerous or non-cancerous. The proposed algorithm has been tested on standard digital images.

Various studies have tested algorithms, including watershed algorithms, for enhancing the sensitivity of mammography or other imaging techniques. However, few incorporated automated methods.

This project shows the watershed transform using foreground markers is able to segment real images[7]. The study showed a similar benefit, as experienced by the above-mentioned clinical study references using various algorithms for processing images. The algorithm successfully resulted in the automatic segmentation of the MRI images. The automatic watershed segmentation method helped identify breast cancer lesions in all 20 MRI images processed. The diagnosis based on this method helped to accurately delineate and detect early breast cancer. The results were cross verified and approved by an independent expert.

In future research, it would be beneficial to verify these results against hospital records. This is to identify 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 as many false positives as possible.

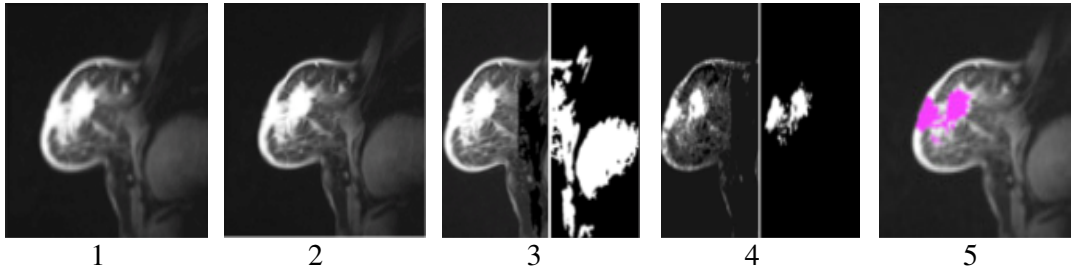


Figure 2. Stepwise image processing using watershed algorithm

1 The original MR Image, 2 after applying median filtering, 3 Pectoral Muscle Identification

4 Watershed and Region growing applied, 4 Final result after masking to the original Image

5. CONCLUSION

The results of MRI segmentation using the watershed algorithm are satisfactory and this can be a beneficial tool for the early detection of breast cancer. The study showed that the automatic results correctly agree with the manual detection. The results were cross verified and approved by an independent expert.

6. REFERENCES

- [1] American Cancer Society, "Breast Cancer Facts and Figures", 2007-08.
- [2] H. D. Cheng et al., "Computer-aided detection and classification of micro calcification in mammograms: a survey," *Pattern Recogn.*, vol. 36, pp. 2967-2991, 2003.
- [3] K. Kinkel and N. M.Hylton, "Challenges to interpretation of breast MRI," *J.Magn.Reson. Imaging*, vol.13, no. 6, pp. 821-829, 2001.
- [4] A. Q. Al-Faris et al., "Computer-aided segmentation system for breast MRI tumour using modified automatic seeded region growing (BMRI- MASRG)," *J. Digit. Imaging*, vol. 27, no. 1, pp. 133-144, 2014.
- [5] S. Gupta et al., "Semiautomatic segmentation of breast cancer for mammograms based on watershed segmentation," in *Proceedings of the 5th National Conference, Computing For Nation Development*, 2011.
- [6] S. R. Shareef, "Breast cancer detection based on watershed transformation," *IJCSI*, vol. 11, no. 1, pp. 237-45, 2014.
- [7] N. R. Raajan, R. Vijayalakshmi, and S.Sangeetha, " Analysis of malignant neoplastic using image processing techniques," *Int. J. Engineer. Technol*, vol. 5, no. 2, pp. 975-4024, 2013.
- [8] A. Hast, "Simple filter design for first and second order derivatives by a double filtering approach," *Pattern Recogn Lett.* vol. 42, pp. 65-71, 2014.
- [9] Z. Qian, "Medical images edge detection based on mathematical morphology," in *Engineering In Medicine And Biology27th Annual Conference*, 1-4 Proc. of The IEEE, 2005.
- [10] J. Sharma and S. Sharma, "Mammogram image segmentation using watershed", *Int. J. Information Technology Know. Manage.*, vol. 4, no. 2, p. 423, 2011.