Automated Region of Interest Detection of Spiculated Masses on Digital Mammograms

Rana Jahanbin[1], Mehul P. Sampat[2], Gautam S. Muralidhar[1],
Gary J. Whitman[3], Alan C. Bovik[4], Mia K. Markey[1]
[1]The University of Texas Department of Biomedical Engineering, Austin, TX, USA
[2]Center for Neurological Imaging, Brigham and Women's Hospital, Boston, MA, USA
[3]The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA
[4]Department of Electrical and Computer Engineering, The University of Texas at Austin, TX, USA
{rana_jahanbin, gautam, mia.markey}@mail.utexas.edu, mehul.sampat@ieee.org

Abstract

We have developed a novel technique to automatically identify a region of interest (ROI) surrounding a spiculated lesion on a mammogram. Our proposed approach for determining the size of the ROI depends on the response of a set of unique Spiculation Filters (SF). The design of these filters is based on manually annotated physical characteristics of spicules. The accuracy of our algorithm is measured in terms of the percentage of spicule pixels located inside the identified ROI. Spicules on each image were identified by an experienced radiologist to serve as a reference to determine the percentage of spicules located in the ROI. On average, 94 *percent of spicule pixels were located inside the ROI identified by our algorithm.*

1. Introduction

The American Cancer Society has estimated that $218,940$ women will be diagnosed with breast cancer in the U.S. in 2007 [1], out of which $40,460$ women are expected to die. Screening mammography, x-ray imaging of the breast, is currently the most effective tool for early detection of breast cancer. However, mammography is not perfect. Detection of suspicious abnormalities is a repetitive and fatiguing task. For every thousand cases analyzed by a radiologist, only 3 to 4 are cancerous and thus an abnormality may be overlooked. Computer-Aided Detection (CADe) systems have been developed to aid radiologists in detecting mammographic lesions that may indicate the presence of breast cancer. These systems act only as a second reader and the final decision is made by the radiologist. Some studies have shown that CADe systems, when used as an aid, have improved radiologists' accuracy for detecting breast cancer.

Radiologists visually search mammograms for specific abnormalities. Some of the important signs of breast cancer that radiologists look for are clusters of microcalcifications, masses, and architectural distortions. Masses with spiculated margins carry a much higher risk of malignancy than calcifications or other types of masses. However, current CADe systems are dramatically better at detecting microcalcifications than masses. The most widely used commercial CADe system is reported to have 98.5% sensitivity at 0.185 false positives per image (FPI) for microcalcification clusters but only 86% sensitivity at 0.24 FPI for spiculated masses [9]. However, the results vary considerably on different datasets. For example, clinical studies to evaluate the performance of commercial CADe systems for mass detection, have reported sensitivities ranging from 67% to 89% with the FPI ranging from 0.40 to 0.74 FPI [10, 2]. For normal images FP rates of 1.3 to 1.8 FPI have been reported [11].

In our previous work, we developed a successful algorithm for the CADe of spiculated lesions [5, 6]. This algorithm entails computing the Radon transform of an image followed by filtering in the Radon domain to enhance the spicules. Once the enhanced image is obtained, the locations where spicules converge are found by using a set of spiculation filters. The parameters of these filters are based on the physical characteristics of the spicules obtained from the measurement study [7].

In this paper, we present a new algorithm for automated detection of region of interest (ROI) in mammogram images that completely contain candidate spiculated lesions. The detection of ROI is based on the responses of spiculation filters applied on mammogram images. The candidate spiculated lesions that are detected can then be subjected to a more intensive, directed and finer scale processing in

978-1-4244-2297-5/08/$25.00 ©2008 IEEE

order to extract features to be used in the classification of normal vs. lesion and, subsequently benign vs. malignant.

2. Background

In our previous work, we developed a detection algorithm comprised of three components. The first component is used to find the central mass region of spiculated lesions by using Gaussian filters. A set of 25 Gaussian filters was applied on each image. The maximum output response across all Gaussian Filters computed at each pixel of the image represents the likelihood of the presence of a mass.

The second component is used to find the spatial locations where spicules converge. The spicules were first enhanced by filtering the mammograms in the Radon domain. Then a set of spiculation filters was applied on the enhanced images. A large overall output of a spiculation filter at a particular spatial location corresponds to a high likelihood of the center of a spiculated lesion. A likelihood map for the locations of spiculated lesions is computed by taking the maximum spiculation filter output across all filters.

The third component is used to reduce the false positives due to normal linear structures in parenchyma. Normal structures such as blood vessels and ducts may be enhanced due to filtering in the Radon domain and these could contribute to large spiculation filter response. Therefore, these linear structures can generate false positive detection. We use the Oriented Difference-of-Gaussian filters to identify and suppress false positives.

3. Methods

3.1. Physical parameter measurements

We asked an experienced radiologist to measure the physical parameters of spiculated lesions for 37 cases. We used the ROI Manager plugin of NIH ImageJ (http://rsb.info.nih.gov/ij/) to enable a user, with minimal training, to place markers at specific locations, and compute the pixel distance between the markers. The radiologist measured the length of the spicules and width of the spicules at their base (where the spicules meet the mass). The radiologist also measured the major axis of the central mass region of each lesion. Finally, the radiologist counted the number of spicules for each mass and this quantity is used to compute angular density. Since the resolution of the images is known, the pixel measurements can be converted into physically meaningful quantities (e.g., mm).

3.2. Region of interest identification

The primary goal of this work is to identify a region in the mammogram image that completely contains the candidate mass and its spicules. It is important that the dimensions of the ROI be not too small, or else some information regarding the candidate spiculated lesion will be lost. Also, it is necessary for the ROI to be not too large since subsequent per pixel directed processing is computationally intensive, and it is crucial not to waste these resources on noncandidate portions of the image. Our approach for detecting the ROI is based on the use of spiculation filters. Spiculation Filters are a new class of complex quadrature filters made up of the cosine, $f_c(r, \theta; r_0, \sigma, \omega)$, and sine, $f_s(r, \theta; r_0, \sigma, \omega)$, components. The spiculation filters have three parameters. 1) radius 2) frequency and 3) σ. The radius of the filter is the size parameter measured in pixels and its value corresponds to the length of the spicules. The frequency corresponds to the number of spicules located per circumference of the central mass region. σ is the standard deviation in pixels. Figure 1 and 2 illustrate examples of sine and cosine spiculation filters.

$$g(r, r_0, \sigma) = \exp(\frac{(r - r_0)^2}{\sigma^2}) \quad (1)$$

$$f_c(r, \theta; r_0, \sigma, \omega) = g(r, r_0, \sigma)\cos(\omega\theta) \quad (2)$$

$$f_s(r, \theta; r_0, \sigma, \omega) = g(r, r_0, \sigma)\sin(\omega\theta) \quad (3)$$

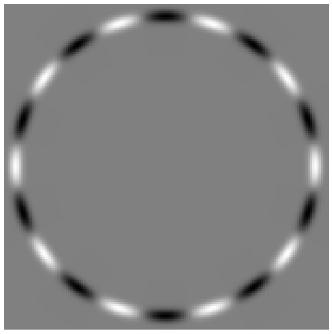

Figure 1. Cosine Spiculation Filter with $\sigma = 2$, $\omega = 10$.

We determine the size of the ROI based on the response of spiculation filters, such as those depicted above. For each candidate spiculated lesion with the true positive location determined by our detection algorithm, we applied a set of spiculation filters with radii increasing from a minimum to a maximum value. The response of these spiculation filters is indicative of the presence of spicules that have a given length. The density of spicules of a given length will determine the response of the spiculation filter as the radii increase. Beyond a certain value of radius, the filter response

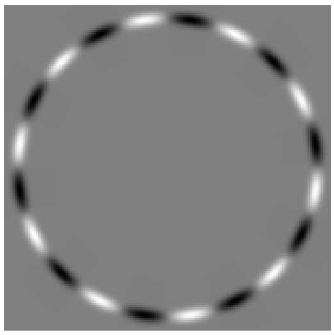

Figure 2. Sine Spiculation Filter with $\sigma = 2$, $\omega = 10$.

becomes negligible owing to the disappearance of spicules. This point will be determined by a simple, empirically calculated threshold value. The outermost above-threshold filter diameter will thus determine the size of the ROI.

It is not possible to know beforehand the number of spicules that are present around the central mass region. Hence we use a set of spiculation filters with each having a different frequency. The frequencies are also chosen based on our ongoing measurement study of the physical characteristics of the spicules [7]. Figure 3 illustrates the results of spiculation filtering of mammogram images. A set of spiculation filters with frequency ranging from 5 to 35 spicules per central mass region was used and radius of each of these filters was increased from 10 to 350 pixels. For each filter with a fixed frequency, we determined the radius corresponding to the first local minimum filter response that was less than a certain threshold. The threshold was determined empirically as half the maximum filter response. The desired radius was then calculated by taking the mean of the radii that correspond to the first minimum value.

4. Results

We tested our ROI detection algorithm on mammograms obtained from the Digital Database for Screening Mammography (http://marathon.csee.usf.edu/ Mammography/ Database.html) [3]. The DDSM is the largest publicly available dataset of digitized mammograms. The mediolateral oblique (MLO) views of 37 cases of spiculated masses were randomly selected from the DDSM. Cases were selected from a single scanner (a pixel resolution of $50\mu m \times 50\mu m$), and we confirmed that a range of density ratings, subtlety ratings, and pathology were represented by the sample.

Figure 4 shows two examples of the detected ROI by our algorithm. The red square represents the border of the ROI detected by this technique. Sides of the square are twice the calculated radius of the spiculated lesion. Blue lines are

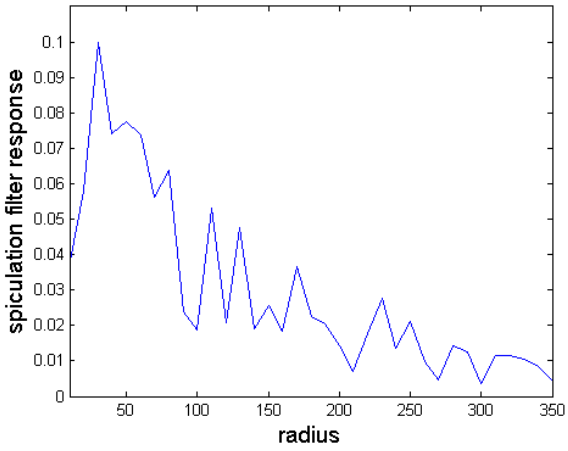

Figure 3. Spiculation filter response at specific pixel (TP location) plotted against increasing radii at a fixed frequency.

spicules marked by an experienced radiologist in the measurement study [7].

In order to evaluate the performance of our algorithm, we calculated the percentage of pixels marked as spicules by the radiologist which are located inside the ROI detected by our algorithm. This value is bound between zero (none of spicules are inside the detected ROI) and one (all of spicules are located inside the detected ROI). The average percentage of pixels inside the ROI is 94%, maximum and minimum values are 100% and 49% respectively. This quantitative metric alone is not sufficient to test the accuracy of the algorithm as it will show a 100% result for the detected ROI when it is much bigger than what is required. Therefore, we used an area overlap criterion as a second metric for evaluating the performance of our algorithm. Overlap criteria have been used by several mass segmentation techniques [4, 8] and is described by equation 4.

$$O = \frac{S \cap T}{S \cup T} \tag{4}$$

S is the area of the box detected by our technique and T is the area of best box, the smallest one that can have all spicules inside. Green squares in figure 4 illustrate the best ROI (T). An overlap percentage of one means an exact match between the two regions. The average overlap percentage of our ROI identification method is 48%.

5. Conclusions

A new automated method for detection of a ROI in a mammogram images has been presented in this paper. The responses of spiculation filters on mammograms are used

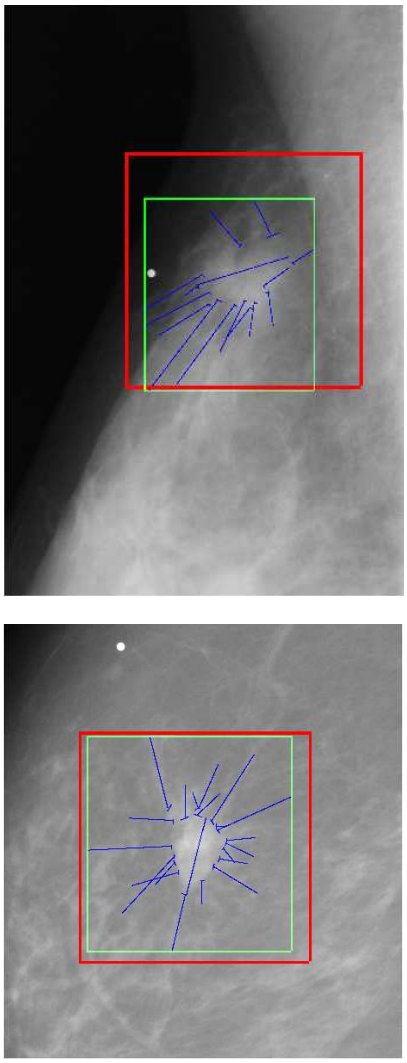

Figure 4. Examples of the detected region of interest (red box) against the best region of interest (green box) are shown on two example mammograms with spiculated lesions.

for ROI identification. Accurate ROI identification is important since it makes it possible to apply a more directed, finer-scale and localized processing in order to identify features that will help in making the classification of lesion vs. normal and benign vs. malignant. Evaluation of the performance of our technique for detecting ROI was done in two different ways. First, we calculated the percentage of pixels marked as spicules by the radiologist which are located inside the ROI detected by our algorithm. On average, 94% of spicule pixels were located inside the ROI. Second, we used an area overlap criteria to determine the overlap in area between the ROI identified by our algorithm and the area that was manually identified as best. On average, overlap percentage of our ROI identification method is 48%. These results show that our approach to automatically detect ROI in mammograms is promising and we plan to use this as a starting point for localized processing.

References

[1] A. American Cancer Society. Cancer Facts and Figures 2007, 2007.

[2] T. W. Freer and M. J. Ulissey. Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center. *Radiology*, 220(3):781–786, 2001.

[3] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer. The digital database for screening mammography. *Int. Work. on Dig. Mammography*, pages 212–218, 2000.

[4] M. A. Kupinski and M. L. Giger. Automated seeded lesion segmentation on digital mammograms. *Medical Imaging, IEEE Transactions on*, 17(4):510–517, Aug 1998.

[5] M. P. Sampat and A. C. Bovik. Detection of spiculated lesions in mammograms. *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, 1:810–813 Vol.1, 17-21 Sept. 2003.

[6] M. P. Sampat, G. J. Whitman, M. K. Markey, and A. C. Bovik. Evidence based detection of spiculated masses and architectural distortions. volume 5747, pages 26–37. SPIE, 2005.

[7] M. P. Sampat, G. J. Whitman, T. W. Stephens, L. D. Broemeling, N. A. Heger, A. C. Bovik, and M. K. Markey. The reliability of measuring physical characteristics of spiculated masses on mammography. *The British Journal Of Radiology*, 79 Spec No 2:PS134 –PS140, 2006.

[8] G. M. te Brake and N. Karssemeijer. Segmentation of suspicious densities in digital mammograms. *Medical Physics*, 28(2):259–266, 2001.

[9] C. J. Vyborny, T. Doi, K. F. O'Shaughnessy, H. M. Romsdahl, A. C. Schneider, and A. A. Stein. Breast Cancer: Importance of Spiculation in Computer-aided Detection. *Radiology*, 215(3):703–707, 2000.

[10] L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino. Potential Contribution of Computer-aided Detection to the Sensitivity of Screening Mammography. *Radiology*, 215(2):554–562, 2000.

[11] S. K. Yang, W. K. Moon, N. Cho, J. S. Park, J. H. Cha, S. M. Kim, S. J. Kim, and J.-G. Im. Screening Mammography-detected Cancers: Sensitivity of a Computer-aided Detection System Applied to Full-Field Digital Mammograms. *Radiology*, 244(1):104–111, 2007.