

AUTOMATED SEGMENTATION AND QUANTITATIVE CHARACTERIZATION OF RADIODENSE TISSUE IN DIGITIZED MAMMOGRAMS

*J.T. Neyhart¹, M. Kirlakovsky¹, L.M. Coleman¹, R. Polikar¹, M.Tseng²
and S.A. Mandayam¹*

*¹Department of Electrical and Computer Engineering
Rowan University, Glassboro, NJ 08028, USA*

²Fox Chase Cancer Center, Philadelphia, PA 19111, USA

Abstract. Mammography has emerged as a reliable non-invasive technique for the early detection of breast cancer – the second leading cause of cancer-related mortality among American women. The radiographic appearance of the female breast consists of radiolucent (dark) regions due to fat and radiodense (light) regions due to connective and epithelial tissue. The amount of radiodense tissue can be used as a marker for predicting breast cancer risk. This paper presents the development of an algorithm for estimating the percentage of radiodense tissue in a digitized mammogram. The technique involves determining a dynamic threshold for segmenting radiodense indications in mammograms. Both the mammographic image and the threshold are modeled as Gaussian random variables. This work is intended to support a concurrent study at the Fox Chase Cancer Center (FCCC) exploring the association between dietary patterns and breast cancer risk. Mammograms have been obtained from an existing cohort of women enrolled in the Family Risk Analysis Program at FCCC.

INTRODUCTION

The usefulness of mammography as a non-invasive, low-cost technique for the early detection of breast cancer in women has been established. Traditionally a trained radiologist has performed the diagnosis of mammography x-rays. The need for reliable techniques for the detection of breast cancer and analysis of risk factors is imperative – breast cancer is the second leading cause of cancer related mortality among American women [1]. The radiographic appearance of the female breast consists of radiolucent (dark) regions due to fat and radiodense (light) regions due to connective and epithelial tissue. The amount of radiodense tissue can be used as a marker for predicting breast cancer risk. Women with radiodense tissue in more than 60-75% of the breast have been shown to be at a four to six times greater risk of developing breast cancer than those with lesser densities [2]. The estimation of radiodense tissue has traditionally been a subjective determination by trained radiologists, with very few published work describing quantitative measures [3, 4]. This paper presents a novel algorithm that employs objective measures for quantifying dense tissue in digitized mammograms. This research project, conducted at Rowan University is intended to support an investigation conducted at Fox Chase Cancer Center (FCCC), examining the correlation between dietary patterns and breast density.

This paper is organized as follows; we first give an introduction to the problem. This is followed by our research goals. The approach outlining the technique we developed is presented. In the results section some of the typical results are presented from exercising our algorithm on the digitized mammogram images. Finally, a summary of the work is provided and future research directions are identified.

RESEARCH OBJECTIVES

The objectives of this research are to develop image processing algorithms that:

- (a) Automatically scan digitized mammogram images to locate the breast tissue region in the x-ray.
- (b) Segment the tissue into radiodense and radiolucent indications.
- (c) Quantify the amount and percentage of radiodense tissue.

In previous work a pair of trained radiologists worked together to classify breast density using a six-category classification system (0%, 0-10%, 10-25%, 25-50%, 50-75%, >75%) [5]. Our intent is to mimic the expert system, the radiologist, and to quantify the percentage of radiodense tissue in the mammogram using objective methods and measures.

APPROACH

Our overall research approach is shown in Figure 1. A flow diagram demonstrating the process of determining the percentage of radiodense tissue is shown in Figure 2. The approach that we have taken addresses two main issues. The first is the actual segmentation of the arbitrarily shaped breast tissue region from within the rectangular shaped x-ray film. This is an edge detection problem that is accomplished using a multiresolution segmentation mask. After the tissue is segmented we then identify and quantify the radiodense tissue indications within the breast region. The challenge here is that the gray-level intensities vary from x-ray to x-ray and locally across the same x-ray. This is a threshold estimation problem. We have developed a dynamic threshold that is modeled as a parametric Gaussian random variable. The mammogram itself is modeled as a Gaussian random image. Details of this research approach are presented in the following sections.

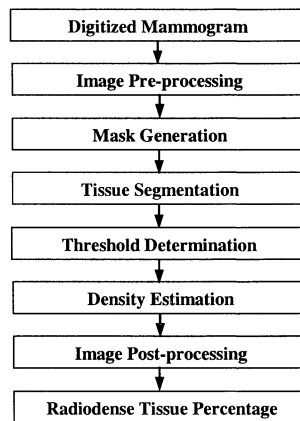


FIGURE 1. Overall approach breast density estimation.

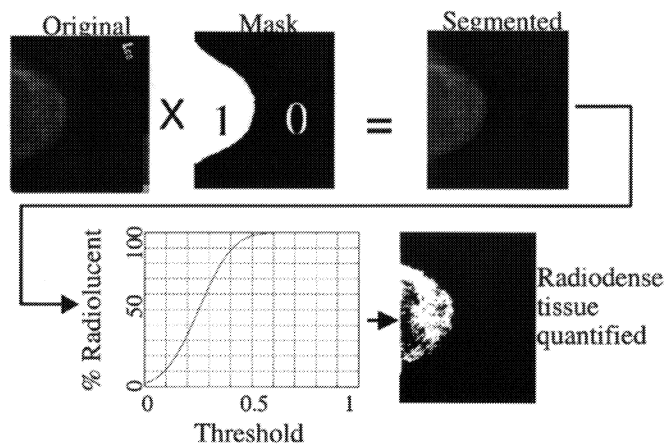


FIGURE 2. Flow diagram of research approach.

Digitization and Pre-processing

The data used in this research comes from the Family Risk Analysis Program at FCCC. This cohort of women includes the daughters, mothers and grandmothers of a population of women that are given regular mammography screenings. Each set of mammograms contains four x-ray films. The x-rays are of the cranio-caudal (CC) and mediolateral oblique (MLO) views of both breasts. The CC view is projected down on the horizontally compressed breast while the MLO view is projected across a breast that is compressed parallel to the patient's pectoral muscle [7]. It has been shown that the separate analysis of the MLO and CC views results in similar findings of breast density [4]. This allows the investigation of only a single view per breast. Because of the inclusion of the pectoral muscle in most MLO views, as shown in Figure 3, analysis is performed only on the CC views of each x-ray set. All x-rays were digitized using a commercial x-ray film scanner. The scanner quantified the intensity of the x-ray at 500 points per linear inch using 8-bits. Each image was then cataloged and placed into a secure database prior to analysis. The first step in image analysis is to map the 8-bit intensity value from the scanner to representative gray scale intensities. Each pixel of data was assigned one of 256 gray-level values ranging from 0 (black) to 1 (white). Each image is oriented to place the breast region on the left side to further aid analysis.

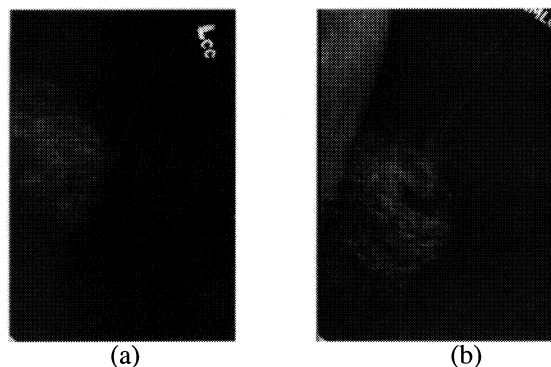


FIGURE 3. (a) CC and (b) MLO views of the same breast, shows the appearance of the pectoral muscle in the MLO view.

Mask Generation

To properly quantify the amount and percentage of radiodense tissue present in the breast there needs to be a technique for determining the tissue region of the x-ray from the film region. To accomplish this a segmentation mask was designed to segment the tissue region from the film region. The mask template is a binary matrix of size equal to that of the original image. The segmentation algorithm described below determines which sections of the image correspond to a tissue region, and assigns the value “1” (white) to the corresponding regions of the matrix. The rest of the matrix, corresponding to the non-tissue region is set to “0” (black). This process allows us to subsequently identify radiodense regions in the image by concentrating on the tissue region only.

The first step in the segmentation algorithm is to scan the x-ray image line-by-line. A scan of a single line in a digitized mammogram can be seen in Figure 4. The goal is to identify the gray-level transition at the boundary of the tissue region. Note that the line-scan exhibits both local and global variations in gray-level. The global variation corresponds to the transition in the x-ray from the tissue to the non-tissue region, which is often obscured by the local variations corresponding to local variations in tissue density. The large variance of the local variations makes it impossible for a preset threshold to identify the tissue region. Furthermore, these local variations within the tissue region make it very difficult to employ standard edge detection algorithms. We therefore use a discrete wavelet transform (DWT) multiresolution decomposition [6] to simultaneously model both these variations in the gray-level for each line-scan in the original x-ray image. A polynomial fit is then used to correlate adjacent line scans to generate the final mask template. The mask template is a binary image consisting of white pixels corresponding to tissue regions and black pixels corresponding to non-tissue regions. The mask template is placed over the original mammogram image. The resulting segmented image contains a gray-level value of 0 (black) on all non-tissue regions. In all regions found to contain tissue the original gray-level value is retained.

The DWT is used for modeling the gray-level values in the line-scan and for identifying the transition from the tissue to non-tissue region. An extensive study of the image heuristics revealed that Daubechies mother wavelet with five vanishing points provided the most optimal model. The fifth level approximation coefficients of the original signal provide space-frequency information that corresponds to the tissue-film boundary of the mammogram image. This model is indicated in Equation (1), $y_0(k)$ is the detail coefficients at a given level and $y_1(k)$ represents the approximation coefficients at that level.

$$y_0(k) = \sum_n x(n)g(-n + 2k) \quad (1)$$

$$y_1(k) = \sum_n x(n)h(-n + 2k)$$

where, $x(n)$ is a single line from the digitized mammogram image. $g(n)$ and $h(n)$, respectively, are the high-pass and low-pass filters. The results $y_0(k)$ and $y_1(k)$ are the detail and approximation coefficients at each level of decomposition. $y_{15}(k)$ is the fifth-level approximation coefficient that is used in determining the tissue-film boundary. Figure 5 shows the associated wavelet decomposition tree for this system.

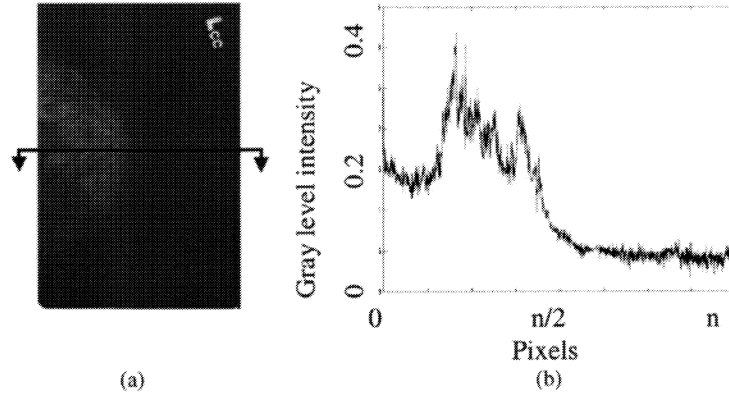


FIGURE 4. (a) Original mammogram image and (b) single line scan.

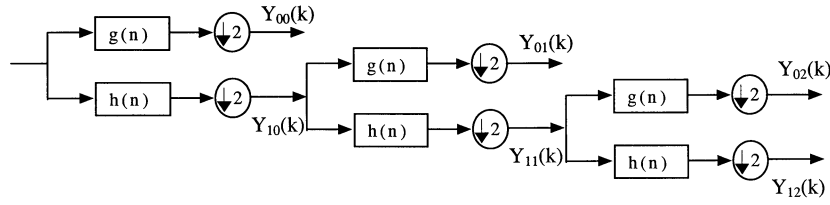


FIGURE 5. Decomposition tree of the discrete wavelet transform.

Tissue Segmentation

The use of a binary mask to perform image segmentation is a common practice and has been implemented here as well. The mask generated using the algorithm described above is a binary matrix equal to the size of the original image. Using the above technique the mask matrix is set to a value of “1” (white) in all tissue regions and “0” (black) in all non-tissue regions. We then multiply the corresponding elements in each array using

$$R_{ij} = O_{ij} * S_{ij}; \quad i=1 \text{ to } m, j = 1 \text{ to } n \quad (2)$$

where O is the original image and S is the segmentation mask, yields the matrix R . The resulting matrix R is of equal size to O and P and contains the original gray-levels of O in all regions designated by S as being of a tissue region and contains a gray-level value of “0” (black) in all regions designated by S as being a non-tissue region.

Threshold Determination

A dynamic threshold for segmentation of the radiodense tissue inside the mammogram film, is described by the equation

$$T_{global} = T_{nominal} + \alpha(T_{local} - T_{nominal}) \quad (3)$$

where T_{local} represents the local gray-level variations, $T_{nominal}$ represents the random image model, α is a parametric weight parameter and T_{global} is the global threshold. If the mammogram was a zero-mean Gaussian random image then

$$T_{global} = T_{local} = T_{nominal} = 50\%. \quad (4)$$

Since this is a real image then we must determine T_{local} . To do this we vary the segmentation threshold across the entire gray scale range. The image is converted to a binary matrix with all pixels above the threshold being set to “1” (white) and all pixels below the threshold being set to “0” (black). The effect of varying segmentation threshold on the percentage of black pixels in a gray-level mammogram image is shown in Figure 6(a). This plot resembles the cumulative distribution function (CDF) of a Gaussian random variable. The probability density function (PDF) can be calculated by differentiating the CDF as shown in Figure 6(b). The mean value of the random variable can be calculated using the PDF. This mean is the local segmentation threshold, T_{local} . The parameter, α , allows the equation to be tuned using a single variable. Validation of this technique will involve the determination of α . The resulting segmentation threshold, T_{global} , is used in the next step of our approach.

Density Estimation and Image Post-processing

Using T_{global} the image is segmented into a binary matrix. All gray-level values that lie above T_{global} are set to “1” (white) and all other values are set to “0” (black). Using this new binary matrix and the segmentation mask matrix we can determine the percentage of radiodense tissue using

$$\% \text{ Radiodense Tissue} = \frac{P_{white}}{P_{total} - P_{film}} \times 100\% \quad (5)$$

where P_{white} is the total number of white pixels in the matrix, P_{total} is the total number of pixels in the matrix and P_{film} is the total number of pixels in the film-only region, as found using the segmentation mask.

RESULTS

Results obtained by exercising the algorithm on typical mammograms that are used in the study shown in Figures 7 and 8. Figure 7 shows the results of the first part of the technique – the dynamic segmentation of the breast region from the film region using a segmentation mask. Figure 8 shows density segmentation from several patients with varying parameter α . Table 1 shows the resulting estimated radiodense tissue percentages for patients from Figure 8.

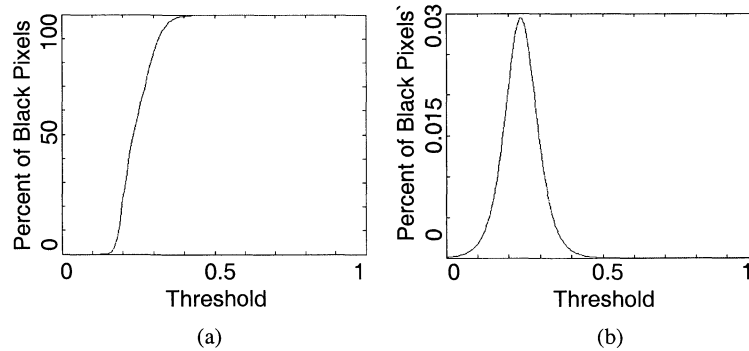


FIGURE 6. (a) Effect of varying segmentation threshold on the percentage of black pixels in a gray-level mammogram image and (b) Probability density function of the threshold random variable.

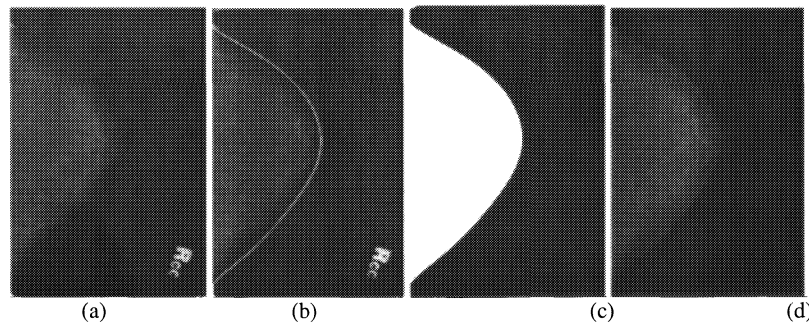


FIGURE 7. Typical results of the mask generation algorithm showing (a) original image, (b) edge line, (c) mask based on edge line and (d) segmented breast region.

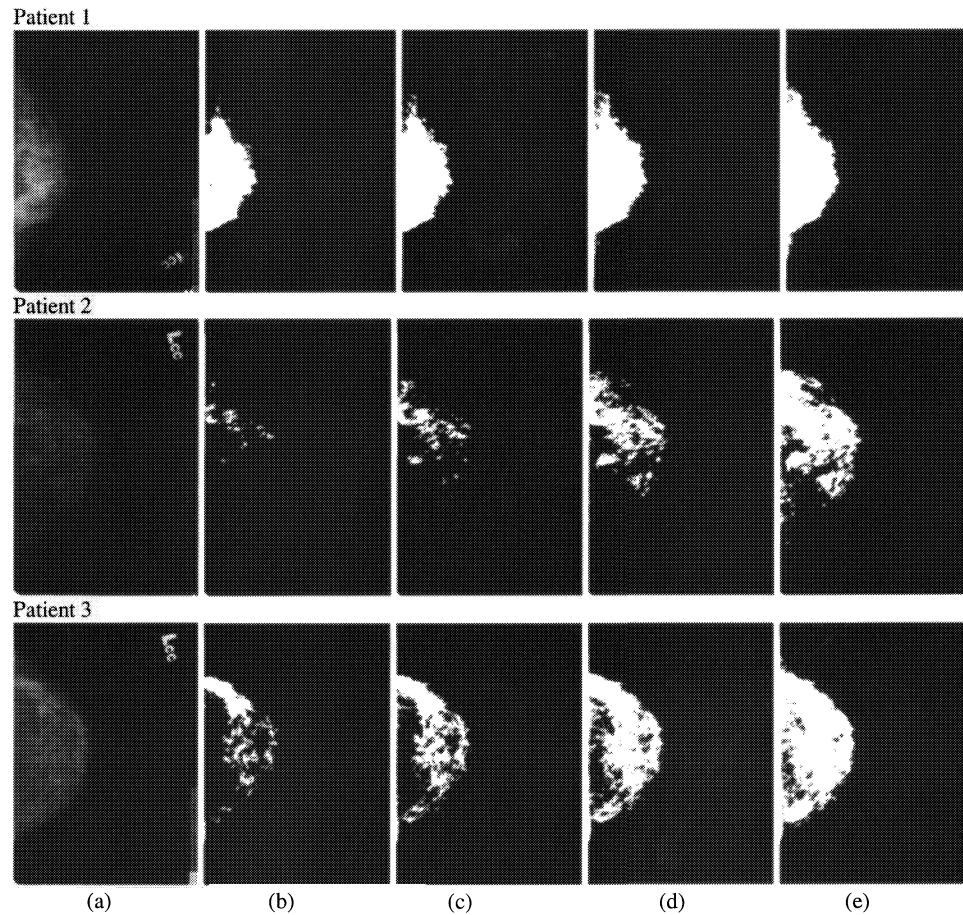


FIGURE 8. Typical results showing (a) the original image and (b-e) the binary segmented tissue region resulting from $\alpha = 0.65, 0.75, 0.85, 0.95$ respectively for three patients.

TABLE 1. Summary of percentages of radiodense tissue for patients from Figure 8.

Patient Number	α , parametric weight			
	0.65	0.75	0.85	0.95
1	33.95 %	37.50 %	41.96 %	45.67%
2	2.69 %	7.76 %	19.11 %	37.31 %
3	13.86 %	23.15 %	40.23 %	55.04 %

SUMMARY

An algorithm for the dynamic segmentation and quantification of the radiodense tissue indications of a digitized mammogram has been presented. We have demonstrated that this technique provides a quantitative estimation of the percentage of radiodense tissue in a digitized mammogram. Future work includes the validation of the results obtained by the segmentation algorithm. We are currently in the process of obtaining validation data.

ACKNOWLEDGEMENTS

This research is supported by grants from the American Institute for Cancer Research and the American Cancer Society. We would also like to thank Dr. Katherine Evers, radiologist, at Fox Chase Cancer and Mr. Dave Chezem of the Rowan University Information Technology Center for their assistance during the course of this project.

REFERENCES

1. S. H. Landis, T. Murray, S. Bolden and P. A. Wingo, "Cancer statistics," *CA Cancer Journal*, Vol 48, pp. 6-29, 1998.
2. N. F. Boyd, G. A. Lockwood, J. W. Byng, D. L. Tritchler and M. J. Yaffe, "Mammographic densities and breast cancer risk," *Cancer Epidemiology, Biomarkers and Prevention*, Vol. 7, pp. 1133-1144, December 1998.
3. N. F. Boyd, J. W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B. Miller, G. A. Lockwood, D. L. Tritchler, M. J. Yaffe, "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study," *Journal of the National Cancer Institute*, Vol. 87, No. 9, pp. 670-675, May 1995.
4. M. J. Yaffe, N. F. Boyd, J. W. Byng, R. A. Jong, E. Fishell, G. A. Lockwood, L. E. Little, D. L. Tritchler, "Breast cancer risk and measured mammographic density," *European Journal of Cancer Prevention*, Vol 7 (suppl 1) pp. S47-S55, 1998.
5. N.F. Boyd, H.M. Jensen, G. Cooke, H. Lee Han, "Relationship between mammographic and histological risk factors for breast cancer," *Journal of the National Cancer Institute*, 84:1170-1179, 1992
6. G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996.
7. G. Wentz, *Mammography for Radiologic Technologists*, McGraw-Hill, New York, NY, 1997.