

Detection of breast cancer by mammogram image segmentation

**Sheshadri HS,
Kandaswamy A**
Department of
Electronics &
Communication
Engineering, PSG
College of Technology,
Coimbatore-641004,
Tamilnadu, India

For correspondence:
Sheshadri HS,
Research scholar,
Department of
Electronics &
Communication
Engineering, PSG
College of Technology,
Coimbatore-641004,
Tamilnadu, India. E-mail:
hssheshadri@hotmail.com

ABSTRACT

An important approach for describing a region is to quantify its structure content. In this paper the use of functions for computing texture based on statistical measures is prescribed. MPM (Maximizer of the posterior margins) algorithm is employed. The segmentation based on texture feature would classify the breast tissue under various categories. The algorithm evaluates the region properties of the mammogram image and thereby would classify the image into important segments. Images from mini-MIAS data base (Mammogram Image Analysis Society database(UK)) have been considered to conduct our experiments. The segmentation thus obtained is comparatively better than the other normal methods. The validation of the work has been done by visual inspection of the segmented image by an expert radiologist. This is our basic step for developing a computer aided detection (CAD) system for early detection of breast cancer.

Key words: Mammograms, CAD, Textural analysis, Segmentation

INTRODUCTION

In the past several years there has been tremendous interest in image processing and analysis techniques in mammography. One common approach for detecting abnormalities in mammograms^[1,2] is to use a series of heuristics, e.g. filtering and thresholding, which may include texture analysis to automatically detect abnormalities.^[3] These heuristic methods suffer from a lack of robustness when the number of images to be classified is large.^[4] Statistical methods have also been developed to address this problem. Brzakovic^[5] used fuzzy pyramid linking to identify homogeneous regions in mammograms, and then used a statistical model to classify regions as non-tumor, benign tumor, or malignant tumor. Kegelmeyer^[4,6] extracted a five dimensional feature vector for each pixel which included edge orientation and the output of four spatial filters. Each feature vector was then classified using a binary decision tree.

In this paper we present a new statistical algorithm^[7,10] which partitions a mammogram into homogeneous texture regions. Our algorithm assigns each pixel in the mammogram membership to one of a eight number of classes depending upon the statistical properties of the pixel and its neighbors. The individual pixel classifications form a two-dimensional labeled which must be

estimated from the observed image. Both the mammogram and its labeled fields are modeled as discrete-parameter random fields. We estimate the pixel classes by minimizing the expected value of the number of misclassified pixels, this is known as the maximizer of the posterior marginals (MPM) estimate. The expectation-maximization (EM) algorithm is employed to estimate from the observed mammogram the unknown parameters needed for the MPM estimate.

MATHEMATICAL MODEL

We model the texture class labels X as a Markov Random Field (MRF) with a four-point nearest-neighbor neighborhood system and probability mass function is defined as

$$p_X(x) = \frac{1}{Z} \exp \left(- \sum_{\{r,s \in C\}} \beta_t(x_r, x_s) - \sum_{\{r\} \in C} \gamma x_r \right) \quad (1)$$

Where β and γx_r are MRF model parameters, Z is a normalizing constant, C is the collection of cliques for the MRF, and $t(m; n) = 0$, if $m = n$; $t(m; n) = 1$, if m is not equal to n . We will assume that the pixels in the observed image Y can be modeled as conditionally independent Gaussian random variables given the pixel labels X , and that the conditional probability density function of the pixel at location r given X depends only on the value of X at the same pixel location, i.e.

$$fY|X(y|x, \theta) = \prod_{r=1}^N fY_r|X_r(y_r|x_r, \theta) \quad (2)$$

where N is the total number of pixels in the image and $\theta = [\mu_1, \sigma_1^2, \dots, \mu_L, \sigma_L^2]$, with μ_i and σ_i^2 being the unknown mean and variance for texture class i , $i = 1, \dots, L$. L is the number of texture classes.

ALGORITHM EMPLOYED

We must assign each pixel in the mammogram membership in one of the L classes subject to minimizing the expected value of the number of misclassified pixels. The estimate which minimizes this expected value is the MPM estimate of X . The MPM estimate at pixel s is given by

$$X_{s,MPM} = \arg \max_k \sum_{X: X_s = k} pX|Y(x|y, \theta) \quad (3)$$

Where

$$k \in \{1, 2, \dots, L\}.$$

A close form solution of the marginal probability mass functions in Equation (3) is infeasible. Our algorithm uses a Gibbs Sampler^[11] to generate a Markov chain $x^{(i)}$ which converges in distribution to a random field with probability mass function $pX|Y(x|y, \theta)$.^[10]

The conditional probability mass functions are the approximated by

$$pX_s|Y(k|y, \theta) \approx \frac{1}{n} \sum_{i=1}^n [1 - t(x_s^{(i)}, k)] \quad (4)$$

where n is the number of iterations performed for the Gibbs Sampler. In order to implement the Gibbs Sampler, we will use the EM algorithm to estimate the value of θ . The EM algorithm is an iterative procedure for approximating maximum-likelihood estimates. Refer^[7] for more details on EM algorithms.

EXPERIMENTAL RESULTS

Images used in this research were obtained from mini MIAS mammogram database. Abnormal mammograms have an interpretation γ_k that indicates the type(s) of abnormalities. In our experiments, images were assumed to consist of three classes: background, normal tissue, and tumor. For the spatial interaction parameter, we shall assume that its value is $2.4^{[7,11]}$. γ_k can be viewed as a cost parameter for class k . If for a given k , k is larger, then class k is less likely to occur than classes with smaller costs. These cost parameters were determined experimentally using a variety of sample mammography images and are set to 1, 2 and 3 for this study, where class 3 is the tumor class. The effect of this is to increase the cost of belonging to the tumor class. Since a tumor is usually associated with higher grayscale values than the other regions, we sort the grayscale values in the observed image and use the sample mean and variance of the largest grayscale values as the initial parameter estimates for the tumor class. The remaining values are then used to obtain initial estimates of the parameters for the normal tissue and background region. In addition to segmenting abnormalities, our algorithm can also indicate the reliability of each classified pixel by using the marginal



Figure 1: a) original mammogram b) segmented image c) microcalcifications given in the interpreted image d) reliability image



Figure 2: a) original mammogram b) segmented image c) circumscribed masses d) reliability image

conditional probability mass function of the label l estimated by Equation (4). This information is displayed as an image where each pixel value is proportional to the marginal conditional probability associated with classified pixel, i.e. larger graylevel indicates higher reliability of classification. We have tested our algorithm on all the mammograms in the database that contain Microcalcifications, circumscribed masses, and stellate lesions. Figures 1 and 2 show some of our segmentation results. Our algorithm detected 100% of the abnormal tissues in the 7 mammograms that contain circumscribed masses, 65% of the 17 mammograms that have microcalcifications, but only 58% of the 12 mammograms that have stellate lesions. The number of false classifications is larger with stellate lesions because our observation model is not very good at modeling locally radiating structures such as stellate lesions.

CONCLUSIONS

We developed the above image processing algorithms using Matlab codes and applied over to several images from the miniMIAS data base. The results are found to be satisfactory and they have been validated by expert radiologists.

Further we are incorporating the above algorithms for the development of CAD system for early detection of breast cancer.

REFERENCES

1. Chan HP, Doi K, Vyborny CJ, Lam KL, Schmidt RA. Computer-aided detection of microcalcifications in mammograms: methodology and preliminary clinical study, *Investigative Radiology* 1988;23:664,671.
2. Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schmidt RA. Computer-aided detection of clustered microcalcifications on digital mammograms, *Medical and Biological Engineering and Computing*, 1995;33:174,178.
3. Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography. *Am J Roentgenol* 1994;162:699,708.
4. Kegelmeyer WP Jr. Computer detection of stellate lesions in mammograms, *Proceedings of the SPIE Conference on Biomedical Image Processing and Three-Dimensional Microscopy*, February 10-13 1992, San Jose, California, pp. 446,454.
5. Brzakovic D, Luo XM, Brzakovic P. An approach to automated detection of tumors in mammograms, *IEEE Transactions on Medical Imaging* 1990;9:233,241.
6. Kegelmeyer WP Jr, Pruneda JM, Bourland PD, Hillis A, Riggs MW, Nipper ML. Computer-aided mammographic screening for spiculated lesions. *Radiology* 1994;191:331,336.
7. Comer ML, Delp EJ. Parameter estimation and segmentation of noisy or textured images using the EM algorithm and MPM estimation, "Proceedings of the 1994 International Conference on Image Processing, November 13-16 1994, Austin, Texas, pp. 650,654.
8. Comer ML, Delp EJ. Multiresolution image segmentation, *Proceedings of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Detroit, Michigan; 1995. pp. 2415-8.
9. Comer ML. Multiresolution Image Processing Techniques with Applications in Texture Segmentation and Nonlinear Filtering. PhD thesis, School of Electrical Engineering, Purdue University, December 1995.
10. Comer ML, Delp EJ. The EM/MPM algorithm for segmentation of textured images: Analysis and further experimental results, to appear *Proceedings of the IEEE International Conference on Image Processing*, September 16-19 1996, Lausanne, Switzerland.
11. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721-741, November 1984.