CLASSIFICATION OF MAMMOGRAPHIC IMAGES USING THE JOINT BEST BASIS AND THE APPROXIMATE KLT

Sol Neeman
Johnson and Wales University
U.S.A

Nathan Intrator
Brown University
U.S.A

ABSTRACT

Breast cancer is currently one of the major causes of death for women in the U.S. Mammography is currently the most effective method for detection of breast cancer and early detection has proven to be an efficient tool to reduce the number of deaths. Mammography is the most demanding of all clinical imaging applications as it requires high contrast, high signal to noise ratio and resolution with minimal x-radiation. According to studies [16], 10\% to 30\% of women having breast cancer and undergoing mammography have negative mammograms, i.e. are misdiagnosed. Furthermore, only 20\%-40\% of the women who undergo biopsy have cancer. Biopsies are expensive, invasive and traumatic to the patient. The high rate of false positives motivate research aimed to enhance the mammogram images, to provide Computer Aided Diagnostics tools that can alert the radiologist to potentially malignant regions in the mammograms and to develop tools for automated classification of mammograms into benign and malignant classes (see for example [4, 8]). In this paper we present classification results of mammographic images from an early stage of malignancy using feature vectors based on wavelet packets, PCA and the Approximate Karhunen Loeve transform. We employ an innovative method that provides classification results better than the average performance of radiologists. The method was tested using database of mammograms from an early stage of malignancy. Correct detection is harder and more important at an early stage of malignancy.

KEY WORDS

Breast cancer, classification of mammograms, Wavelet analysis.

Introduction

Wavelet and wavelet packet analyses found to provide orthogonal transformations with high value of TCG (Transform Coding Gain). While other transforms (e.g. the Discrete Cosine Transform) may achieve a better compression ratio, they do not have the adaptive property of the wavelet transform. In other transforms, the rule of zeroing small coefficients is applied evenly and globally over all detail coefficients while the wavelet transform is adaptive in this respect. It allows preserving small coefficients that may account for 'important' minute features and may be useful for feature extraction and classification. In this work, we use three transformations to derive feature vectors for the classification of mammographic images.

Best Basis and the Joint Best Basis

The first set of feature vectors is based on the wavelet packet analysis. Compared to wavelet analysis, wavelet packet analysis provides a finer partition of the frequency space. This results in a richer family of basis functions (with respect to a certain wavelet function) offering a larger number of bases, some of them orthogonal. The family of basis functions can be searched for a 'best basis' with regard to some information cost function. We will use entropy as the cost function, as it will provide a good measure of the compression property of the basis. A fast algorithm developed by Coifman and Wickerhauser [14] searches for the best basis with computational complexity of *ON(logN)* steps where *N* is the length of the data. This basis has the smallest reconstruction error among all possible bases in the family of basis functions when the signal is reconstructed from a subset of its wavelet packet coefficients (a subset with the largest magnitude).

The concept of best basis can be extended to a family of signals or vectors. Given an ensemble of vectors, a wavelet packet analysis can be applied to the ensemble. The wavelet packet coefficients can be used to construct a wavelet packet table from which one may derive the Joint Best Basis [14]. This basis best represents the ensemble (with respect to some information cost function) among all possible bases offered in the joint packet table.

Principal Component Analysis (PCA)

Also known as the Karhunen-Loeve transform (KLT), is the best-known tool for multivariate analysis. It provides an orthogonal transformation in which the original set of observations is transformed to a new set of de-correlated coordinates [5]. This results in a new distribution of the population's variance, where most of the variance is concentrated in a fewer number of coordinates (called the principal components). This reduction of dimensionality is useful for compression and in certain cases some properties or features of the population may be associated with a smaller number of coordinates. In the case of multivariate normal distribution, PCA will provide the highest TCG (Transform Coding Gain) among all orthogonal transformations.

The third transformation used in this work is the *Approximate Karhunen Loeve Transform* [14], which combines the Joint Best Basis, and the PCA transforms. Given a vector population $X=\{X_1, X2, XN\}$, first its Joint Best Basis is computed, and then the coefficients of the vector population are derived in the Joint Best Basis. A subset of these coefficients, corresponding to the coordinates with the largest variance values, is truncated from the complete set of coefficients and PCA is applied to the truncated subset of coefficients to further achieve dimensionality reduction. The Approximate KLT provides a basis that may preserve features of interest (due to the Joint Best Basis), while achieving a good compression ratio (due to the application of PCA to the coefficients in the Joint Best Basis).

Feature Extraction and Classification

In mammographic images, some characteristics such as sharpness, regularity of the border lines of the calcification points, etc., are important factors in discriminating between benign and malignant tissues. These characteristics are local both in space and frequency, and the conventional Fourier analysis techniques are not useful in detecting them. Extracting the relevant features from the class of images is the first and the crucial step in the classification problem. Images of 128x128 pixels cannot be analyzed directly due to their high dimensionality. We will provide a method for image representation that will be useful for feature extraction and classification. As feature vectors, we use the *accumulated variance* of an ensemble of vectors associated with a mammographic image.

Formally, given an ensemble of vectors $V=\{v_1, v2, vN\}$, and assuming the average of the vector ensemble is zero, the variance of the pth coordinate is given by:

$$\sigma_p^2 = (1/N)\sum_{i=1}^{N} v_p^i$$

where the sum is over the pth coordinate of the vector population. The accumulated variance, *AccVar*, is a vector whose k^{th} entry, *AccVar(k)*, is given by:

$$AccVar(k) = \sum_{p=1}^{k} \sigma_p^2$$

METHOD

Mammographic Data Base

The method of classification used in this work was applied to a set of processed mammograms from a well-known mammographic database from Nijmegen, the Netherlands, which can be found in the Digital Data Base for Screening Mammography (DDSM) of the University of South Florida Digital Mammography Home Page. The mammograms we use consist of 105 ROI's (regions of interest) contributed by the University of Bologna, Italy to Dr. Nathan Intrator from Brown University. Each of the 105 ROI's, is of size 128x128 pixels derived from screen film mammograms with a pixel size of 0.1 mm and a 12-bit gray scale and is large enough to contain a few micro calcifications or the majority of micro-calcifications in a cluster. The mammographic images contain 29 benign and 76 malignant regions. The mammograms are from an early stage and come from general screening of women population and not from a population with pathological indicators (e.g. pain, lump in breast, asymmetry in breasts).

Indicators for Breast Cancer

The two main indicators associated with breast cancer are micro calcification clusters and masses. Micro calcifications appear in mammograms as tiny areas (with a size of a few pixels in digitized images or about .2mm in diameter) that are slightly brighter than the background. Micro calcifications clusters are not always easy to detect. Radiologists observe them in 30%-50% of all malignant mammograms, but in pathological examination, 80% of breast carcinomas contain micro calcifications [2].

Feature vectors based on shift invariant statistics

Since the number of mammograms is not large, there is no point in analyzing an image as a whole, e.g. with Principal Component Analysis (PCA) or wavelet packet analysis, as the high dimensional space for such representation is extremely sparse. Rather, we would represent each mammographic image by a collection of segments sampled with overlapping regions to capture the shift invariant statistics of the image.

Most of the background structure in both classes (benign and malignant) is similar, and small segments of size 8x8 pixels are sufficiently large to contain differences relevant to classification points (differences between various characteristics of calcification points such as shape, irregularity, etc.).

We experimented with 3 methods of deriving feature vectors for the mammographic images. In the first, a common Joint Best Basis was derived from a very large number of segments sampled from a training subset of benign and malignant mammograms. Then the mean and variance of the wavelet coefficients of the segment samples are computed for each class (benign and malignant). To classify a test image we first compute the wavelet coefficients of the collection of segments sampled from the image. We then classify each segment based on the distance (normalized by the variance) of its wavelet coefficients from the mean of each class. The image is classified to be benign if the majority of its segments are classified as benign.

In the second method, two Joint Best Bases were computed, one derived from a large number of segments sampled from a training subset of benign images and the other from a training subset of malignant images. To classify a test image, we first compute the wavelet coefficients of its segments in each base, and then classify each segment based on the distance of its wavelet coefficients in each base from the mean of that class. The image is classified to be benign if the majority of its segments are classified as benign.

In the third method, each image was represented individually by a large collection of overlapping segments. The accumulated variance of the segment population in a certain base is a 'signature' of that image. In this method the accumulated variance in various bases serves as a feature vector in the classification phase. Figure 1 is the plot of the accumulated variance in different bases of the segment population sampled from one of the benign images.

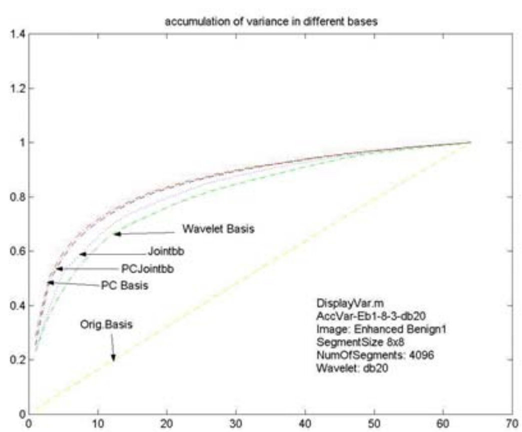

Figure 1: Accumulated variance of a collection of segments sampled from a benign mammogram, in different bases

Image Enhancement as a Preprocessing Step

To enhance features of interest and reduce the background structure, we apply a variant of *local averaging normalization*. We first remove the DC component at the image level, and then apply two local neighborhood normalizations to the image.

Classification Framework for Mammographic Images

The classification framework involves two steps. The first is to find a transformation that provides coefficients (feature vectors) that can be used to discriminate between benign and malignant images. The second step is applying a classifier to the extracted feature vectors. We experimented with both Fisher's Linear Discriminant Analysis (LDA) and K-nn (K-Nearest Neighborhood, the multivariate version).

Evaluation of Feature Vectors

To evaluate the performance of each feature vector used for classification, we applied the *Jackknife method* [16]. We created 50 sets of training and test images, where we use 70% of the 29 benign and 76 malignant mammograms for training and the rest for testing the classification performance. The results of the 50 experiments was averaged and presented as the *average error of misclassification,* the *sensitivity,* and the *specificity* of the classification.

To compare the performance of the various feature vectors, we compare their average results.

Experimental Results

Effects of image enhancement

Image enhancement is the first step we apply to improve the performance of classification. Figure 2 shows an unprocessed and enhanced mammogram along with their frequency spectrum. Note that the strong white background structure is attenuated in the enhanced version, emphasizing the details. In the frequency domain, the very low frequency components in the unprocessed image are dominant. The low frequency components account for the background structure. The high frequency components account for the details including the characteristics of the calcification points (e.g. intensity, sharpness, smoothness or irregularity in the border line of the calcification points) and borderlines of masses. The enhancement of the image distributes its energy more evenly along the frequency spectrum.

The effect of image enhancement can also be seen in the histogram of pixels' intensities. Figure 3 is a histogram plot of the third benign and malignant mammograms, both for the unprocessed and the enhanced images. The enhancement has a normalization effect on the distribution of the pixel intensities of the image.

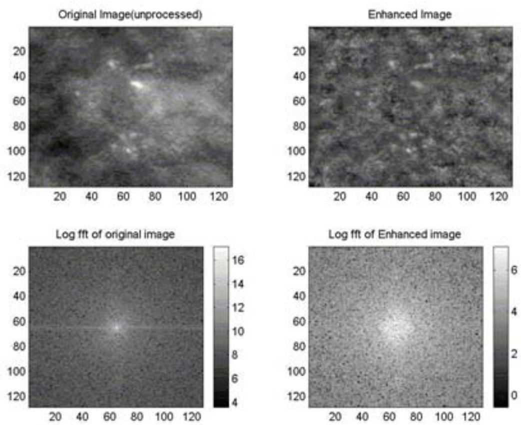

Figure 2: Top row: unprocessed and enhanced mammograms. Bottom row: their respective spectrum

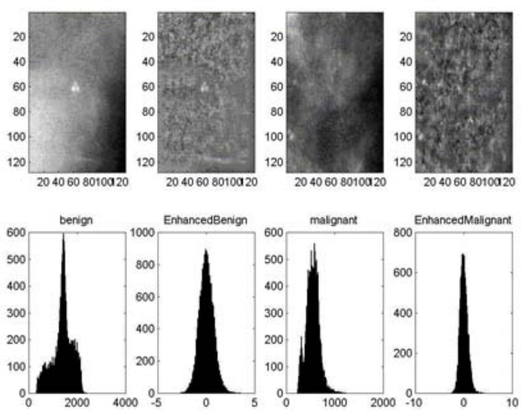

Figure 3: Top row from left: unprocessed benign, enhanced benign, unprocessed malignant, enhanced malignant. Bottom: their respective intensity histogram

Classification results

Feature vectors based on the Common Joint best basis and a set of two Joint Best Bases did not provide useful results for classification. The reason for that may be the great variance in the location (in the spatial-frequency space), of the discriminating features.

The third method, in which each image was represented individually by a collection of overlapping segments sampled from the image, provided the best results. The feature vectors for each image is the accumulated variance of the wavelet coefficients of the collection in the bases corresponding to PCA, the Joint Best Basis and the Approximate KLT transforms.

We experimented with various wavelets, segment size for both the unprocessed and processed images. The best results were achieved using feature vectors based on the db20 wavelet. With this wavelet we achieved a sensitivity of 88.1%, specificity of 41.1% and an average error of 24.2%.

The next table compares the sensitivity and specificity of our method to the average performance of radiologists and figure 4 provides the results of 50 experiments in graphical form.

	Sensitivity	Specificity
Our results	88	41
Average performance of radiologists	80	20

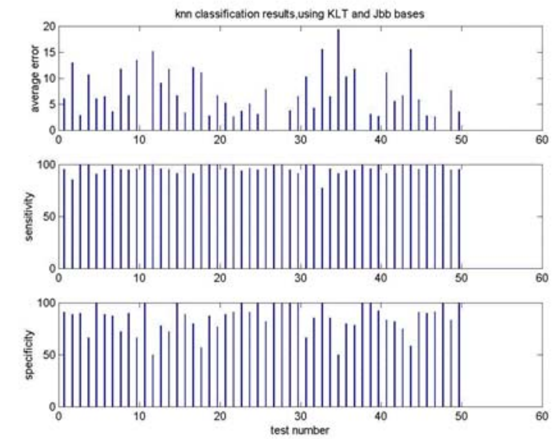

Figure 4: Classification results of 50 experiments. From top: average error, sensitivity and specificity

We emphasize that the mammograms used in these experiments are from a general screening of women population. Therefore they are harder images (in terms of discrimination) when compared to mammograms taken by women due to some pathological indications (e.g. pain, lumps in the breast, asymmetry in the breasts).

Summary

We have employed an innovative method to classify mammographic images from an early stage of malignancy. The method is based on enhancing each image and representing it by an ensemble of segments that capture its shift invariant statistics. The ensemble is used to extract feature vectors based on coefficients in various bases: the Approximate KLT basis, the KLT basis and the Joint Best Basis. When combined with the knn classifier we achieved results that are better than the average performance of radiologists.

We experimented with various wavelets and found the db20 to provide the best results, compatible with previous research that suggested this filter is best for detecting calcification points [16].

The database used in this study was taken from a general population of women. Also the tumors in the malignant cases are very small and therefore very difficult to detect.

Radiologists' performance at this stage of the disease is significantly lower than the performance of more advanced stages. Detection of breast cancer at early stage enables treatment which is much more effective, less invasive and inexpensive. Having achieved encouraging classification results, our method can provide a second opinion to radiologists in the diagnosis process.

REFERENCES

[1] Bird, R.G., T.W. Wallace, and B.C. Yankaskas, Analysis of cancers missed at screening mammography, *Radiology*, 184:613-617, 1992.

[2] Cox, J.D., S.R. Sharma and R.B. Schilling, Advanced Digital mammography, *Proceedings of Computer Aided Radiology '* 97, 11-16.

[3] FM, Hall , storella JM, Silverstone DZ, Wyshak G., Nonpalable breast lesions: recommendation for biopsy based on suspicion of carcinoma at mammography, *Radiology* 1988, 167:353-358.

[4] Giger, M.L., R.M.Nishikawa, M.Kupinski, U.Bick, et al., Computerized detection of breast lesions in digitized mammograms and results with a clinically-implemented intelligent workstation, *Proceedings Computer Aided Radiology* ' 97, 325-329.

[5] Gonzalez, Rafael C. and Richard E. Woods, *Digital Image Processing*, Addison Wesley, 1992.

[6] Jolliffe, I.T., *Principal Component Analysis*, Springer-Verlag, 1986.

[7] Kallergi, M. , L. Clarke, W. Qian, M. Gavrielides, P. Venugopal, C. Berman, S. Holman-Ferries, M. Miller, R. Clark, Interpretation of Calcifications in Screen/Film, Digitized, and Wavelet-Enhanced Monitor-Displayed Mammograms: A ROC Study, *Academic Radiology* 1996, 3:2285-293

[8] Mascio, L., J. Hernandez and C. Logan, Lawrence Livermore National Laboratory, Automated Analysis for Micro-calcifications in High Resolution Digital Mammograms, *http://www-sed.llnl.gov/documents /imaging/jmhsie93.html*

[9] Nishikawa, R.M. , M.L. Giger, et al., Automated Classification of breast Lesions on Digital Mammograms, *Computer Aided Radiology* '97, 347-351.

[10] Ritter, Gerhard X. and Joseph N. Wilson, *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, 1996.

[11] Russ, John C., *The Image Processing Handbook*, CRC Press, 1995

[12] Scharf, Louis L., *Statistical Signal Processing*, Addison Wesley, 90'.

[13] Smith, John R. and Shih-Fu Chang, Transform Features for Texture Classification and Discrimination in Large Image Databases, *Proceedings of IEEE*, Austin, Tx, Nov., 1994

[14] Wickerhauser, Mladen V., *Adapted Wavelet Analysis from Theory to Software*, A K Peters, 1994.

[15] Wu, Y. C. , M.T. Freedman, A. Hasegawa, R. Zurbier, Shih-Chung, S.K.Mun, Classification of Micro-calcifications in Radiographs of Pathologic Specimens for the Diagnosis of Breast Cancer, *Academic Radiology* 1995; 2:199-204

[16] Yoshida, H., K. Doi, R. Nishikawa, M. Giger and R. Schmidt, An improved Computer-Assisted Diagnostic Scheme Using Wavelet Transform for Detecting Clustered Micro-calcifications in Digital Mammograms, *Academic Radiology* 1996; 3:621-627

[17] Yoshida, H., R. Nishikawa, M. Giger and K. Doi, Signal /background separation by wavelet packets for detection of micro-calcifications in mammograms, *the International Society for Optical Engineering (SPIE)* 96' Vol 2825, pp 805-811.

[18] Yoshida, H., R. Nishikawa, M.Giger, and K.Doi. Optimally Weighed Wavelet Packets for Detection of Clustered Micro-calcifications in Digital Mammograms, *DIGITAL MAMMOGRAPHY* '96. pp 317-322.