Evaluation of Bayes Risk Weighted Vector Quantization with Posterior Estimation in the Detection of Lesions in Digitized Mammograms

Cheryl L. Nash Keren O. Perlmutter Robert M. Gray
Information Systems Laboratory, Stanford University, Stanford, CA 94305-4055

Abstract

The automated detection of suspicious tissue in digital mammograms can provide a useful aid to diagnosis by permitting a radiologist to see all regions deemed suspicious by the computer. We apply to digital mammography a method that combines aspects of data compression techniques based on clustering and decision trees together with algorithms for classification and regression. The idea is to use a distortion measure in a clustering algorithm that includes both squared error for general appearance and average Bayes risk for classification accuracy. The algorithm structure is that of a vector quantization compression system that incorporates Bayes risk into the optimization algorithm.

1 Background

Currently X-ray-mammography is the best method for early detection of breast cancer [1] having a detection accuracy of 85–95% for small lesions. These small lesions are typically in the form of tiny non-palpable calcifications. Unfortunately radiologists miss an average of 10% of these cancers when present on the film during everyday screenings. These oversights tend to be caused by fatigue, distraction, or other reasons [4] of a similar nature. It is in this initial screening phase that computer-aided diagnosis (CAD) techniques can greatly assist the radiologist in the detection of abnormalities such as microcalcifications and masses.

In order to make use of CAD techniques, the mammogram images need to be in digital format. Digital X-ray-mammograms are most commonly produced by digitizing the analog films; however, there currently exist devices that will acquire the image directly in a digital format. Both methods produce large files that can quickly compound as different views are taken for each patient over two or more years. A single full frame digitized mammogram can range from 7 Megabytes (low resolution) to 175 Megabytes (high resolution) and a typical screening images two views of each breast. Storage and transmission of large quantities of these digital images are two important reasons for use of compression algorithms in addition to the CAD. The ability to simultaneously compress and classify thus can allow a physician to view a compressed image with the suspected anomalies highlighted, and provides the ability to do tele-radiology and to have easier storage capabilities.

2 Compression and Classification

Classification and compression both consist of the mapping of real valued vectors (such as pixel intensity blocks in an image) into a finite set. If the goal is classification, the set is a collection of classes such as tumor and nontumor. However, if the goal is compression, the set is a collection of templates or reproduction codewords. The history of these two fields is intertwined and many similar algorithms have been developed for the two applications, such as the CART^{TM} [2] (classification and regression tree) algorithm for classification trees and tree-structured vector quantization for compression. Both are designed by optimally trading off a measure of quality, such as mean squared error for compression or Bayes risk for classification, with a measure of rate or complexity, such as the average number of bits required to make a decision or represent the template.

By incorporating a Bayes risk term into the usual distortion measure for a compression system, one can simultaneously optimize a code with respect to compression, classification or any weighted combination of the two. This is done within the framework of vector quantization (VQ), since there exists an intimate connection between the algorithms used to design and implement vector quantizers for compression and those for statistical classification that provides a natural

1058-6393/95 $4.00

means of jointly optimizing the two goals. The input is a joint random process $\{X(n), Y(n); n = 0, 1, \ldots\}$, where the $X(n)$ are k-dimensional real-valued vectors and the $Y(n)$ designate membership in a class and take values in a set $\mathcal{H} = \{0, 1, \cdots, M-1\}$. The VQ-based classifier operates solely on the observed sequence X and consists of an encoder α, which views only X and outputs a binary index $i = \alpha(X) \in \mathcal{I}$ and a decoder β, which maps the indices into the reproduction vectors $\beta(i) = \hat{X}_i$ and a class label $\delta(i) \in \mathcal{H}$. Because the index i is used to both decompress and classify the vector, the classification is implicit in the compression, and hence no additional computation or bits are required.

The quality of the reproduction $\hat{X} = \beta(\alpha(X))$ for an input X is measured by a nonnegative distortion $d(X, \hat{X})$. Squared error distortion, $d(X, \hat{X}) = ||X - \hat{X}||^2$, is considered for simplicity. The average distortion for compression

$$D(\alpha, \beta) = E[d(X, \beta(\alpha(X)))] \tag{1}$$

is then the mean squared error (MSE). The quality of the classifier is measured by the Bayes risk, defined as

$$\begin{aligned} B(\alpha, \delta) &= \sum_{i=0}^{N-1} P(\alpha(X) = i) \\ &\quad \times \sum_{j=0}^{M-1} C_{j,\delta(i)} P(Y = j | \alpha(X) = i) \end{aligned} \tag{2}$$

where the cost $C_{j,k} \geq 0; k = \delta(i)$ represents the cost incurred when a class j vector is classified as class k, where $C_{j,k} = 0$ if $j = k = \delta(i)$. These costs can be chosen to reflect the fact that different classification errors can have different consequences.

3 Bayes Risk Weighted VQ with Posterior Estimation

The Bayes VQ system [5, 7, 9, 8] consists of two cascaded vector quantizers. The design of these vector quantizers is based on the empirical distribution $P_{\mathcal{L}}$ induced by a training set of labeled data $\mathcal{L} = \{(x_n, y_n); n = 1, \ldots, |\mathcal{L}|\}$. The first stage in the system, a tree-structured vector quantizer (TSVQ), provides an estimate of the posterior conditional probabilities required in (2). It is designed in a manner analogous to the CART^{TM} algorithm to generate estimates of $P(Y = l | X = x) = P_{\mathcal{L}}(l|x)$. The TSVQ is designed to minimize the distortion given by the average relative entropy $D(P_{\mathcal{L}}||\hat{P}) = \sum_{x \in \mathcal{L}} P_{\mathcal{L}}(x) \sum_{l \in \mathcal{H}} P_{\mathcal{L}}(l|x) \log \frac{P_{\mathcal{L}}(l|x)}{\hat{P}(l|x)}$ between the estimated probability and the empirical probability. Squared error determines which node is used to encode a vector within the tree. The probability estimate for any node is then given by the relative frequencies of the class labels given the encoder output. These encoded estimates are used in creating the second VQ and are not communicated to the decoder.

The second VQ, either full search or tree-structured, incorporates the simultaneous optimization of both compression and classification by using a modified distortion measure that contains both squared error for general appearance and Bayes risk for classification accuracy. These two error measures are combined with a Lagrangian importance weighting to form the modified distortion measure

$$\rho_{\lambda,\hat{P}}(x, \hat{x}, l) = ||x - \hat{x}||^2 + \lambda \sum_{j=0}^{M-1} C_{j,l} \hat{P}(Y = l | x)$$

so that

$$\begin{aligned} J_{\lambda,\hat{P}}(\alpha, \beta, \delta) &= E[\rho_{\lambda,\hat{P}}(X, \beta(\alpha(X)), \delta(\alpha(X)))] \\ &= D(\alpha, \beta) + \lambda B(\alpha, \gamma), \end{aligned} \tag{3}$$

where the Bayes risk $B(\alpha, \delta)$ is given by (2) and the average distortion $D(\alpha, \beta)$ by (1). This VQ is designed using a descent algorithm, analogous to the Lloyd algorithm, that iteratively optimizes the encoder, decoder, and classifier for each other. The classifier is a minimum average Bayes risk classifier, defined by $\delta_{\text{Bayes}}(i) = \arg\min_k \sum_{j=0}^{M-1} C_{j,k} \hat{P}(Y = j | \alpha(x))$. The costs, $C_{j,k}$, are particularly important in the classification of mammograms since the consequences of misclassifying an abnormality (i.e. missing a tumor) are quite different than those of a false alarm. The Lagrangian parameter λ in (3) provides a flexible tradeoff between compression and classification priorities. In particular, when $\lambda = 0$, the focus for designing the encoder and decoder is purely on compression; we thus create an ordinary minimum MSE VQ. If a class label, optimized for the VQ encoder output, is subsequently attached to these vectors, the system is simply an independent design of a VQ and a classifier [6]. When $\lambda \to \infty$, we obtain a minimum Bayes risk classifier.

4 Results

In our study, we use 12 bit 100 micron resolution digitized mammograms that contain calcifications and/or masses. Two different training sets were formed. One training set consisted of mammograms

with calcifications and the other training set contained mammograms with masses. Each training vector consisted of a 2x2 pixel block of intensity values and a class label. We considered two test images taken from outside of the training sets. One test image contained calcifications and was encoded using the calcification codebook. The other test image contained masses and was encoded using the mass codebook.

We measure the compression performance of the encoded images by SNR, where $\text{SNR} = 10\log_{10}(D_0/D)$, D is the distortion measured by MSE, and D_0 is the distortion obtained on the test sequence using the best zero rate code. The classification performance is measured by three parameters: Bayes risk, sensitivity and specificity. Sensitivity is the percentage of tumor vectors that are correctly classified as tumor. Specificity is the percentage of nontumor vectors that are correctly classified as nontumor (i.e. it indicates the extent of false highlighting in a classified image).

In this study, we use a tree-structured compression code for the Bayes VQ system (BTSVQ design). We compare our results with an independent tree-structured design of classifier and quantizer (independent TSVQ design). Remember that this is the special case of $\lambda = 0$ in a Bayes VQ system. Note, however, that in the independent design, we use the empirical distribution $P_{\mathcal{L}}$ to compute the probabilities required by the Bayes classifier defined in (2) rather than the posterior estimates provided by the posterior estimating TSVQ. We also compare the performance of our system to Kohonen's "learning vector quantizer" (LVQ) [3]. LVQ implicitly designs a full search codebook to reduce classification error rather than reducing compression error. The encoder operates as an ordinary minimum squared error selection of a representative from the codebook. Because the algorithm does not explicitly consider compression in its design, we use the codebook only for classification purposes. An optimized version of this codebook, which replaces encoder codewords produced by LVQ by the centroids of all training vectors that map into them, is then used to provide the reproduction vectors [8]. For the LVQ design, the codebook was initialized using the LVQ PAK *propinit* algorithm and then designed using the *olvq1* with the modification discussed above. We note that LVQ does not have the capability to incorporate unbalanced costs into its design.

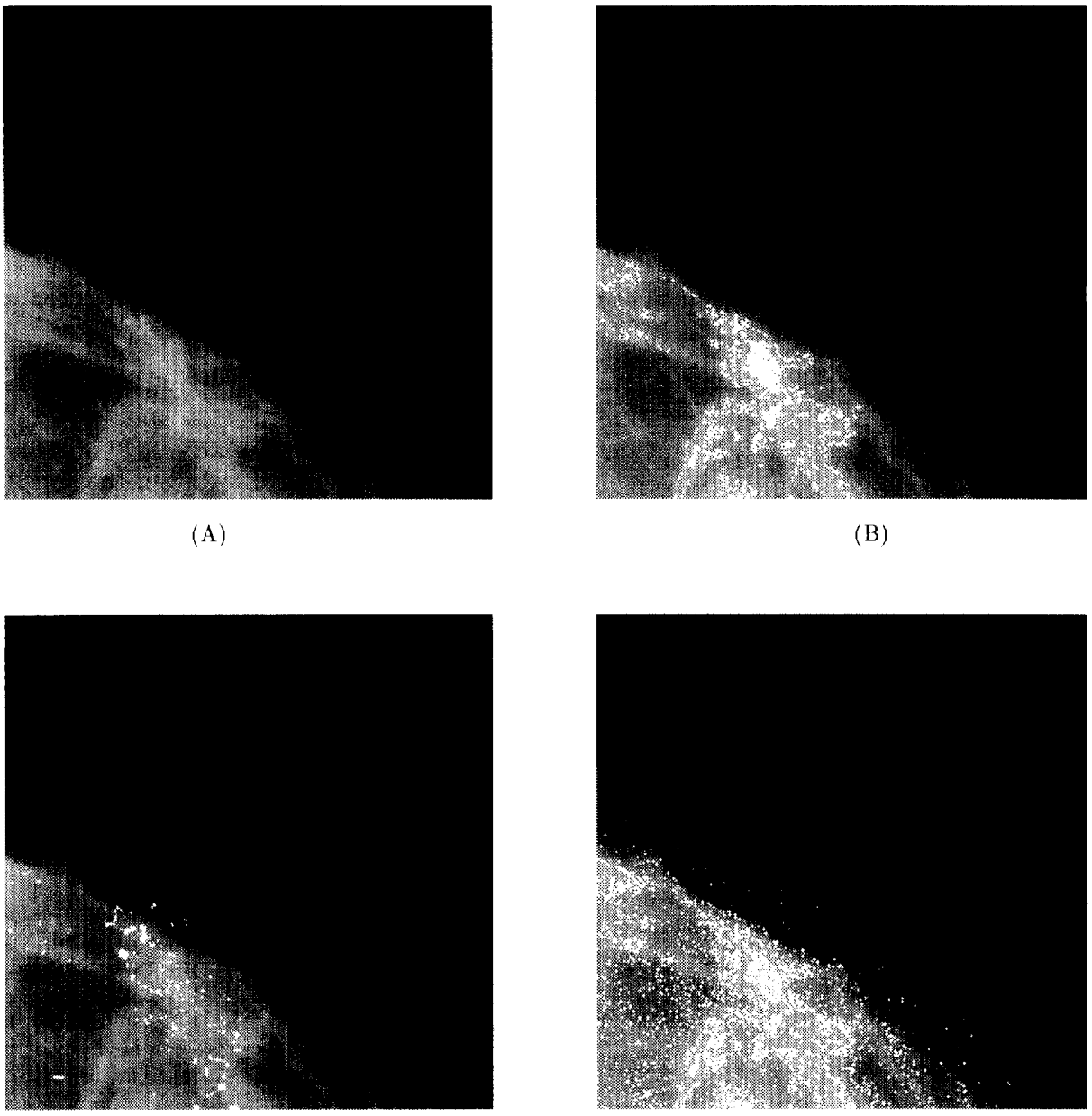
For the analysis of the mammogram with calcifications, we designated the cost of misclassifying a tumor vector as 50 times more detrimental than misclassifying a nontumor vector. We selected $\lambda = 10^5$ for the BTSVQ design. Figure 1 presents results obtained on a portion of the test image containing calcifications at 2bpp. Table 1 provides the corresponding statistical results.

The table indicates that the BTSVQ design produced lower Bayes risk and comparable SNR to the independent TSVQ design for the test image containing calcifications. The table also indicates that the BTSVQ design significantly outperformed LVQ with respect to both SNR and Bayes risk. The BTSVQ design yielded higher specificity but lower sensitivity to that obtained with the independent TSVQ design. Because of the preponderance of nontumor vectors, the 3.15% difference in specificity can affect the visual quality of the images considerably. This is illustrated in Figure 1, where we observe less false highlighting with the BTSVQ design than with the independent TSVQ design. Although the sensitivity obtained for the encoded calcification image using the Bayes TSVQ method seems low, if we use a criterion that is less stringent than the *pixel-by-pixel* definition that asks whether the algorithm has detected enough of the lesion to signal the radiologist, the answer to such a question would be affirmative.

For the analysis of the mammogram with a mass, we designated the cost of misclassifying a mass vector as ten times more harmful than a false alarm. The BTSVQ design again used $\lambda = 10^5$. Table 2 provides the statistical results obtained using the BTSVQ and independent TSVQ designs on the test image containing masses at 2 bpp. LVQ did poorly on this image (none of the mass vectors were classified correctly and its compression performance was very low), and as such a comparison with LVQ here is not meaningful.

We again observe that the the BTSVQ design produced lower Bayes risk and comparable SNR to the independent TSVQ design. In addition, the BTSVQ design yielded higher specificity but lower sensitivity to that obtained with the independent design. The BTSVQ design correctly identified the one mass in the corresponding test image section including its spiculation. The independent design, however, did not clearly point out the mass area. Although both algorithms produced false highlighting, the amount engendered by the BTSVQ design was considerably less than that of the independent design (a 5.1% difference).

For both calcification and mass images, the BTSVQ and independent designs produced a tradeoff between specificity and sensitivity, i.e., an increase in performance in one of the measures resulted in a decrease in performance in the other. We note, however, that only the Bayes TSVQ design has the ability to select a desirable ratio between these two measures using both



(A) (B)

(C) (D)

Figure 1: Compression and classification of digitized mammograms at 2 bpp for calcifications: (A) Portion of Compressed Mammogram using BTSVQ with posterior estimation (B) Compressed/Classified image using BTSVQ with posterior estimation (white highlighted areas denote pixel blocks classified as microcalcifications) (C) Original 12 bit image with microcalcifications highlighted in white (D) Compressed/Classified image using independent TSVQ design (white highlighted pixel areas denote pixel blocks classified as microcalcification)

Codebook design method:	SNR(dB)	Sensitivity	Specificity	Bayes Risk
Bayes TSVQ	29.2	41.19	92.60	0.106
Independent TSVQ	29.2	47.92	89.45	0.134
Kohonen's LVQ	19.3	59.48	55.08	0.471

Table 1: Statistical results of algorithms on mammogram images containing calcifications coded at 2 bpp.

Codebook design method:	SNR(dB)	Sensitivity	Specificity	Bayes Risk
Bayes TSVQ	32.96	49.45	70.80	0.351
Independent TSVQ	32.93	53.24	65.66	0.397

Table 2: Statistical results of algorithms on mammogram images containing masses coded at 2 bpp.

costs and the λ parameter.

5 Conclusions

The BTSVQ design produced lower Bayes risk and visually superior compressed/classified images compared to those obtained with the independent TSVQ design and with LVQ on the two test images we investigated. Although, the general area of the lesions were identified by the BTSVQ algorithm, we obtained some false highlighting as well – particularly with the mammograms containing masses. We note, however, that the λ value and costs inherent in the BTSVQ design allow a flexibility in the compression and classification performance. We are currently investigating these tradeoffs more thoroughly and believe (based on results on both simulated and real data [7, 9]) that the appropriate selection will provide modest or possibly significant improvement in the results. In addition, we believe that more sophisticated posterior estimators that use features that better characterize the mammogram images may help classification performance.

References

[1] I. Andersson. Mammography in clinical practice. *Med Radiography and Photography*, 62(2):2, 1986.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees.* Wadsworth, Belmont, CA, 1984.

[3] T. Kohonen, J. Kangas, J. Laaksonen, and K. Torkkola. LVQ_PAK: The learning vector quantization program package, version 2.1. Technical report, Helsinki University of Technology, Laboratory of Computer and Information Science, Finland, Oct 1992.

[4] J.E. Martin, M. Moskowitz, and J.R. Milbrath. Breast cancer missed by mammography. *Amer. J. Roentgenology*, 132:737–739, 1979.

[5] K. L. Oehler. *Image Compression and Classification using Vector Quantization.* Ph.D. Dissertation, Stanford University Electrical Engineering Department, September 1993.

[6] K. L. Oehler and R. M. Gray. Combining image compression and classification using vector quantization. *IEEE Trans. PAMI*, 1994. To appear.

[7] K. O. Perlmutter, R. M. Gray, K. L. Oehler, and R. A. Olshen. Bayes risk weighted tree-structured vector quantization with posterior estimation. In J. A. Storer and M. Cohn, editors, *Proceedings of the 1994 IEEE Data Compression Conference (DCC)*, pages 274–283, Snowbird, Utah, March 1994. IEEE Computer Society Press.

[8] K. O. Perlmutter, C. L. Nash, and R. M. Gray. A comparison of bayes risk weighted vector quantization with posterior estimation with other VQ-based classifiers. In *Proceedings of the IEEE International Conference on Image Processing*, Austin, Texas, 1994. To appear.

[9] K. O. Perlmutter, S. M. Perlmutter, R. M. Gray, R. A. Olshen, and K. L. Oehler. Bayes risk weighted vector quantization with posterior estimation for image compression and classification. Submitted to TIP, 1994.