Quantization and classification

Quantification et classification

par Robert M. GRAY[1]

Information Systems Laboratory, Department of Electrical Engineering, Stanford, CA 94305

résumé et mots clés

The problem of quantizer design for detection or classification has a long history, with classical contributions by Kassam, Poor, Picinbono, Bucklew and others. The goal was to design a quantizer such that a detection rule based on the quantized information was optimized. During recent years an alternative approach has been developed which seeks to jointly optimize quantization and classification by incorporating the Bayes risk resulting from the quantizer into the quantizer optimization. In this paper the general classical approach of Picinbono and Duvaut is compared contrasted with the joint approach and illustrated by a simple example.

Signal quantization, signal detection, optimal classification, density estimation

abstract and key words

Il existe une importante littérature traitant du problème de la conception d'un quantificateur pour un système de détection ou de classification. A l'origine, les travaux menés dans ce domaine – notamment par Kassam, Poor, Picinbono et Bucklew – ont pour but de concevoir un quantificateur qui optimise une règle de décision basée sur l'information quantifiée. Rompant avec cette approche classique, ces dernières années ont vu l'émergence d'une approche alternative dont l'objectif est d'optimiser conjointement les opérations de quantification et de classification. L'optimisation conjointe est réalisée par minimisation d'un critère Lagrangien comprennant l'erreur quadratique moyenne (quantification) et le risque de Bayes (classification). Dans cet article, nous proposons de comparer l'approche conjointe à l'approche classique, plus courante, de Picinbono et Duvaut. Nous illustrons les deux méthodes à l'aide d'un exemple simple.

Quantification, détection, classification, classification optimale par risque de Bayes, estimation de densité.

1. introduction

A classical problem in signal detection or statistical classification arises when one is provided quantized observables instead of the original observables. If the quantizer is out of the control of the detection system designer, then the solution is to simply replace the original observations with the quantized observations in any statistical analysis. Although simple in principle, this can seriously complicate the analysis, leading to the common assumption that the quantization is high resolution and leaves the observations asymptotically unchanged or that one can use the Bennett high rate quantization approximations to analyze the problem. More importantly, the quantization might be part of the design so that it can be optimized for the specific application. In this case it is clearly suboptimal to simply design a minimum mean squared error quantizer (e.g., using the Lloyd algorithm) and then design an optimal detector for the output of the quantizer. The cascade of two separate optimization problems is unlikely to provide an overall optimal solution to the true goal of providing an overall design for quantizer and detector which is optimal given whatever rate constraint is placed on the quantizer. Alternatively, a quantizer designed to minimize mean squared error may inadvertantly lose information necessary for good detection performance.

A variety of early solutions to this problem considered ways of designing quantizers so as to preserve the needed information for detection, usually by attempting to ensure that the quantizer

1. This work was supported by the National Science Foundation under NSF Grant No. MIP-931190 and by gifts from Hewlett-Packard, Inc., the Okawa Foundation, and SK Telecom.

preserved distinct conditional probability distributions under the hypotheses to be decided as in [2, 3, 4, 5, 10] or by using Bennett-style high rate approximations to approximate the conditional densities resulting from quantization [6, 7]. Important to the nonasymptotic approaches of interest here was the fact that quantization was taken in the general sense of being described by a partition $\mathcal{S} = \{S_i\}$ of the real line $\Re$ for a scalar quantizer or $\Re^k$ for a vector quantizer and a corresponding collection of output labels $\mathcal{C} = \{c_i\}$ so that the quantizer produced an output of c_i if the input $x \in S_i$. In particular, the quantizer did not necessarily operate as a nearest neighbor or minimum distortion with respect to squared error or Euclidean distance, more general (non-polygonal) partition cells S_i were permitted. Picinbono and Duvaut [8, 9] solved this general problem by designing the quantizer specifically to optimize the detector in the sense of minimizing the probability of error. This was accomplished by optimizing the partition by maximizing its "deflection" - a measure of how spread out the distributions of the quantizer output distributions were, and then optimizing the quantization values for the partition. The partition optimization was shown to be effectively a quantization of the likelihood ratio for the given hypotheses and the observed input vector. This reduces to a partition of the input space that is a refinement of the partition of the space by a Bayes-optimal classifier. The formulation permits a comparison with the performance of suboptimal quantizers constrained to have rectangular cells and an analysis of the special case of vectors with independent components.

The idea that a vector quantizer with a suitable encoder can be used in detection or classification problems has been much studied in recent years. Historically, the nearest neighbor classifiers popular in statistics [12, 13] use a minimum mean squared error vector quantizer for classification, as does Kohonen's "learning vector quantizer" [14, 15, 16]. These quantizers all use a minimum squared error encoder (MSE-encoder), but they are designed so as to minimize empirical classification error. The asymptotics of such quantizer-based classifiers or partition-based classifiers has been treated in great detail in Pollard [17] and Devroye et al. [18] An obvious shortcoming of this approach is that the Voronoi cells of an MSE quantizer are polygonal, which means they are not always a good fit to the optimal cells from a detection standpoint. If the quantizer rate is high, there will be many such cells and a good approximation can be achieved, but in the nonasymptotic situation the fit can be quite poor.

Two situations of interest are not covered by the general solution of Picinbono and Duvaut. The first is the case where the distributions are not known a priori and must be deduced from training or learning data. The second is the situation where detection or classification is important, but it is not the only operation to be applied to the quantized data. An example of both situations is the detection of anomalies in a medical image. The multidimensional distributions of groups of pixel intensities and the conditional distributions of, say, microcalcifications are not known in advance, but there is a wealth of labeled examples available (e.g., http ://marathon.ccsee.usf.edu/Mammography/Database.htm). Furthermore, the quantization occurs in order to speed transmission and minimize storage requirements, but it is not only the detection information that is important : a radiologist is likely to inspect the reconstructed image for other things as well, so it is important that the reproduction look as good as possible for the given bit rate. These considerations motivate the joint design of a quantizer and classifier not only for the usual goal of quantization (minimum average squared error or similar distortion measure) or of classification (minimum probability of error or, more generally, Bayes risk), but with the dual goal of minimizing both cost functions. The idea of joint classification and compression dates back at least to Hilbert in 1977 [19].

Our focus here is on a reformulation of the quantization problem which includes the detection or classification aspect, but also retains a squared error (or other) distortion measure. The basic idea is simple : a distortion measure to measure general image quality such as squared error is combined with a Bayes risk term to measure classification accuracy using the quantized output. Varying the Lagrange multiplier provides a user to balance the relative importance of the two tasks and the extreme cases yield the obvious special cases : Bayes optimal classification with incidental quantization or minimum distortion quantization with incidental classification. An example shows that in some cases one can have essentially optimal performance with respect to one cost function (e.g., Bayes risk) with only slightly suboptimal performance with respect to the other (e.g., mean squared error). The code structure is called "Bayes-risk weighted vector quantization" or simply "Bayes vector quantization" and many of the issues treated here are considered in more detail along with applications to image segmentation in [20, 21, 22, 23, 24, 25, 26, 27].

2. Bayes Vector Quantization

A training sequence $\mathcal{L} = \{(x_n, y_n),\ i = 1, 2, \ldots, L\}$ which is a sample of a random process $\{(X_n, Y_n),\ i = 1, 2, \ldots\}$ is observed, and the individual (X_n, Y_n) are assumed to have a common, but unknown, distribution P_{XY} on a generic $(X, Y) \in A_X \times A_Y$, where A_X might be k-dimensional Euclidean space and $A_Y = \{1, 2, \ldots, M\}$. Typically P_X is absolutely continuous and is described by some pdf f_X on $\Re^k$, and P_Y is discrete, described by some pmf p_Y. We will later consider the case where P_{XY} is not known and must be estimated based on $\mathcal{L}$. The goal is to design a vector quantizer for data X based on $\mathcal{L}$ that provides a good tradeoff among average distortion, bit rate, and the Bayes risk entailed by guessing Y from the encoded X.

A fixed rate Bayes vector quantizer (BVQ) consists of the following components :

Encoder/Partition $\tilde{\alpha} : A_X \to \mathcal{Z}$, where $\mathcal{Z} = \{0, 1, 2, \ldots, |\mathcal{Z}| - 1\}$. Equivalently, the encoder is described by a partition $\mathcal{S} = \{S_i;\ i \to \mathcal{Z}\}$ where $\tilde{\alpha}(x) = i$ if and only if $x \in S_i$.

Decoder/Reproduction Codebook $\tilde{\beta} : \mathcal{Z} \to \mathcal{C}$, where $\mathcal{C} = \{\beta(i);\ i \in \mathcal{Z}\}$ where $\beta(i)$ is the label or reproduction codeword corresponding to index i (and the corresponding partition cell S_i).

Classifier/Detector $\kappa : \mathcal{Z} \to A_Y$.

We assume a distortion measure $d(x, \hat{x}) = ||x - \hat{x}||^2 = \sum_{l=0}^{k-1} |x_l - \hat{x}_l|^2$. The average distortion resulting for a given code is $D(\tilde{\alpha}, \tilde{\beta}) = E[d(X, \tilde{\beta}(\tilde{\alpha}(X)))]$. The classification cost is given by the Bayes risk :

$$B(\tilde{\alpha}, \kappa) = \sum_{k=1}^{M} \sum_{j=1}^{M} C_{j,k} \Pr(\kappa(\tilde{\alpha}(X)) = k \text{ and } Y = j),$$

where $C_{j,k}$ is the cost of guessing $Y = k$ based on the encoded X when the true class is $Y = j$. For simplicity we assume that $C_{j,k} = 1 - \delta_{k,j} = 0$ for $k = j$ and 1 otherwise. In this case the average Bayes risk reduces to the probability of classification error.

A Lagrangian cost function is formed to incorporate the separate costs of the two "decoders" : squared error for the reproduction decoder and Bayes risk for the classifier. Given decoder $\tilde{\beta}$ and classifier κ, define the Lagrangian distortion between an input x and an encoder output i as

$$\rho_{\lambda,P}(x, i) = d(x, \tilde{\beta}(i)) + \lambda \sum_{j=1}^{M} C_{j,\kappa(i)} P(Y = j | X = x)$$

$$J_{\lambda,P}(\tilde{\alpha}, \tilde{\beta}, \kappa) = E[\rho_{\lambda,P}(X, \tilde{\alpha}(X))] = D(\tilde{\alpha}, \tilde{\beta}) + \lambda B(\tilde{\alpha}, \kappa)$$

There is anecdotal evidence to suggest that this joint approach may be useful even when the goal is purely classification or compression. The intuition is that the second distortion measure helps avoid local minima and uses an auxiliary criterion to help cluster the data.

3. Optimality properties of Bayes VQ

Following the Lloyd clustering approach to quantizer design, we describe necessary conditions for overall optimality. These properties yield a descent algorithm for designing the code [22]. The components of the code are the encoder $\tilde{\alpha}$, the decoder $\tilde{\beta}$, and the classifier κ. The goal is to minimize the Lagrangian distortion

$$J_{\lambda,P}(\tilde{\alpha}, \tilde{\beta}, \kappa) = D(\tilde{\alpha}, \tilde{\beta}) + \lambda B(\tilde{\alpha}, \kappa).$$

Optimal Decoder Given $\tilde{\alpha}$, κ, the optimal decoder is

$$\tilde{\beta}(i) = \operatorname*{argmin}_{y \in \hat{A}} E[d(X, y) | \tilde{\alpha}(X) = i],$$

the *Lloyd centroids* with respect to P_X (conditional means for MSE).

Optimal Classifier Given $\tilde{\alpha}$, $\tilde{\beta}$, the optimal classifier is

$$\kappa_{\text{Bayes}}(i) = \operatorname*{argmin}_{k} \{\sum_{j=1}^{M} C_{j,k} P(Y = j | \tilde{\alpha}(X) = i)\},$$

that is, the Bayes optimal classifier given the encoded input.

Optimal Encoder Given the κ, $\tilde{\beta}$, then the optimal encoder is

$$\begin{aligned}\tilde{\alpha}(x) &= \operatorname*{argmin}_{i} \rho_{\lambda,P}(x, i) \\ &= \operatorname*{argmin}_{i} \{d(x, \tilde{\beta}(i)) + \lambda \sum_{j=1}^{M} C_{j,\kappa(i)} P(Y = j | X = x)\}.\end{aligned}$$

Iterating the three optimality properties provides a descent algorithm based on learning set (a generalized Lloyd algorithm), and pairwise application yields a tree-structured VQ (TSVQ). Unlike the usual Lloyd algorithm, however, the optimal encoder requires the class posterior probabilities $P(Y = j | X = x)$, which must be estimated from $\mathcal{L}$. Unlike the distribution P_X used in the usual Lloyd algorithm, these probabilities cannot be estimated simply by the empirical distribution implied by the training set. The empirical distribution defines these conditional probabilities in an obvious fashion *only for the x contained in the training set*, yet the designed code will have to be well defined for all future x.

4. BVQ and density estimation

In the absense of a known joint distribution, the proposed strategy is to first design the estimator $\hat{P} = \{\hat{P}_{Y|X}(k|x),\ k \in \mathcal{H}; x \in A\}$ based on labeled learning set $\mathcal{L}$. Then design $(\tilde{\alpha}, \tilde{\beta}, \kappa)$ using $\rho_{\lambda,\hat{P}}$ and the implied encoder. The two-step procedure yields a descent algorithm which should produce a good code if the density estimator is good. Bayes' rule implies that the distribution estimation can be accomplished either by estimating pmfs or pdfs. In particular, if we estimate class conditional pdfs $f_k(x) = f_{X|Y}(x|k)$ by $\hat{f}_k(x)$, $k \in A_Y$, then

$$\hat{P}_{Y|X}(y|x) = \frac{\hat{f}_k(x) \hat{P}_Y(k)}{\sum_m \hat{f}_m(x) \hat{P}_Y(m)},$$

where $\hat{P}_Y(k)$ = relative frequency of the class k in $\mathcal{L}$. The density estimate itself need not be sent to the decoder as it is not needed for decoding, only for encoding.

Density estimation is a much studied topic in the statistical literature. See, e.g., Silverman [31] or Scott [29]. See also Devroye et al. [18] for theoretical aspects of vector quantization-based ("partition-based") density estimators. Well known approaches also include kernel estimators, projection pursuit, and tree-structured algorithms such as the CART or "classification and regression trees" algorithm [11]. If a lattice quantizer is used, these can also be viewed as a form of "histogram density estimator," where the density in each Voronoi cell is the value of the histogram divided by the common cell volume.

A variety of algorithms for designing Bayes vector quantization with posterior estimation have been developed and applied to artificial and real-world examples in the previously cited references. Examples include the development of methods for nonparametric estimation of the posterior class probabilities required for the Bayes VQ and studies of both tree-structured and full search Bayes VQs [6]. Specific techniques for estimating posterior probabilities include a tree-structured vector quantizer that is grown based on a relative entropy splitting rule [20] and two estimators [24] that are designed using the BFOS [30] variations on the CART algorithm [11]. Specific examples include the combined compression and classification for segmentation of computerized tomographic images, aerial images, mammograms, and documents [25, 21, 26, 24, 23]. A method of [31] for estimating class conditional probabilities in simple two-dimensional quantization using a fast Fourier transform FFT was applied to BVQ in [27]. This simple Gaussian example will be considered here as it illustrates several properties of the method.

5. Kohonen's example

Consider a random vector $X = (X_0, X_1)$ described by a probability density function (pdf) [16]

$$f_{X_0,X_1}(x_0,x_1) = \frac{1}{2} f^{(0)}_{X_0,X_1}(x_0,x_1) + \frac{1}{2} f^{(1)}_{X_0,X_1}(x_0,x_1),$$

where $f^{(k)}_{X_0,X_1} = \mathcal{N}(0, \sigma_k^{(2)} I)$ is the conditional pdf given that the class label $Y = k$, where $P(Y = 0) = P(Y = 1) = 1/2$ and I is the identity matrix. The Bayes rule for classification is simply the maximum *a posteriori* (MAP) classifier :

$$\kappa_{\mathrm{MAP}}(x_0,x_1) = \max_k{}^{-1} \frac{f^{(k)}_{X_0,X_1}(x_0,x_1)}{f^{(0)}_{X_0,X_1}(x_0,x_1) + f^{(1)}_{X_0,X_1}(x_0,x_1)},$$

which can be explicitly solved for specific distributions and the resulting minimum Bayes risk computed. No VQ based classifier can provide lower Bayes risk. The optimal encoder needs to compute $P(Y = l|X = x) = f_{X|Y}(x|l)p_l/\sum_j f_{X|Y}(x|j)p_j = f_{X|Y}(x|l)/\sum_j f_{X|Y}(x|j)$, since the priors are equal, where $f_{X|Y}(x|l) = e^{-\frac{1}{2\sigma_l^2}||x||^2}/\sqrt{2\pi\sigma_l^2}$. Because equal costs are assumed, $C_{01} = C_{10} = 1$, the optimal encoder, $\alpha^*(x)$, then becomes

$$\alpha^*(x) = \min_i{}^{-1} \Bigg\{ ||x - \beta(i)||^2 + \lambda \left[\frac{1(\{\kappa(i) = 1\}) e^{-\frac{3}{8}||x||^2} + 1(\{\kappa(i) = 0\})\frac{1}{2}}{e^{-\frac{3}{8}||x||^2} + \frac{1}{2}} \right] \Bigg\}$$

The Bayes decision rule is a circle about the origin of radius 1.923, yielding an error probability of 0.264.

The Voronoi diagram for the parametric case where the the pdf's are actually known is shown in figure 1. Here the inner cells coincide exactly with the actual Bayes region. The MSE Voronoi regions resulting from the same codebook is also shown. Here the polygonal regions do not well approximate the circle and the Bayes risk increases.

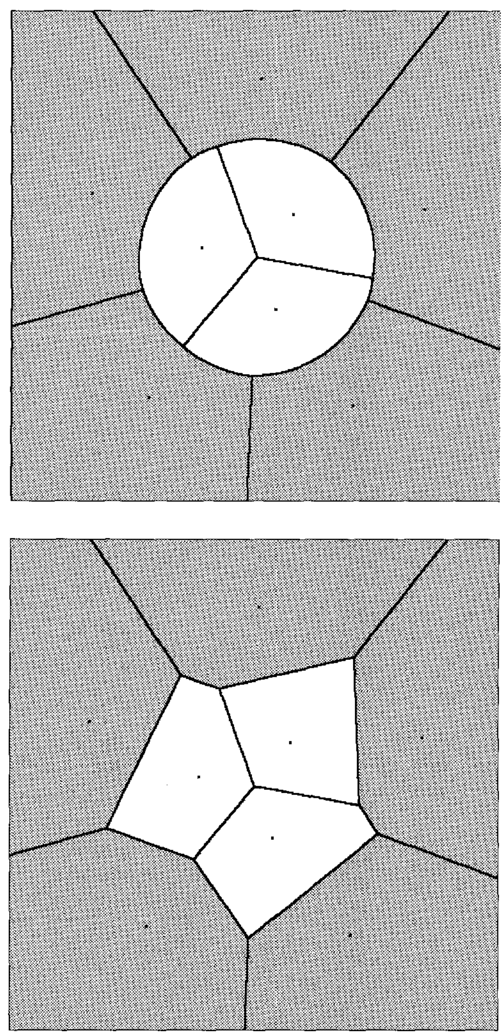


Figure 1. – Parametric BVQ Voronoi diagram (left) and MSE encoded Voronoi diagram

An example of a trained VQ (using an FFT-based density estimator) based on five trials with a training set of 10,000 vectors and a test set of the same time is shown in Figure 2. For comparison the following table shows the average squared error of the parametric and trained BVQ schemes along with several other methods. The parametric BVQ is equivalent for classification to Picinbono and Duvaut's scheme.

We also observed that the BVQs with the inverse halftoning estimator and with the CART-based estimator performed better with lower values of λ. For example, with $\lambda = 10$, the (MSE,P_e) results for the two BVQs were (0.633, 0.266) and (0.647, 0.273), respectively.

In general, we note that the average classifier performance for the density estimating BVQ is very close to that achieved by

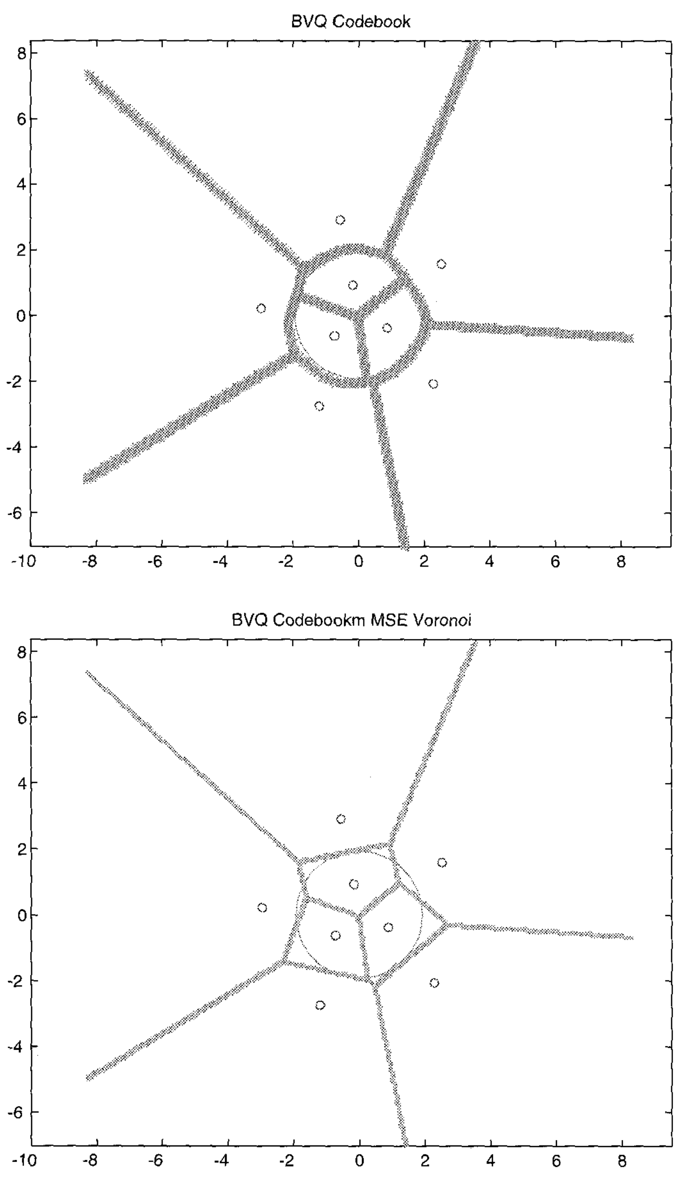


Figure 2. – (left) density estimating BVQ (right) using MSE encoder

the parametric BVQ where the density is known, and both are extremely close to the optimal classifier performance *based on the original, unquantized, inputs*, the performance that would be achieved by Picinbono and Duvaut's quantizer.

Tableau 1. – MSE and P_e for Kohonen's Example.

BVQ : Inverse halftone estimator	0.655	0.269
BVQ : CART-based estimator	0.653	0.274
Kohonen LVQ	0.725	0.279
Parametric BVQ	0.620	0.264
MSE Quantizer/Classifier Cascade	0.598	0.295
BVQ : TSVQ pmf estimator	0.630	0.270

Acknowledgments

The work described here has been a cooperative effort with current and former students and colleagues in the Compression and Classification Group of Stanford University

http ://www-isl.stanford.edu/~gray/compression.html

The material has been particularly influenced by Karen Oehler, Richard Olshen, and Keren Perlmutter.

BIBLIOGRAPHIE

[1] S. A. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Comm.*, vol. COM-25, p. 479–484, May 1977.

[2] H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Comm.*, vol. COM-25, p. 893–900, September 1977.

[3] H. V. Poor and J. B. Thomas, "Optimal quantization for local decision based on independent samples," *J. Franklin Inst.*, vol. 303, p. 549–561, 1977.

[4] H. V. Poor and D. Alexandrou, "A general relationship between two quantizer design criteria," *IEEE Trans. Inform. Theory*, Vol. IT-26, p. 210–212, 1980.

[5] D. Alexandrou and H. V. Poor, "The analysis and design of data quantization schemes for stochastic signal detection systems," *IEEE Trans. Comm.*, vol. COM-28, p. 983–981, September 1980.

[6] J. A. Bucklew, "Multidimensional digitization of data followed by a mapping," *IEEE Trans. Inform. Theory*, vol. 30, p. 107–111, January 1984.

[7] B. Aazhang and H. V. Poor, "An optimum and nearly optimum data quantization for signal detection," *IEEE Trans. Comm.*, vol. COM-32, p. 745–751, September 1984.

[8] B. Picinbono and P. Duvaut, "Quantification et détection," *Ann. Télécommun.*, vol. 41, p. 1246–251, 1986.

[9] B. Picinbono and P. Duvaut, "Optimum quantization for detection," *IEEE Trans. Comm.*, vol. 36, p. 1254–1258, November 1988.

[10] G. R. Benitz and J. A. Bucklew, "Asymptotically optimal quantizers for detection of i.i.d. data," *IEEE Trans. Inform. Theory*, vol. IT-35, p. 316–325, 1989.

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

[12] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inform. Theory*, vol. IT-13, p. 21–27, 1967.

[13] C. J. Stone, "Consistent nonparametric regression," *Annals of Statistics*, vol. 5, p. 595–645, 1977.

[14] T. Kohonen, "An introduction to neural computing," *Neural Networks*, vol. 1, p. 3–16, 1988.

[15] T. Kohonen, *Self-organization and associative memory*, Berlin : Springer-Verlag, third ed., 1989.

[16] T. Kohonen, G. Barna, and R. Chrisley, "Statistical pattern recognition with neural networks : benchmarking studies," in *IEEE International Conference on Neural Networks*, p. 61–68, July 1988.

[17] D. Pollard, "Quantization and the method of k-means," *IEEE Trans. Inform. Theory*, vol. 28, pp. 199–205, Mar. 1982.

[18] L. Devroye, L. Györfi, and G. Lugosi, *A Probabalistic Theory of Pattern Recognition*, Springer, New York, 1996.

[19] E. E. Hilbert, "Cluster compression algorithm : a joint clustering/data compression concept," Publication 77-43, Jet Propulsion Lab, Pasadena, CA, Dec. 1977.

[20] K.O. Perlmutter, R.M. Gray, K.L. Oehler, R.A. Olshen, "Bayes risk weighted tree-structured vector quantization with posterior estimation," *Proceedings Data Compression Conference (DCC)*, IEEE Computer Society Press, March 1994, p. 274–283.

[21] C. L. Nash, K. O. Perlmutter, and R. M. Gray, "Evaluation of bayes risk weighted vector quantization with posterior estimation in the detection of lesions in digitized mammograms," In *Proceedings of the 28th Asilomar Conference on Circuits Systems and computers*, Pacific Grove, CA, October 1994, p. 716–20 vol. 1.

[22] K.L. Oehler and R.M. Gray, "Combining image compression and classification using vector quantization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 17, p. 461–473, May 1995.

[23] K. O. Perlmutter, "Compression and Classification of Images using Vector Quantization and Decision Trees," Ph.D. Thesis, December 1995.

[24] K. O. Perlmutter, R. M. Gray, R. A. Olshen, S. M. Perlmutter, "Bayes Risk Weighted Vector Quantization with CART Estimated Posteriors," *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing* (ICASSP), p. 2435-8 vol. 4, May 1995.

[25] 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 1994 International Conference on Image Processing* (ICIP), volume 2, p. 217–221, Austin, TX, Nov. 1994.

[26] 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," *IEEE Transactions on Image Processing*, vol. 5, no. 2, p. 347–60, February 1996.

[27] R.M. Gray, K.M. Perlmutter, and R.A. Olshen, "Quantization, classification, and density estimation for Kohonen's Gaussian mixture," *Proceedings 1998 IEEE Data Compression Conference,* J.A. Storer and M. Cohn, Eds, IEEE Computer Society Press, Los Alimitos, Calif., p. 63–72, March 1988.

[28] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Boston : Kluwer Academic Publishers, 1992.

[29] D.W. Scott, *ultivariate Density Estimation : Theory, Practice, and Visualization*, Wiley, New York, 1992.

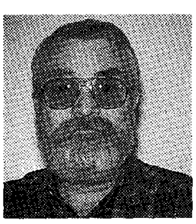
[30] P.A. Chou, T. Lookabaugh, and R.M. Gray, "Optimal pruning with applications to tree-structured source coding and modeling," *IEEE Transactions on Information Theory*, Vol. IT-35, p. 299-315, March 1989.

[31] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability 26, Chapman & Hall, New York, fourth printing, 1994.

Manuscrit reçu le 19 mars 1999.

L' AUTEUR

Robert M. GRAY



Robert M. Gray (B.S, M.S. MIT 1966, Ph.D. U.S.C. 1969) is Professor and Vice Chair of the Department of Electrical Engineering, Stanford University. His current research interests include the theory and design of systems for image compression and segmentation an related areas in coding and signal processing.