PRUNED TREE-STRUCTURED VECTOR QUANTIZATION OF MEDICAL IMAGES WITH SEGMENTATION AND IMPROVED PREDICTION

BY

GIOVANNI POGGI and RICHARD A. OLSHEN

TECHNICAL REPORT NO. 165

DECEMBER 1993

PREPARED UNDER THE AUSPICES

OF

PUBLIC HEALTH SERVICE GRANTS

9 R01 CA59039-18 AND 5 R01 CA55325

DIVISION OF BIOSTATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Pruned Tree-structured Vector Quantization of Medical Images with Segmentation and Improved Prediction

By

Giovanni Poggi and Richard A. Olshen

Technical Report No. 165

December 1993

Prepared Under the Auspices

Of

Public Health Service Grants

9 R01 CA59039-18 and 5 R01 CA55325

Division of Biostatistics

Stanford University

Stanford, California
Pruned Tree-structured Vector Quantization of Medical Images with Segmentation and Improved Prediction

Giovanni Poggi* Richard A. Olshen**

* CORRESPONDING AUTHOR, Dipartimento di Ingegneria Elettronica, Università di Napoli Federico II, via Claudio 21, 80125 Napoli, Italy, FAX: +39 81 7683149, phone: +39 81 7683151, e-mail: poggi@nadis.dis.unina.it.

** Division of Biostatistics, Department of Health Research and Policy, Stanford University, Stanford, CA 94305-5092.

EDICS NUMBER IP 1.1

Abstract

In this work we use predictive pruned tree-structured vector quantization for the compression of medical images. Our goal is to obtain a high compression ratio without impairing the image quality, at least so far as diagnostic purposes are concerned. We use a priori knowledge of the class of images to be encoded to help us segment the images and thereby to reserve bits for diagnostically relevant areas. Moreover, we improve the quality of prediction and encoding in two additional ways: by increasing the memory of the predictor itself and by using ridge regression for prediction. The improved encoding scheme was tested via computer simulations on a set of mediastinal CT scans; results are compared with those obtained using a more conventional scheme proposed recently in the literature. There were remarkable improvements in both the prediction accuracy and the encoding quality, above and beyond what comes from the segmentation. Test images were encoded at 0.5 bit per pixel and less without any visible degradation for the diagnostically relevant region.

This work was supported in part by the National Institutes of Health under Grants CA49697-02 and 5 R01 CA55325 and the National Science Foundation under Grants DMS-9101528 and MIP-9016974
1 Introduction

Digital images are used widely in medical practice. Their use will only increase with the diffusion of powerful and affordable computers that have large storage and computational power, as well as newly found medical indications for them. Several imaging modalities are already genuinely digital, for example computerized tomography (CT) and magnetic resonance (MR); others produce analogue images that are subsequently digitized. The digital format allows for the efficient storage, retrieval, and transmission of images, and therefore the generation and use of large image banks. Digital processing techniques, such as enhancement and automatic classification, can be important diagnostic tools for the physician.

Despite the technical advances that steadily increase the performance of processing, storage, and communication facilities, the huge amount of data produced in medical imaging is likely to remain a bottleneck for many applications. An X-ray image, for example, is usually converted to a 2 Mbyte file; and even less data-intensive modalities like MR, for which 128 kbytes per image are required, can easily overflow the memory/channel capacity when used in a 3-D fashion, as when a sequence of axially displaced slices is generated.

As a consequence, some form of data compression may be imperative for many applications. Often lossless compression is preferred because of legal constraints or simply to avoid any risk, however small, of impairing diagnoses. The compression ratios it provides, however, are quite limited (typically 2:1 or 3:1) and are inadequate for many applications. By now, several independent research groups [1, 2, 3, 4] working with different compression algorithms have shown that lossy compression, at least up to a ratio of 10:1, does not significantly impair the diagnostic value of the images. Moreover, there are applications for which less restrictive criteria can be adopted (storage of data post diagnosis and instructional data bases being two such) and for which data size is a major problem.

In this paper we deal with a particular data compression algorithm, predictive pruned tree-structured vector quantization (P-PTSVQ), which has already been applied successfully [5, 6] to the compression of medical images. In previous work much emphasis was given to the design of tree-structured codebooks for vector quantization, and effective and fast algorithms were
developed to this end. Quite simple prediction schemes were adopted since more sophisticated schemes seemed to provide little improvement in the accuracy of the prediction and, what is more important, even less in the quality of the encoding.

Here, we propose a new prediction scheme that improves in several respects on the one used in [5] and [6]. We show that it leads to a considerable gain in the quality of encoding. To begin, note that typical images are composite sources, made up of many objects having different characteristics. A predictor that is not adaptive must confront very different situations and perform reasonably well in all of them, so there is little flexibility for optimization. On the other hand, attempts to build adaptive predictors that are based on local statistics of the image may be frustrated by unreliable small-sample estimates, not to speak of computational complexity.

To improve prediction, we take a less ambitious but more realistic approach that consists of segmenting the image into homogeneous regions and designing a predictor for each region. This way reasonable sample sizes are available for estimation. We devise predictors with large memory that are reliable and fit well. Our approach is especially suited to medical images where, given the imaging modality and subject matter, a large amount of a priori information is available; and segmentation can be carried out easily. In particular, for now we deal with CT scans of the mediastinum (chest cavity). The segmentation just singles out the body, classifies the remainder as background, and discards it.

In dealing with more homogeneous data and designing predictors with large memory, it is fundamental to have reliable estimates of the covariance matrices that bear upon the prediction coefficients. The problem of estimating vectorial parameters has been a major area of recent research, especially in statistics, but also in numerical analysis. One line of activity began in earnest in 1956 when Charles Stein showed that maximum likelihood is not admissible for quadratic loss in the problem of simultaneously estimating at least three independent Gaussian means (with common variance) [7]. His results have been extended to regression, and thereby to prediction. Separately, inverting badly conditioned positive definite, symmetric matrices has been studied as part of numerical linear algebra. These two areas converge in part in their concluding that the inversion and prediction problems can both be improved by, in effect,
reducing the ratio between large eigenvalues and small ones, all necessarily real, of the matrix whose inverse figures in best linear unbiased estimation. Variations of how to do that are sometimes termed "ridge regression" [8]. We show that a very simple version of ridge regression with validated choice of ridge parameter not only improves prediction in P-PTSVQ but also diminishes coding error.

The rest of the paper is organized as follows. In next Section lossy compression of medical images is discussed, and the segmentation algorithm for chest CT scans is presented. P-PTSVQ is reported in some detail in Section 3, and in Section 4 the new approach to the design of the predictor is introduced and motivated. In Section 5 the simulation experiments are described and results are discussed. Finally, we draw conclusions.

2 Compression of Medical Images

Lossy compression, by definition, leads to the loss of some information; that is, it is irreversible. Only an approximate version of the original data can be obtained from the compressed data. The interest in lossy compression is justified by the large compression ratios that can be achieved, perhaps 10:1 or better, as compared to 2:1 to 3:1 provided in the best cases by lossless coding, at least in medical applications like ours. Frequently, the loss of information is acceptable in applications, for example in video or speech coding, where limited resources are available in terms of memory or bandwidth, and one is interested in reaching the best possible reproduction quality for a given encoding rate. This is less the case with medical images because compression might diminish the diagnostic value of the images and possibly lead to wrong diagnoses. Such an occurrence is unacceptable if it alters treatment, and there is a reluctance in much of the medical community to accept any form of lossy compression.

On the other hand, even in the medical field there are applications where some form of compression ought to be welcomed. Compression can significantly increase the storage capacity of an image data base and allow a hospital to preserve data about the medical history of patients indefinitely. Furthermore, at present, electronic transmission of images over low-rate channels for remote diagnosis or instruction can be accomplished in reasonable time only through some
form of compression.

Another argument in favor of lossy compression is that medical images have to be seen and interpreted by physicians, human users, whose sensitivities are limited. If an expert viewer finds the compressed version of an image indistinguishable from the original, then for him or her it has the same diagnostic value for all practical purposes. In addition, even when an image is perceived to have been degraded from an aesthetic point of view, its diagnostic value is not necessarily affected. In a recent study [3, 4] on PTVQ-compressed medical images it was found that compression rates up to 10:1 did not affect the diagnostic accuracy in a statistically significant way, and that both objective (signal-to-noise ratio) and subjective quality degrade sooner than the diagnostic accuracy.

It is evident that the encoding strategy itself should use priori knowledge on the input image to concentrate resources on those parts that are significant for making diagnoses. Indeed, medical images are highly structured; for a given imaging modality (CT, radiography, MR, etc.) and given subject matter (chest, head, etc.) there are many predictable features in the images that can be taken into account to improve the encoding performance.

An extreme example is given by sequences of CT-slices of a body where displacement estimation techniques can be used to exploit the high redundancy among spatially close slices [9]. However, even when images of different patients are considered, there is a significant amount of mutual information to exploit. In [10] MR images of the head are segmented with a fixed mask extracted from the analyses of similar training images, and more resources are devoted to those regions where the cortex and the cerebellum are expected to be.

In this work we consider a well defined set of images, mediastinal (chest cavity) CT scans, and perform a segmentation guided by the common features of the whole class but dependent on the particular image. This way, each image is divided in diagnostically significant and irrelevant regions, the latter coded to black without any impairment of the subjective quality and particularly of the diagnostic value. An important consequence, besides saving bits for relevant areas, is that both the predictor and the codebook will be better matched to the chest region with heavy influence on the encoding and, eventually, evident improvement of the quality
of reproduction of the chest for the same encoding resources.

For mediastinal CT scans, the task of selecting the regions of the image containing relevant data is very well defined, since all the information of possible diagnostic use is in the chest region, while other objects, like body supports, and the background have no importance. As a result, the segmentation process is straightforward and requires negligible processing time. One discards between 40% and 60% of the original data. The outer circle, about 22% of the whole square image, can be discarded a priori, since it contains no useful information. Of course, it is necessary to transmit information regarding the segmentation and, since we are considering a predictive encoding algorithm, regarding those parts of the image for which no prediction is possible. The cost of this information is extremely limited.

The segmentation process is performed in the three steps described below:

• **THRESHOLDING**

  The luminance histogram of CT scans we study are clearly multimodal, due to the various subsources present in the image: muscles, bones, lung cavity, background, objects in the background. For our purposes, however, it is convenient to think of the image as bimodal and cluster together all the "bright" and all the "dark" modes. These two groups of modes are distinct and a threshold level between them can be found easily. (We use the Lloyd algorithm for clustering [11].) By thresholding the image at such a level the bright boundary of the chest is accurately singled out against the dark background. Further processing is required, however, since the lung cavity is also classified as dark, while external objects to be discarded are bright.

• **SEGMENTATION REFINEMENT**

  To refine the segmentation map, all connected regions (4-pixel connectivity is considered) are extracted and labeled. A dark region is considered to be internal, and hence turned to bright, if it does not include any of the four corners of the image (which are always dark). This is certainly the case with all the chest cavity regions since a thick, bright muscle wall separates them from the outer circle. At this point, several bright regions can be present
in the image besides the chest, due to noise or to the presence of other objects. These regions, which are without exception much smaller than the chest, are easily detected and erased.

- SHAPE SMOOTHING

The segmentation map must be encoded and transmitted in order to synchronize the operations of encoder and decoder. For these kinds of bilevel images, the best encoding strategy is no doubt the chain coding of the object boundary [12]. To prevent any ambiguity in the boundary following process, a preliminary morphological filtering [13] (dilation and erosion with a 3x3 window) was performed. This way, the boundary curve, obtained as the difference between the dilated and the basic shape, is a chain of pixels, each with exactly two 4-connected neighbors in the chain.

Although this segmentation algorithm is clearly tailored to chest CTs, similar algorithms can be devised for other classes of images. The key point is to focus attention on a given class and exploit the knowledge of its invariant features.

3 Predictive PTSVQ

Vector quantization (VQ) [14] is a lossy coding technique that extends the concepts of scalar quantization to multidimensional spaces. Both encoder and decoder are provided with a set of template vectors (codevectors) called codebook \( C \equiv \{ \hat{x}_1, \ldots, \hat{x}_N \} \) which is designed so as to well approximate the distribution of the input vectors. For each input vector \( x \) the encoder selects the codevector \( \hat{x}_i \) that minimizes the assigned distortion measure \( d(\cdot, \cdot) \) over the whole codebook

\[
d(x, \hat{x}_i) \leq d(x, \hat{x}_j) \quad j = 1, \ldots, N
\]

and transmits the corresponding index \( i \). With such an index the decoder addresses its local copy of the codebook to recover the corresponding codevector and perform the approximate reconstruction of the input vector.

For a given vector dimension, VQ provides the best performance among all block coding techniques (which can be seen as special cases of VQ), and the performance improves with
increasing dimension. Unfortunately, the encoding complexity increases as well, exponentially, with the vector dimension forcing one to use small vectors (typically, blocks of no more than 4 \times 4 pixels in image coding) and therefore only partially to exploit the potential advantages of VQ.

It is possible, and often convenient, to trade off performance for reduced complexity by constraining the structure of the codebook, and hence adopting a suboptimal encoding strategy. Indeed, a simpler constrained encoder can work with larger vectors and hence provide an overall better performance despite its lack of optimality. A good compromise between performance and complexity is offered by tree-structured VQ (TSVQ).

As the name suggests, in TSVQ the codebook is constrained to have a tree structure, not necessarily balanced, where the codevectors are associated with their respective terminal nodes (leaves) of the tree. A vector is associated with each internal node as well; this vector is a weighted average of all the codevectors belonging to the subtree originating from that node. The encoding is performed with a sequence of binary decisions (our trees are all binary by design): starting from the root the encoder chooses the child node with the best reproduction vector (the one with the smallest distortion) and moves to that node; the process is then repeated until, eventually, a leaf node is reached and the corresponding codevector selected.

A well designed variable rate TSVQ can significantly outperform fixed rate full-search VQ and have a much reduced computational complexity, the only drawback being an increased memory requirement. In [15] an effective design algorithm, called pruned TSVQ (PTSVQ) was devised by generalizing the BFOS algorithm [16, 14] for the design of binary classification and regression trees. It consists of two steps: growing a large unbalanced tree in a greedy fashion, and optimally pruning it to obtain the subtree that satisfies the design requirements.

- **GROWING**

  Given a tree \( T \), let \( \tilde{T} \) be the set of its leaves. We can define the average distortion \( D \) and rate \( R \) of the tree as

  \[
  D = \sum_{i \in \tilde{T}} p(i) d(i) \tag{2}
  \]
\[ R = \sum_{i \in \mathcal{T}} p(i) l(i) \]

where \( p(i) \) is the probability that node \( i \) be selected; \( l(i) \) is its depth; and \( d(i) \) is the associated average distortion. Since \( p(i) \) and \( d(i) \) are not known in practice, they are usually estimated from a training set.

Whenever a leaf is split, the rate is increased, since longer words are produced; the distortion is decreased, since better codevectors are designed for each subset of training vectors. Let \( t \) be the leaf to be splitted, \( t_L \) and \( t_R \) its left and right children, and \( p_L \) and \( p_R \) the proportion of training vectors in \( t \) that go to the left and right children respectively. Then the variation in average distortion and rate is

\[
\begin{align*}
\Delta D &= p(t)[p_L d(t_L) + p_R d(t_R) - d(t)] \quad (4) \\
\Delta R &= p(t)[p_L l(t_L) + p_R l(t_R) - l(t)] = p(t) \quad , (5)
\end{align*}
\]

and we can define the ratio of the decrease in distortion to the increase in rate as

\[ \lambda = \frac{\Delta D}{\Delta R} = p_L d(t_L) + p_R d(t_R) - d(t) \]

Choosing the split that gives rise to the largest \( \lambda \), as originally proposed in [5], we optimize the growth of the tree. This optimization, however, is greedy, since it does not take into account the future consequences of a particular choice. A split which is sub-optimal in one step, and hence discarded, can be the best in the long run. On the other hand, a lookahead strategy would entail an exponential growth in complexity and not necessarily provide significant improvements, as was noted in [17] where two-step lookahead growing was considered.

*PRUNING*

To obtain some advantages of lookahead without the computational overhead one can grow larger trees than desired and then prune them back to the desired size (or rate, or distortion). The pruning is done by performing, at each step, the cut that provides the smallest \( \lambda \), namely, the smallest increase in distortion per decrease in rate [5]. This
approach has proven to be extremely powerful since entire branches of the tree, and not only single nodes, can be pruned at once, and therefore truly optimal subtrees can be reached. Moreover, since the optimal subtrees are nested, as was proved in [16], the computational complexity of the pruning is limited, making the algorithm in its entirety computationally practical.

As all block coding techniques, PTVS suffers from the inherent limitation of exploiting the dependence among just the small number of samples of a block. Sources like images or speech exhibit a high degree of memory and hence a considerable inter-block correlation which ought to be taken into account to improve the performance. This can be accomplished effectively by performing a prediction of the block to be encoded based on transmitted blocks, and applying PTVS to the residuals. Since the energy of the residuals is usually much smaller than that of the original samples, a codebook of smaller rate can then be used without affecting the reconstruction quality.

The problem arises of how to design the predictor so that the energy of the residuals is small. To limit complexity the constraint of linear (or affine) prediction is usually imposed (although nonlinear prediction is sometimes adopted [18]) and, small prediction configurations are considered. Then, the coefficients are evaluated via the Wiener-Hopf equations. Usually, the covariances necessary for the design of the predictor are (Gaussian) maximum-likelihood, estimated on some finite sample size. We will show, however, that better results can be obtained when a suitable bias term is introduced.

4 Prediction

Our discussion of prediction begins somewhat abstractly, though we will come shortly to a precise description of our implementation. In what follows, readers should think of \( \mathbf{x}_1 \) as a vector of pixel intensities being predicted and \( \mathbf{x}_2 \) as a vector of pixel intensities that are used to predict \( \mathbf{x}_1 \).

Suppose that the \((p_1+p_2)\)-dimensional random vector \( \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \), where \( \mathbf{x}_1 \) is \( p_1 \)-dimensional and \( \mathbf{x}_2 \) is \( p_2 \)-dimensional, has a Gaussian distribution. Conformably partition the mean \( \mu = \)
and covariance $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Then the conditional distribution of $x_1$ given $x_2$ is also Gaussian, with mean

$$E\{x_1|x_2\} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

and covariance

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21};$$

that $\Sigma_{22}^{-1}$ must be defined carefully when $\Sigma_{22}$ is singular does not concern us here. Even when $x$ is not Gaussian, (7) is the wide sense conditional expectation of $x_1$ given $x_2$, and is sometimes called the best affine predictor of $x_1$ given $x_2$. The estimate (7) is widely used in practice, though typically $\mu$ and $\Sigma$ are estimated from data. Thus, if $x^{(1)}, x^{(2)}, \ldots, x^{(N)}$ are independent and each is distributed as $x$, then the maximum likelihood estimates of $\mu$ and $\Sigma$ are, respectively,

$$\hat{\mu} = \frac{1}{N} \sum_{j=1}^{N} x^{(j)}$$

and

$$\hat{\Sigma} = \frac{1}{N} \sum_{j=1}^{N} (x^{(j)} - \hat{\mu})(x^{(j)} - \hat{\mu})^T,$$

where superscripted $T$ denotes transpose.

From statistical decision-theoretic points of view neither (9) nor (10) is a very good estimate; by most formulations $\hat{\mu}$ is "inadmissible", (meaning that another estimator exists that has no larger risk for any $\mu$, and smaller risk for some $\mu$) when $p_1 + p_2 \geq 3$, and $\hat{\Sigma}$ is inadmissible even when $p_1 + p_2 = 1$ (see, for example, [7, 8] and [19, 20, 21, 22] and their cited references). Note that the inadmissibility of $\hat{\Sigma}$ is not because it is slightly biased; in fact the admissible estimates are all biased, and are obtained by shrinking the unbiased estimate towards some meaningful value, usually zero. Since (9) and (10) admit improvement, especially in high dimensions, it is plausible that estimates of (7) based upon them also can be improved.

Statisticians have thus tried to improve upon the estimate of the matrix $\Sigma_{22}^{-1}$ that appears in (7). A simple way to obtain this is to add a multiple of the $p_2 \times p_2$ identity matrix $I$ to $\hat{\Sigma}_{22}$ before inverting it. This technique is called ridge regression, and is the subject matter of
[8, 21, 22] and many other papers. Because (10) is a symmetric, nonnegative definite matrix, its eigenvalues are real and nonnegative, and are the reciprocals of the eigenvalues of its inverse. Adding a scaled identity matrix $\alpha I$, all eigenvalues are increased by the quantity $\alpha$, and therefore the eigenvalues of the inverse are shrunk towards the origin. It turns out that with a proper choice of $\alpha$ the estimate of $\hat{\Sigma}_{22}$ is significantly improved. For each positive number $\alpha$, we write

$$
\hat{E}\{x_1|x_2\} = \bar{\mu}_i + \hat{\Sigma}_{12}(\hat{\Sigma}_{22} + \alpha I)^{-1}(x_2 - \bar{\mu}_2).
$$

In our application $\mathbf{x}$ is surely not Gaussian. We have no theoretical grounds on which to model its distribution; and, anyway, our purpose is practical, not theoretical. Thus we estimate $\alpha$ empirically by what in [16] is called test sample estimation.

Besides improving the estimate of the inverse matrix, ridge regression decreases the ratio of largest to smallest eigenvalues, making the matrix better conditioned. The use of ridge regression, eliminating the risk of inverting a badly conditioned matrix, encourages us to increase the memory of the predictor, with a relevant effect on performance. Therefore, we predict a block of pixels from all the neighboring blocks which have already been transmitted.

Given this prediction geometry, we used the empirical distribution of $4 \times 8$ pixel blocks pooled from four CT images of the mediastinum to estimate $\mu$ and $\Sigma$. Then $\alpha$ was chosen experimentally as the value for which the mean squared error for prediction was minimized. Readers will see that suitably chosen $\alpha$'s offer substantial improvement in the ultimate Peak Signal-to-noise Ratio (PSNR) of the coding.

But the algorithm thus far might be criticized for at least two reasons. First, $\hat{\mu}$, $\hat{\Sigma}$ and $\alpha$ are chosen in an "open loop" manner. That is, our estimates devolve from "true" pixel blocks predicting "true" pixel blocks. In practice, however, we apply the predictor in the "closed loop" scenario, where predictions are based on previously encoded samples. Since we deal only with high quality encoding, original and encoded images are quite similar, and we believe that the "closed loop" approach would entail limited improvements, not worth the much increased complexity.

Perhaps the more telling criticism is that our ultimate goal is not really prediction at all, but accurately and usefully encoded images. Prediction that is good but not perfect need not
yield small encoding errors, or, what is the same, high PSNR. Since we have only plausibility arguments for our improvements, we turn now to actual experiments to strengthen our case.

5 Experiments and Results

In our experiments we used a training set of four mediastinal CT scans (Fig.1) to design both the PTVQ codebook and the predictor, and evaluated performance on a different mediastinal CT scan. Each image is a squared array of $512 \times 512$ values with 11 bits per pixel (bpp); the actual image, however, occupies a circle inscribed in the square, and the undefined values are set to zero.

Rather small blocks were considered, $2 \times 4$ pixels, in order to contain the complexity and guarantee good reproduction; and the prediction configuration (Fig.2a) was chosen to be the set of already transmitted pixels belonging to a two-pixel wide strip surrounding the given block (for a total of 20 pixels).

All the images were segmented, as is described in Section 2; and in all cases the chest was accurately selected (see Fig.3) resulting in about 50% of the whole image classified as background. The encoding of the boundary called for an average of 0.01 bpp, which is negligible for the rates in which we are interested.

The predictor was designed following the approach of Section 4. In particular, only the chest region of each image was considered to estimate the mean vectors and covariance matrices in (11), and the weight $\alpha$ was chosen, on the basis of preliminary experiments, as the one corresponding to the minimum mean-square prediction error. This predictor was then applied to the training images in order to obtain the blocks of residuals to be used as a training set for the generation of the codebook. A large tree-structured codebook was grown and subsequently pruned back in order to obtain a sequence of nested optimal subcodebooks, from rate 1.5 to rate 0. Eventually, these codebooks were used, with the selected predictor, to perform P-PTSVQ on the selected test image.

The results obtained with an out-of-training test image (shown in Fig.4a) are reported in Fig.5a, for the prediction alone, and in Fig.5b, for the whole encoding process. For comparison
purposes, we also report the results obtained with a simple reference scheme, similar to the one adopted in [6], with a small predictor (Fig.2b) optimized over the whole image, no segmentation and no ridge regression. All error measures, of course, are computed only on the part of the image, the chest, that is actually encoded.

The improvement obtained is striking: the mean-square prediction error (MSPE) is at least halved, and also the peak signal-to-noise ratio (PSNR) defined as

\[ \text{PSNR} \triangleq 10 \log_{10} \frac{z_{\text{MAX}}^2}{\text{MSEE}} \]  \hspace{1cm} (12)

where MSEE stands for mean-square encoding error, increases of about 6 dB over the whole range of rates.

The PSNR is an objective measure of the quality of encoding (what we are interested in, eventually) however, as we stressed the importance of the subjective judgement of the encoding quality for medical images, we also compare in Fig.4 the original and encoded version (at 0.5 bpp) of the test image. Apart from the obvious differences due to segmentation, like the erasure of the body supports, the images appear almost indistinguishable, which is reasonable with a PSNR as high as 43 dB.

The comparison proposed in Fig.5 might be sensed as unfair, since in the reference encoding scheme a conspicuous amount of data is spent for the background, completely neglected with the new approach. Let us consider, then, a smarter reference scheme, where predictor and PTVSQ codebook are the same as before but nothing is transmitted for the background. The new comparison is shown in Fig.6b, and even in this “equalized” condition the superiority of the new scheme is obvious, although the improvement in the PSNR is now reduced to about 3 dB. In order to help explain such a result, two more graphics are given, corresponding to two intermediate schemes so as to obtain a path from the reference to the new scheme by first using a larger prediction configuration, then adding segmentation, and lastly making use of the ridge regression. It appears that each step contributes individually to improve the encoding quality although the relative importance of the contributions is different.

As an example, increasing the memory of the predictor from 4 to 20 pixels is by itself a very helpful step. Note that this result is partially hidden if the whole image is considered rather
than just the chest since on a flat background all predictors work reasonably well and differences in performance are reduced on the average. This is probably the reason why such a step was not taken in previous research.

The use of segmentation in the design phase leads to further improvements, especially for the lower bit-rates. This is not surprising as in this case both predictor and codebook are optimized for the kind of data on which they operate. Suppose, for example, that in the codebook designed on full images the codeword with constant null value has such a high probability, due to a perfectly flat background, that it is assigned a binary index of length one (Fig.7a), while it is never used for the chest. Working on segmented images it can be dropped, obtaining a more efficient codebook to be used on the chest (Fig.7b). Of course the real situation will be different, and more complex, but this kind of behavior is to be expected.

Finally, note that the use ridge regression, leading to the design of a more reliable predictor, allows one to squeeze out a half dB more, on top of the improvements already obtained.

It is natural to wonder whether the order with which the single steps are considered is important or not: as an example, the ridge regression could help when using the larger prediction configuration but hurt with the smaller one. To gain some insight about this we conducted experiments with all the prediction schemes arising from the combination of the three choices, segmentation vs whole image, 4-point vs 20-point configuration, maximum likelihood vs ridge regression. In Fig.8 we show the PSNR obtained in all these situation at the 0.5 bpp rate. The figures confirm that each of the modification introduced consistently improve the performance; for example, moving from the upper to the lower facet of the cube (namely, using ridge regression) the PSNR is always increased, up to a maximum of 1.2 dB, and similar considerations apply in all cases.

6 Conclusions

We have used Predictive-PTSVQ for the lossy compression of medical images, introducing a segmentation step to avoid coding pixel blocks that are pure background or are otherwise irrelevant, and several refinements to the prediction schemes used in previous work. These include
increasing the number of pixels used by the predictor, and improving the prediction scheme by employing ridge regression adapted to our specific problem. We establish that, besides the obvious advantages provided by segmentation, the new predictor reduces the mean-square prediction error for a given bit-rate, at least for the class of images that we considered. What is more important, significant improvements of the encoding quality are obtained, with an increase of the PSNR of 3 dB or more, above and beyond the improvement from segmentation. As a result, we obtained images encoded at 0.5 bpp which are virtually indistinguishable from the original, except for the erasure of irrelevant background objects.

Besides its value for the specific application, this result alerts users to the fact that algorithms for PTVQ published thus far, though probably valuable in applications [3, 4, 5, 6, 9, 10], can be improved considerably. Much can be gained by working out more sophisticated prediction schemes. We feel that Predictive-PTSVQ will be a convenient alternative to the more popular DCT-based schemes, like JPEG, when all its building blocks are suitably designed, especially, though not exclusively, the prediction itself. For example, region-based coding has shown to have significant potential [10], and clever choice of block geometry helps avoiding other problems [23]. At least two topics deserve further investigation: the design of a smarter segmentation scheme, based for example on the estimated activity and edge orientation of each block, and the analysis of other prediction schemes in place of the simple ridge regression considered here.

Acknowledgments

We thank Robert Gray for the stimulating environment he provides for investigations like ours, and Pamela Cosman for helping with the exposition of PTVQ.
References


Captions to figures

Fig.1 Images forming the training set. All images are CT scans of the mediastinum, 512x512 pixels, 11 bit/pixel (8 most significant bits displayed).

Fig.2 Prediction configurations adopted in the new scheme (a) and in the reference scheme (b). The “X” indicates the pixel to be predicted, and the “p’s” indicate the predictor pixels.

Fig.3 Segmentation of the images of the training set. The chest region (white) is accurately selected in all cases.

Fig.4 The test image: original (a) and encoded at 0.49 bit/pixel (b). The Peak Signal-to-noise Ratio (PSNR) in the chest region is 42.92 dB.

Fig.5 Mean-square Prediction Error (MSPE) versus bit-rate (a) and PSNR versus bit-rate (b) for the new and the reference scheme.

Fig.6 MSPE versus bit-rate (a) and PSNR versus bit-rate (b) for the new scheme, the reference scheme, and two intermediate schemes. In all cases the bit-rate takes into account only the bits spent to encode the segmented region.

Fig.7 Two examples of tree-structured codebooks. The code (b) is better matched to the chest region than the code (a) and provides a lower bit-rate.

Fig.8 Effects of the improved prediction scheme on the encoding PSNR. All paths going from the reference scheme (max likelihood & whole image & 4-point prediction) to the new scheme (ridge regression & segmentation & 20-point prediction) exhibit a monotonic increase of the PSNR. The overall improvement exceeds 3dB.
Fig. 1
Fig. 2
NULL VECTOR, USED ONLY FOR THE BACKGROUND

Fig. 7
Fig. 8