Practical Multi-Resolution Source Coding: TSVQ Revisited *

Michelle Effros

California Institute of Technology

Abstract – Consider a multi-resolution source code for describing a stationary source at L resolutions. The description at the first resolution is given at rate R_1 and achieves an expected distortion no greater than D_1. The description at the second resolution includes both the first description and a refining description of rate R_2 and achieves expected distortion no greater than D_2, and so on. Recently derived multi-resolution source coding bounds describe the family of achievable rate and distortion vectors $((R_1, R_2, \ldots, R_L), (D_1, D_2, \ldots, D_L))$. By examining these multi-resolution rate-distortion bounds, we gain insight into the problem of practical multi-resolution source coding. These insights lead to a new multi-resolution source code based on the tree-structured vector quantizer. This paper covers the algorithm, its optimal design, and preliminary experimental results.

I Introduction

With advances in communications media and technologies come needs for communications techniques that take full advantage of the capabilities particular to those technologies. A prime example is the medium of internet communications. With the growth of internet communications comes an increased need for techniques whereby a single user can simultaneously communicate the same information to a wide array of other users with vastly varying bandwidth resources, computational capabilities, and performance requirements. This application, among others, has inspired a surge of interest in multi-resolution or progressive transmission source coding.

Multi-resolution source codes are data compression algorithms in which simple, low-rate source descriptions are embedded in more complex, high-rate descriptions. Use of multi-rate source codes allows users with severe bandwidth constraints or low performance requirements to achieve a low quality data representation by only incorporating a fraction of the original coded bit stream. Users with greater capabilities or needs can achieve more precise data representations by using larger fractions of the same bit stream. Further, users uncertain of their precision needs can progressively

reconstruct the data to higher and higher accuracy – stopping the communication process when the desired accuracy is achieved. The world wide web, with its vast array of users with widely varying computational capabilities, bandwidth, and functional requirements, represents just one example of an application where progressive source codes can have an enormous impact.

Interest in multi-resolution or progressive transmission source coding has inspired an enormous amount of research into practical multi-resolution source coding algorithms (e.g., [1, 2, 3, 4]). Theoretical questions regarding the optimal performance achievable in multi-resolution source coding have, until recently, attracted considerably less attention. Hence there exists a gap in understanding between practical multi-resolution source codes and the theoretical bounds that govern their performance. The aim of this paper is to move towards closing that gap.

A description of multi-resolution source coding theory, including a brief historical review and description of the multi-resolution source coding theorems for fixed- and variable-rate multi-resolution source codes, is contained in Section II. Section III contains a description of a simple, practical multi-resolution source code – the tree-structured vector quantizer (TSVQ). In section IV, we introduce a new variation of the TSVQ algorithm which employs the insights of Section II in a new variation of the practical TSVQ multi-resolution source coding algorithm of Section III. Section V contains experimental results.

II   Multi-Resolution Source Coding Theory

In single-resolution source coding, a source is broken into contiguous vectors of length N symbols, and each source vector is mapped to a binary description with some average per-symbol description length, say R. The source coding theorem and its converse describe the optimal distortion $D(R)$ theoretically achievable at the given average rate R. Source coding theorems and their converses for stationary ergodic and stationary nonergodic sources on complete separable metric spaces, known as Polish alphabets, appear in [5, Theorems 7.2.4, 7.2.5] and [6] respectively.

According to the source coding theorem, given an average rate of R bits per symbol, the distortion $D(R)$ is achievable in the limit of infinite vector dimension (and complexity). Imagine now stripping off a fraction of the above R bits per symbol to leave some lower average rate of r bits per symbol. While traditional (single-resolution) rate-distortion theory *bounds* by $D(r)$ the optimal distortion theoretically achievable with these r bits per symbol, the single-resolution source coding theorem does not prove the achievability of this bound subject to our constraint on the higher resolution code. Further, given a rate r source code achieving, to arbitrary accuracy, the distortion-rate bound $D(r)$, the source coding theorem bounds, but does not prove achievable, the number of bits which, when appended to our rate r description, would yield a final distortion $D(R)$.

In [7], Gray and Wyner bound the optimal rate-distortion performance of a three-channel communication system with one transmitter and two receivers. Of the three channels, channel one delivers equal rate to both receivers, channel two delivers rate

only to the first receiver, and channel three delivers rate only to the second receiver. The first channel is used for shared information that describes a vector (X, Y) to both receivers. The second channel is used for an incremental description of X to the first receiver. The third channel is used for an incremental description of Y to the second receiver. Under these assumptions Gray and Wyner derive a variety of rate-distortion bounds, treating memoryless Gaussian and binary sources as special cases and deriving conditions for achievement of the "Pangloss bound," where all descriptions simultaneously meet the rate distortion bound associated with their total rates. When $X = Y$ and the rate of either channel two or channel three is zero, this problem becomes the multi-resolution source coding problem and the Pangloss bound becomes the successive refinement condition later described by Koshelev [8, 9], where a source is said to be successively refinable if a code giving a nested two-stage description of that source can always achieve the rate-distortion bound at both rates.

In [10], Equitz and Cover prove that successive refinement is achievable for independent and identically distributed sources if and only if X, $\hat{X}_1$, and $\hat{X}_2$ form a Markov chain, where we here use $\hat{X}_1$ and $\hat{X}_2$ to represent the high- and low-resolution reproductions of X respectively. They then go on to give three examples showing rate-distortion problems that are successively refinable as well as a counterexample to show that not all independent identically distributed sources are successively refinable. In [11], Rimoldi generalizes Equitz and Cover's result by finding the achievable rate pairs for a given pair of distortions and then generalizing to L-resolution codes for $L > 2$. In [12], Effros derives equations describing the achievable rate-distortion region for fixed- and variable-rate L-resolution source codes on stationary ergodic and stationary nonergodic sources with Polish alphabets. The approach taken in deriving the optimal performance bounds of [12] generalizes the variable rate and distortion Lagrangian approach made popular by [13] to the multi-resolution problem.

As in the case of single-resolution source coding, the rate-distortion regions achievable by both fixed- and variable-rate multi-resolution source codes is a closed convex set, describable by its convex hull. The convex hull, in turn, is described by a weighted sum of the rate and distortion terms associated with each of the L resolutions.

More formally, let $(A^\infty, \mathcal{A}^\infty, \mu, T)$ be a stationary dynamical system with Polish alphabet A. That is, let A be a complete separable metric space, let $\mathcal{A}$ be the Borel σ-algebra generated by the open sets of A, let A^∞ be the set of one-sided sequences $x = (x_1, x_2, \ldots)$ from A, let $\mathcal{A}^\infty$ be the σ-algebra of subsets of A^∞ generated by finite-dimensional rectangles with components in $\mathcal{A}$, let T be the left shift operator on A^∞, and let μ be a measure on the measurable space $(A^\infty, \mathcal{A}^\infty)$, stationary with respect to T. We abbreviate the description of the source to $[A, \mu]$.

Now for any fixed integer $L \geq 1$ and each $\ell \in \{1, \ldots, L\}$, where L is the number of levels of resolution in our desired multi-resolution source code, let $\rho_{(\ell)}(x_1, y_1) < \infty$ be a real-valued nonnegative distortion measure for $x_1 \in A$, $y_1 \in \hat{A}_{(\ell)}$, where $\hat{A}_{(\ell)}$ is the ℓth of L abstract reproduction alphabets. Assume that $\rho_{(\ell)}(x_1, y_1)$ is continuous in x_1 for each $y_1 \in \hat{A}_{(\ell)}$ and that there exists a reference letter $y_1^{*(\ell)} \in \hat{A}_{(\ell)}$ such that $E_\mu \rho_{(\ell)}(x_1, y_1^{*(\ell)}) = d^{*(\ell)} < \infty$. Define $\rho_{(\ell)}(x^N, y^N) = \sum_{i=1}^N \rho_{(\ell)}(x_i, y_i)$ Typically $\hat{A}_{(\ell)} = \hat{A}$ and $\rho_{(\ell)}(x_1, y_1) = \rho(x_1, y_1)$ for all $\ell \in \{1, \ldots, L\}$ and all $(x_1, y_1) \in A \times \hat{A}_{(\ell)}$.

Given this observation and our desire for notational simplicity, we drop the (ℓ) from $\hat{A}$, y_1^*, d^*, and ρ, but note that the results apply to the most general case.

The achievable rate-distortion region is the set of rate-distortion points that can be approximated with arbitrary accuracy by a code of arbitrarily high dimension.

Let $\mathcal{Q}^{\mathrm{fr}}(L, N)$ and $\mathcal{Q}^{\mathrm{vr}}(L, N)$ denote the class of fixed-rate dimension-N L resolution codes and the class of variable-rate dimension-N L resolution codes respectively. ($\mathcal{Q}^{\mathrm{fr}}(L, N)$ and $\mathcal{Q}^{\mathrm{vr}}(L, N)$ are described in greater detail in [12].) Then for any $Q^{L,N} = (Q_{(1)}^N, \ldots, Q_{(L)}^N) \in \mathcal{Q}^{\mathrm{fr}}(L, N) \cup \mathcal{Q}^{\mathrm{vr}}(L, N)$, we denote the expected rate and distortion of $Q^{L,N}$ with respect to stationary source μ by

$$
\begin{aligned}
\mathbf{R}(\mu, Q^{L,N}) &= (E_\mu|Q_{(1)}^N(X^N)|, \ldots, E_\mu|Q_{(L)}^N(X^N)|), \\
\mathbf{D}(\mu, Q^{L,N}) &= (E_\mu\rho(X^N, Q_{(1)}^N(X^N)), \ldots, E_\mu\rho(X^N, Q_{(L)}^N(X^N))).
\end{aligned}
$$

For any integer $N > 0$ and any μ, the *L-resolution fixed-rate achievable rate-distortion region* (for a given rate constraint R^L) is

$$
\mathcal{R}^{\mathrm{fr},L}(\mu) = \overline{\bigcup_N \mathcal{R}_N^{\mathrm{fr},L}(\mu)}.
$$

Here $\mathcal{R}_N^{\mathrm{fr},L}(\mu)$ is the associated *Nth-order L-resolution fixed-rate achievable rate-distortion region* defined as

$$
\mathcal{R}_N^{\mathrm{fr},L}(\mu) = \overline{\left\{ (r^L, d^L) : \exists Q^{L,N} \in \mathcal{Q}^{\mathrm{fr}}(L, N) \text{ s.t. } \frac{1}{N}\left(\mathbf{R}_L(\mu, Q^{L,N}), \mathbf{D}_L(\mu, Q^{L,N})\right) \leq (r^L, d^L) \right\}},
$$

where $\overline{S}$ denotes the closure of S with respect to the Euclidean norm and $x^k \geq y^k$ if and only if $x_i \geq y_i$ for all $1 \leq i \leq k$. The *L-resolution variable-rate achievable rate-distortion region* is similarly defined over the class of variable-rate multi-resolution codes. Thus the Nth-order rate-distortion region describes the collection of rate-distortion pairs (r^L, d^L) such that there exists an N-dimensional code with per symbol rate and distortion less than or arbitrarily close (in Euclidean distance) to (r^L, d^L).

Since $\mathcal{R}^{\mathrm{fr},L}(\mu)$ and $\mathcal{R}^{\mathrm{vr},L}(\mu)$ are convex and closed [12], they are entirely characterized by their support functionals [14, p.135] $j^{\mathrm{fr}}(\alpha^L, \beta^L, \mu)$ and $j^{\mathrm{vr}}(\alpha^L, \beta^L, \mu)$, called the *weighted fixed- and variable-rate operational rate-distortion functions*, where

$$
j^{(\mathrm{fr}|\mathrm{vr})}(\alpha^L, \beta^L, \mu) = \inf_{(r^L, d^L) \in \mathcal{R}^{(\mathrm{fr}|\mathrm{vr}),L}} \sum_{\ell=1}^{L} (\alpha_\ell d_\ell + \beta_\ell r_\ell).
$$

Without loss of generality [12] we restrict our attention to $\alpha^L, \beta^L \geq \mathbf{0}$. Thus (r^L, d^L) is *achievable by a fixed-rate code* if $\sum_{\ell=1}^{L}(\alpha_\ell d_\ell + \beta_\ell r_\ell) \geq j^{\mathrm{fr}}(\alpha^L, \beta^L, \mu)$ and *achievable by a variable-rate code* if $\sum_{\ell=1}^{L}(\alpha_\ell d_\ell + \beta_\ell r_\ell) \geq j^{\mathrm{vr}}(\alpha^L, \beta^L, \mu)$ for all $\alpha^L, \beta^L \geq \mathbf{0}$.

The main results from [12] for stationary ergodic sources are as follows. If μ is stationary and ergodic, A is Polish, $\rho(x, y)$ is continuous in x for each y, and there exists a reference letter y^* such that $E_\mu\rho(X, y^*) < \infty$, then the following results hold.

Theorem 1 $j^{\mathrm{fr}}(\alpha^L, \beta^L, \mu) = J(\alpha^L, \beta^L, \mu)$.

Theorem 2 $j^{\mathrm{vr}}(\alpha^L, \beta^L, \mu) = J(\alpha^L, \beta^L, \mu)$.

Here $J(\alpha^L, \beta^L, \mu)$ called the *weighted rate-distortion function*, is defined as

$$J(\alpha^L, \beta^L, \mu) = \inf_N J_N(\alpha^L, \beta^L, \mu),$$

where the *Nth-order weighted rate-distortion function* $J_N(\alpha^L, \beta^L, \mu)$ equals

$$J_N(\alpha^L, \beta^L, \mu) = \inf_q \frac{1}{N} \sum_{\ell=1}^{L} (\alpha_\ell E_{\mu^N q^N} \rho(X^N, \mathbf{Y}_\ell) + \beta_\ell I_{\mu q}(X^N; \mathbf{Y}_\ell | \mathbf{Y}^{\ell-1})),$$

with $E_{\mu^N q^N} \rho(X^N, \mathbf{Y}_\ell)$ and $I(X^N; \mathbf{Y}_\ell | \mathbf{Y}^{\ell-1})$ denoting the expected distortion and conditional mutual information, respectively, between the source vector X^N and its resolution-ℓ N-dimensional reproduction $\mathbf{Y}_\ell$.

The proof uses random coding arguments reminiscent of those used to prove the single-resolution source coding theorem and therefore does not immediately yield a simple algorithm for optimal multi-resolution source code design. Further, achievement of the above rate-distortion bounds generally requires infinite dimensional source codes and thus infinite complexity and delay. Thus multi-resolution source coding theory does not explicitly provide the practical algorithms needed for applications like internet communications. We next consider a simple example from the literature of a multi-resolution code used for practical applications.

III Tree-Structured Vector Quantization

The TSVQ algorithm [15, 16, 1] is an example of a simple multi-resolution source code. In TSVQ, the unstructured codebook and exhaustive search encoder of the basic VQ algorithm are replaced with a tree-structured codebook and tree-search encoder. The TSVQ encoder and decoder both contain copies of a (typically binary) tree with a single vector codeword at each node. The encoder's search strategy is greedy. To encode a given source vector, the encoder starts at the root of the tree and maps the source vector to the child of that node that reproduces the vector to the lowest distortion. The encoder then moves to that node and repeats the process – again mapping the source vector to the closest child of the current node. The process repeats until a leaf is reached. The encoder describes the traversed path using either the path's natural binary description or an entropy coded version of that description.

The decoder maps path descriptions back to reproduction vectors. Since a codeword is located at each tree node, partial source descriptions allow for intermediate reproductions in the process of reconstructing the final higher resolution reproduction.

Most TSVQ design algorithms (e.g., [15, 16, 1]) use either top-down or bottom-up greedy designs. Top-down approaches optimize and fix the low-resolution descriptions then optimize each subsequent higher resolution description. Bottom-up greedy approaches optimize and fix the low-resolution descriptions first and then do the same for higher and higher resolutions.

For variable-rate codes, which often use trees with leaves of varying depths, the inclusion of pruning in the tree design results in a code which does not strictly fall into either the bottom-up or top-down design paradigms. The goal of the tree pruning algorithm is to choose the optimal sequence of subtrees of a given tree to give the best possible rate-distortion performance over all of the resolutions. Nonetheless, as the initial tree to which the tree pruning is applied is typically designed with either a bottom-up or top-down approach, the overall design process is again greedy in nature.

IV   Algorithm

As discussed in Section II, multi-resolution source coding theory provides proofs regarding the existence of multi-resolution source codes achieving, to arbitrary accuracy, the optimal bounds, but does not describe practical multi-resolution source coding algorithms. In contrast, algorithms like TSVQ as described in Section III provide explicit practical codes but are difficult to pin down in terms of "optimal" performance.

By extending our understanding of multi-resolution source coding algorithms, multi-resolution source coding theory offers a variety of lessons useful to the design of practical progressive transmission codes. First, $j^{\mathrm{fr}}(\alpha^L, \beta^L, \mu) = J(\alpha^L, \beta^L, \mu) = j^{\mathrm{vr}}(\alpha^L, \beta^L, \mu)$ implies that on stationary ergodic sources, the achievable rate-distortion region for fixed-rate multi-resolution source codes is identical to that for variable-rate multi-resolution source codes. While this fact does not imply that a given L-dimensional rate-distortion vector (R^L, D^L) will be achievable at the same coding dimension or complexity for both families of codes, but only that fixed-rate codes can potentially offer the same rate-distortion performance as variable-rate codes given sufficient computational and memory resources. As fast algorithms like hierarchical VQ [17] make high dimensional codes increasingly practical, the knowledge that fixed-rate codes are not fundamentally inferior to variable-rate codes, when combined with the properties that make fixed-rate source coding attractive (e.g., greater resilience to channel errors and removal of the buffering problem), makes fixed-rate multi-resolution source coding a desirable option for a variety of applications.

Second, the convexity of the achievable rate-distortion region and the description of its convex hull provide a new performance measure for evaluating multi-resolution source codes. In particular, to design source codes with rate-distortion performance as close as possible to the convex hull of the achievable rate-distortion region, a code $Q^{L,N}$ should be designed to explicitly minimize the weighted performance measure

$$\alpha^L \mathbf{D}^t(\mu, Q^{L,N}) + \beta^L \mathbf{R}^t(\mu, Q^{L,N}).$$

Thus the Lagrangian approach, crucial to the derivation in [12] of the source coding theorems, have strong footing in the realm of practical source code design and implementation. What's more, careful examination of the weighted distortion-rate function indicates that for any point on the above convex hull, α_ℓ / β_ℓ equals the negative slope of the rate-distortion function at distortion D_ℓ. Thus in choosing α^L, β^L, this ratio should increase as a function of ℓ.

The above proposal differs markedly from the sequential minimizations of the top-down or bottom-up approaches of the previous section. Top-down approaches sacrifice high resolution performance for low-resolution performance and are thus most appropriate when focusing on the low-end user, who values the quality of the low-rate reproduction over the long term coding performance. Bottom-up approaches, sacrifice low resolution performance for high-resolution performance and are thus best suited for applications where the priority falls with the high-end user, who is willing to sacrifice initial reproduction quality for the fastest possible high-resolution service. Alternate algorithms like those of [18] take this emphasis on the high end user one step further by combining a bottom-up approach with a full search (rather than tree search). Between these two extremes lies a continuum of other possibilities neglected by the vast majority of multi-resolution codes but addressable using the above Lagrangian approach. We next discuss the design of source codes with arbitrary priority schedules between the low- and high-resolution coding quality.

The proposed multi-resolution coding performance measure

$$J_{\alpha^L,\beta^L}(\mu, Q^{L,N}) = \alpha^L \mathbf{D}^t(\mu, Q^{L,N}) + \beta^L \mathbf{R}^t(\mu, Q^{L,N}).$$

equals the weighted sum of the distortions and rates of every resolution of a source code. By using this measure in both code design and implementation, we can optimize the entire multi-resolution code with respect to a priority schedule of our choosing. Just as the performance measure $D + \lambda R$ may be interpreted as a Lagrangian for minimization of the distortion subject to a constraint on the rate, the above proposed weighted performance measure likewise has a variety of Lagrangian interpretations dependent on the choice of α^L and β^L. In particular, the proposed performance measure may be viewed as a Lagrangian for minimizing the weighted sum of distortions subject to a collection of constraints on the rates. The same function could likewise be interpreted as a Lagrangian for the minimization of the weighted sum of rates subject to a collection of constraints on the distortions, a minimization of the weighted sum of rate-distortion Lagrangians, or in fact a minimization of any combination of rates, distortions, or Lagrangians subject to constraints on the remaining quantities. Thus, for example, when $L = 2$, the weighted rate-distortion function can be used to find: the minimal R_2 needed to achieve distortion $D_2 < D_1$ given $(R_1, D_1) = (R(D_1), D_1)$; the minimal R_1 needed to achieve distortion $D_1 > D_2$ given that $(R_2, D_2) = (R(D_2), D_2)$; the minimal $D_1 + \lambda_1 R_1$ given that $D_2 + \lambda_2 R_2 \leq D_2 + \lambda R_2(D_2)$; and so on, thereby encompassing a variety of multi-resolution source coding problems considered by previous authors into a single formulation. This variety of interpretations illustrates one of the benefits of the Lagrangian approach.

In general, the $2L$-dimensional vector (α^L, β^L) plays a role in multi-resolution source coding analogous to the role played by λ in entropy-constrained vector quantization (ECVQ) [13]. In ECVQ, we design and implement our codes to achieve the lowest possible value of $D + \lambda R$. Just as λ represents the (negative) slope of the line tangent to the rate-distortion curve $R(D)$ at the target rate and distortion, (α^L, β^L) describes the "angle" of the hyperplane tangent to the achievable rate-distortion region at the target L-dimensional rate-distortion vector.

To achieve points as close as possible to the bounds discussed in Section II, we use the given performance measure for both source code design and encoding. The design algorithm is a variation on the generalized Lloyd algorithm [19] which simultaneously optimizes all of the codewords in a tree-structured code with respect to the given performance measure. The algorithm is initialized with an arbitrary tree-structured codebook along with appropriate fixed- or variable-rate prefix codes for describing each possible transition within the given tree. Each iteration in the algorithm is accomplished in three steps, enumerated below.

1. *Nearest Neighbor Encoding.* Optimize the encoder for the given tree-structured codebook and prefix codes. The optimal first-stage encoder maps each source vector to the path with the minimal weighted performance.

2. *Decoding to the Centroid.* Optimize the codebook for the given encoder. Each codeword in the code is redesigned such that it sits at the centroid (with respect to the employed distortion measure) of all training data mapped to that codeword.

3. *Optimizing the Prefix Code.* Optimize the prefix codes associated with each transition. For fixed-rate coding no change occurs. For variable-rate coding, the optimal prefix code is the prefix code mapped to the probability of a given transition.

The design is not performed layer by layer and is not greedy. When run to convergence, the above algorithm guarantees a locally optimal solution over the entire tree – a guarantee not made by either top-down or bottom-up greedy approaches.

One unfortunate consequence of the above algorithm is that it employs the weighted performance measure in its optimal encoder. As a result, the proposed algorithm not only sacrifices the linear search complexity from which TSVQ's original popularity derived, but actually almost doubles the number of codeword distortion calculations required by a full-search vector quantizer. We therefore propose an alternate in addition to the above "optimal" scheme. The alternate code is designed using precisely the algorithm described above but implemented using the traditional greedy TSVQ encoder rather than the above optimal encoder. The performance of such "hybrid" codes will be explored in later work.

V   Results

For simplicity, we include results from fixed-rate multi-resolution codes only. For any fixed-rate quantizer $Q^{L,N}$ and any stationary source μ, $\beta^L \mathbf{R}^t(\mu, Q^{L,N})$ is constant. Thus, in this scenario, application of the weighted performance measure in system design is equivalent to optimization of the weighted sum of the distortions associated with each resolution of a binary tree code. Since it is the relative values of $\alpha_1, \ldots, \alpha_L$ which interest us, we restrict our attention to α^L such that $\sum_i \alpha_i = 1$. When $\alpha^L = (1, 0, \ldots, 0)$, designing a source code to minimize $J_{\alpha^L, \beta^L}(\mu, Q^{L,N})$ is a true top-down approach, sacrificing all higher-resolution performances for the best possible rate-1 performance. When $\alpha^L = (0, \ldots, 0, 1)$, designing a source code to minimize $J_{\alpha^L, \beta^L}(\mu, Q^{L,N})$ is a bottom-up approach, yielding the best possible high-resolution performance at the possible expense of low-resolution performance as in [18]. Notice that according to the above discussion for $L > 2$, neither of these points will sit on
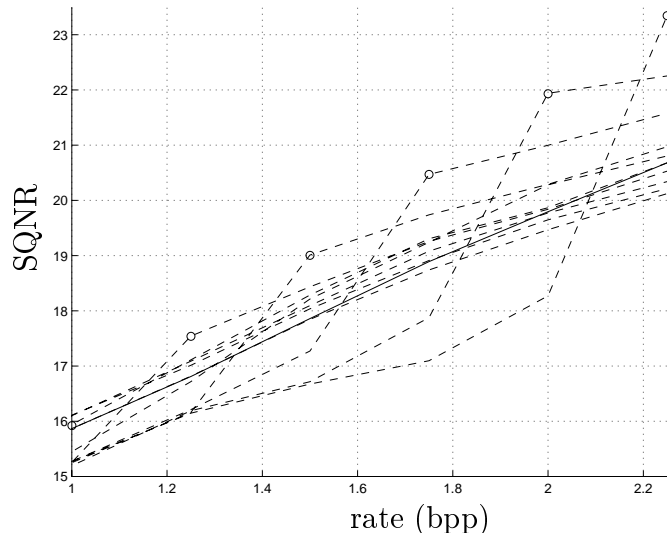
Figure 1: SQNR vs. rate results for fixed-rate multi-resolution VQ using the weighted performance measure (dashed lines) and non-embedded VQ (circles) and TSVQ (solid line).

the convex hull of the rate-distortion region. Values of α^L between these extremes lead to intermediate solutions.

We here report performance results on a test set of 5 medical brain scans using a code designed on a 20 image training set. The training and test sets do not overlap. The codes considered use vector dimension 4. The results are shown in Figure 1, where each dotted line shows the rates and signal to quantization noise ratios (SQNRs) associated with the above described code for a single value of α^L. Distortion is measured as squared error. A range of α^L vectors is shown, not all of which should give values residing on the convex hull of the rate-distortion region. The solid line shows the performance of TSVQ while the performance of a (non-embedded) collection of full search VQs is given by circles. While gains of .5 to .75 dB over TSVQ can be achieved at some rates with almost no loss at other rates, achieving the performance of a full search VQ at the higher rates seems to require corresponding losses at lower rates.

References

[1] E. A. Riskin and R. M. Gray. A greedy tree growing algorithm for the design of variable rate vector quantizers. *IEEE Transactions on Signal Processing*, 39:2500–2507, November 1991.

[2] W. B. Pennebaker and J. L. Mitchell. *JPEG Still Image Compression Standard*. Van Nostrand Reinhold, New York, 1993.

[3] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41:3445–3462, December 1993.

[4] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3):243–250, June 1996.

[5] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[6] M. Effros, P. A. Chou, and R. M. Gray. Variable-rate source coding theorems for stationary nonergodic sources. *IEEE Transactions on Information Theory*, IT-40(6):1920–1925, November 1994.

[7] R. M. Gray and A. D. Wyner. Source coding for a simple network. *Bell Systems Technical Journal*, 53(9):1681–1721, November 1974.

[8] V. Koshelev. Hierarchical coding of discrete sources. *Probl. Pered. Inform.*, 16(3):31–49, 1980.

[9] V. Koshelev. An evaluation of the average distortion for discrete schemes of sequential approximation. *Probl. Pered. Inform.*, 17(3):20–33, 1981.

[10] W. H. R. Equitz and T. M. Cover. Successive refinement of information. *IEEE Transactions on Information Theory*, 37(2):269–275, March 1991.

[11] B. Rimoldi. Successive refinement of information: characterization of achievable rates. *IEEE Transactions on Information Theory*, 40(1):253–259, January 1994.

[12] M. Effros. Rate-distortion bounds for fixed- and variable-rate multi-resolution source codes. 1997. In preparation.

[13] P. A. Chou, T. Lookabaugh, and R. M. Gray. Entropy-constrained vector quantization. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(1):31–42, January 1989.

[14] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley and Sons, New York, 1969.

[15] A. Buzo, A. H. Gray Jr., R .M. Gray, and J .D. Markel. Speech coding based upon vector quantization. *IEEE Transactions on Acoustics Speech and Signal Processing*, 28:562–574, October 1980.

[16] 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*, IT-35(2):299–315, March 1989.

[17] M. Vishwanath and P.A. Chou. An efficient algorithm for hierarchical compression of video. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 275–279, Austin, TX, November 1994. IEEE.

[18] E. A. Riskin, R. Ladner, R.Y. Wang, and L. E. Atlas. Index assignment for progressive transmission of full-search vector quantization. *IEEE Transactions on Image Processing*, 3(3):307–312, May 1994.

[19] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28:84–95, January 1980.