Joint Source and Channel Coding for Image Transmission Over Lossy Packet Networks

Geoffrey Davis and John Danskin

Email: {*gdavis, jmd*}*@cs.dartmouth.edu*
6211 Sudikoff Laboratory, Dartmouth College, Hanover, NH 03755

August 5, 1996

ABSTRACT

We describe a joint source/channel allocation scheme for transmitting images lossily over block erasure channels such as the Internet. The goal is to reduce image transmission latency. Our subband-level and bitplane-level optimization procedures give rise to an embedded channel coding strategy. Source and channel coding bits are allocated in order to minimize an expected distortion measure. More perceptually important low frequency channels of images are shielded heavily using channel codes; higher frequencies are shielded lightly. The result is a more efficient use of channel codes that can reduce channel coding overhead. This reduction is most pronounced on bursty channels for which the uniform application of channel codes is expensive. We derive optimal source/channel coding tradeoffs for our block erasure channel.

Keywords: joint source/channel coding, erasure channel, lossy image transmission

1 INTRODUCTION

With the rising popularity of Web browsers, image transmission has become one of the largest uses of Internet bandwidth [5]. Responsiveness is essential in interactive applications such as browsers, even more so than perfect image fidelity, since many images have already been distorted by lossy compression techniques.

The standard method for transmitting images over the Internet is to first apply a lossy subband-based compression scheme such as JPEG and to then transmit the compressed images across the intrinsically lossy Internet using the automatic repeat request (ARQ) based TCP/IP protocol. While ARQ-based protocols transmit images losslessly, the resulting packet retransmissions can lead to excessive transmission delays in networks with high latency. ARQ-based protocols are also ill-suited for multicasting applications. Because losses experienced by different multicast recipients will in general be different, retransmission of lost packets becomes impractical.

An alternative strategy is to apply forward error correction (FEC) to packets. Forward error correction must be applied with care, however, since the added redundancy increases network loads and can in some cases actually degrade overall network performance [3].

In this paper we consider the problem of how to make the most efficient use of forward error correction for transmission of images over lossy packet networks. We examine and compare three basic strategies: the uniform application of channel codes, a joint optimization the distribution of source and channel codes, and the utilization of robust encodings. In each case we provide motivation based on optimal information-theoretic considerations.

1.1 Related Work

A number of redundant lossy packet transmission schemes have been examined for images and video. The algorithms of [23] and [12] make use of naturally occurring redundancy within images to recover from packet losses. The number of losses that can be sustained is highly image dependent in this case, and only limited compression can be used. Our allocation schemes allow for precise control of the distribution of redundancy.

The Priority Encoding Transmission (PET) scheme [1][13] allows the user to set different levels of error protection for different portions of the MPEG stream, but unlike this paper provides no explicit mechanism for allocating these levels. Layered transmission schemes such as [8][15] incorporate similar ideas, but require networks which treat packets differently according to their priorities. Our allocation scheme can be used with any network supporting a simple datagram protocol.

Our contribution is to provide a simple, low-complexity mechanism for obtaining an optimized partitioning of bits between image quantization and redundancy for a given set of image transform quantizers, an error correction scheme, and a packet loss model. We analyze the asymptotics of this tradeoff and show connections between our methodology and existing work on vector quantization for noisy channels.

2 OPTIMAL TRADEOFF BETWEEN SOURCE AND CHANNEL CODING

2.1 Asymptotic Results

The problem we will address is that of transmitting images as collections of b bit packets over a lossy network. The class of protocols we consider has two important properties. First, packets may be delivered out of order, so each packet contains a unique identifier. Second, the contents of all packets are verified during transmission. Thus our channel is essentially a block erasure channel for which b bit erasures occur only at packet boundaries.

To simplify our analysis we will first consider independent packet erasures that occur with a constant probability $1 - p$. We will examine more realistic models of Internet loss behavior below. Furthermore, we will limit our attention to forward error correction in the form of linear block codes.

Suppose we wish to transmit n samples of an independent, identically distributed (i.i.d.) process over a lossy packet network using N packets. We desire that the total expected error, both from packet erasures and from quantization, be minimized. Clearly there is a tradeoff between source and channel coding. If we increase the number of bits devoted to error correction, we increase the probability of receiving a message correctly, but we lose quantizer resolution. In the extreme case we send no information, but it is received with perfect fidelity. On the other hand, increased quantizer resolution comes at the expense of increased probability of uncorrectable erasures.

In this section we estimate the rate at which some specific source and channel coding strategies approach the asymptotic limits imposed by channel capacity. We use our estimates to derive an asymptotically optimal tradeoff between source and channel coding for high rates. We then use our estimates to motivate approaches to improving transmission rates in more practical regimes. For our analysis we assume that source coding is performed using an optimal n dimensional vector quantizer and that channel coding is performed using (N, k, d) block codes that are optimal in the sense that they achieve the Singleton bound, $d \leq N - k + 1$. In addition, these codes allow us to approach the channel capacity arbitrarily closely for N sufficiently large and with sufficiently large packets. We discuss the implementation of such optimal codes, called maximum distance separable (MDS) codes, in the next section.

Let r be the fraction of bits devoted to source coding and let R be the number of bits per sample. In our framework we have $rR = \frac{rNb}{n}$ bits per sample available for source coding and $(1-r)R$ bits per sample devoted to channel coding. Our channel codes permit the correction of up to $(1-r)N$ erasures. Our goal is to minimize the expected squared error of our received signal by choosing an optimal r. For simplicity, we will use a squared error distortion measure, and we will take our signal to be n samples of an i.i.d. mean 0 Gaussian source with variance σ^2. The calculations below may be generalized in a straightforward manner to non-Gaussian sources as well as to errors of the form $d(\mathbf{x}, \hat{\mathbf{x}}) = ||\mathbf{x} - \hat{\mathbf{x}}||^\alpha$ where the norm is the standard l^2 norm. Our development is motivated by related estimates of.[11]

We obtain the quantization error for our source using Zador's estimate [25][9]. For high rates, the average squared quantization error per symbol for an optimal n dimensional vector quantizer using rR bits per sample is given by

$$D(r) = 2\pi e \sigma^2 B(n,2) 2^{-2Rr} \tag{1}$$

where $B(n,2)$ is a packing constant satisfying

$$\frac{1}{1+\frac{2}{n}} \omega_n^{-\frac{2}{n}} \leq B(n,2) \leq \Gamma(1+\frac{2}{n}) \omega_n^{-\frac{2}{n}} \tag{2}$$

and ω_n is the volume of the n-dimensional unit sphere.

We obtain a complete quantized vector whenever there are $(1-r)N$ or fewer erasures. Let p be the probability that a single packet is lost. The number of packets successfully transmitted has a mean of Np and a variance $s^2 = \frac{p(1-p)}{N}$. For N large, we use the central limit theorem to estimate the probability $P(r)$ of a transmission failure,

$$P(r) \approx \int_{\frac{p-r}{s}}^{\infty} e^{-x^2} dx. \tag{3}$$

For now we assume that all data is lost if there are any unrecoverable packet errors. A number of recently developed image compression techniques, such as Shapiro's EZW scheme [19], are extremely sensitive to erasures, so this is not an entirely unreasonable distortion model. We will consider better decoding strategies below in section 3.4. The expected squared error per-sample under these assumptions is

$$E(D) = P(r)\sigma^2 + (1 - P(r))D(r). \tag{4}$$

We now find the optimal tradeoff between source and channel coding for large numbers of samples n. We will keep R, the total number of bits per sample, fixed as we take n to infinity.

When R and n are sufficiently large we have either $P(r) << 1$ or $D(r) << \sigma^2$ (or both). For the high rate case, then, we ignore the $P(r)D(r)$ term in $E(D)$ and examine the sum (actually an upper bound on $E(D)$) given by

$$E(D) \approx P(r)\sigma^2 + D(r). \tag{5}$$

Setting the r derivative of (2.1) to 0 and taking the log, we find for large N that

$$(p-r)^2 = 2p(1-p)\frac{1}{N}\left\{\ln(\sqrt{N}) - \ln\left[\frac{\sqrt{2\pi}}{p(1-p)}\right] + 2\ln(2)Rr - \ln\left[4\pi e \ln(2) B(n,2) R\right]\right\}. \tag{6}$$

For N sufficiently large the right hand side is dominated by the $\ln(N)$ term, and we find that distortion is minimized for

$$r \approx p - \sqrt{p(1-p)\frac{\ln N}{N}}. \tag{7}$$

For this rate we have

$$D(r) \approx \sigma^2 2^{-2Rr}. \tag{8}$$

Related results for the binary symmetric channel may be found in [11].

We see that as N tends to infinity, r tends to the channel capacity, p, and the overall distortion tends to the rate-distortion curve for a set of i.i.d. Gaussian samples. Even away from the asymptote, when N is sufficiently large the dependence of the tradeoff parameter r on the overall rate R becomes negligible. If our goal is to encode realizations of a collection of independent mean zero Gaussian sources, rather than a single Gaussian source, then for large N the optimal solution is to divide the rNb source coding bits using the standard reverse water-filling algorithm [4]. When N is large there is virtually nothing to be gained by varying the allocation of channel codes by subbands in an image.

The overall rate R does play a role in determining r for smaller values of N, however. A more careful examination of the right hand side of (6) reveals that when the rate R is large we require a higher proportion of channel code. This suggests that for small N we will obtain reduced errors when coding collections of Gaussians by allocating higher levels of channel codes to the high variance variables.

The independence of r from R is due to the asymptotic efficiency of the codes we use. We expect gains from joint source/channel coding for bursty channels for which our codes are less effective, and if we are forced to limit code block lengths due to constraints on computational complexity. As we see in the next section, our experimental results bear out these hypotheses.

2.2 Maximum Distance Separable Codes

The codes we use in this paper are based on Rabin's Information Dispersal Algorithm[17] (IDA) and its variants [16]. The basic idea is a simple one: our codes consist of N dimensional vectors over a finite field $GF(2^c)$, where c is a constant that divides b, the total number of bits in a packet. The data we wish to transmit occupies a subspace of rank k. We recover data by projecting a vector of k received c-bit words onto the k-dimensional source code subspace.

We embed the k-dimensional data space into the N dimensional space using an $N \times k$ matrix A. To ensure that we can recover the data from any k words, A must have the property that any k of its columns are independent. One suitable matrix is the Vandermonde matrix with entries $A_{ij} = \alpha^{i+j}$, where α is an element of $GF(2^c)$ of order $2^c - 1$.

Security considerations in the design of IDA scheme warrant the use of a matrix A such that *no* data values can be recovered with fewer than k words. As we discuss below, we would like to be able to make use of partial data even if there are more than $N - k$ losses. We therefore modify Rabin's scheme by placing the matrix A in systematic form.

In addition to increasing our ability to use partial data, our modification reduces the overall complexity of the scheme. The dominant cost in Rabin's implementation is a matrix multiplication requiring k operations per word. For our modified implementation the cost is l operations per sample where l is the number of lost data values. Since we have $E(l) = (1-p)k$, this represents a large savings for small loss rates. In our implementation we use the extension field $GF(2^8)$, so each word corresponds to a byte. Additions are done using exclusive-or operations, and multiplications of field elements are stored in a lookup table, so the calculations required for reconstructing lost samples are inexpensive.

3 JOINT SOURCE AND CHANNEL CODING

We use a simple wavelet transform coding scheme for our numerical experiments. We perform a 5-level wavelet transform with symmetrized boundaries using the 7/9-tap biorthogonal wavelet of [2]. Coefficients are quantized using the embedded quantization scheme of [22] and entropy coded using an adaptive arithmetic coder. There is no training involved. We describe bit allocation in more detail below. This simple scheme, despite its lack of higher level data structures such as zerotrees, yields PSNR's roughly comparable to those of Shapiro's embedded zerotree wavelet coder[19] due to its more efficient bit allocation procedure.

We have chosen to work with a wavelet-based scheme because of its simplicity and good performance at low bit rates. The ideas presented below can be generalized to other subband-based schemes such as JPEG. Adapting the ideas to DCT-based schemes, for example, requires replacing wavelet subbands in the discussion below with blocks of DCT coefficients of comparable frequency.

3.1 Bit Allocation for Source Coding

The discrete wavelet transform partitions an image into a set of subbands ranging from fine scales (high frequency) to coarse (low frequency). In natural images the bulk of the visually important information is concentrated in the coarse-scale subbands, with the fine-scale subbands contributing primarily near sharp edges. We obtain a compressed image by finely quantizing coefficients that contribute heavily to image fidelity and coarsely quantizing others. Determining the quantization resolution of each subband is a problem of resource allocation. We have a tradeoff between quantization error and total storage cost, and we must allocate quantizer resolutions to obtain minimal distortion for our given bit expenditure.

Our task is to select one of a family of quantizers $\mathcal{Q}_0 \ldots \mathcal{Q}_K$ for each image subband. The quantizers are arranged from coarsest ($\mathcal{Q}_0$) to finest ($\mathcal{Q}_K$) and are have bin widths that are scaled according to the range R_j of coefficients in each subband. Quantizer $\mathcal{Q}_k$ has $2^k - 1$ bins. One bin, of width $2^{-k+1}R_j$ is centered at the origin. The other $2^k - 2$ bins are spaced uniformly and symmetrically around the center bin and have width $2^{-k}R_j$. This family of quantizers has the important property that quantizer bins are nested, i.e. each bin of $\mathcal{Q}_k$ can be decomposed into either two or three bins in $\mathcal{Q}_{k+1}$.

Except for the central double-width bin around 0, each quantizer bin in $\mathcal{Q}_k$ is refined to two equal-sized bins in the next finer quantizer $\mathcal{Q}_{k+1}$. The central width w "dead-zone" bin around 0 in quantizer $\mathcal{Q}_k$ is refined into 3 bins in quantizer $\mathcal{Q}_{k+1}$, a central bin of width $\frac{w}{2}$ and two side bins of width $\frac{w}{4}$. We express the output of the quantizer $\mathcal{Q}_k$ as a string of refinements $(r_0, r_1, \ldots, r_k)$, where each of the r_i's is a 0, 1, or 2. The sets of refinements are essentially the bitplanes of the coefficients ordered from the most significant bit to the least, and we will refer to them as such for convenience. By conditioning the refinement r_j on the previous refinement r_{j-1} we obtain the embedded representation with no increase in storage cost over a non-embedded coder. (On the contrary, in practice we find that the estimates of the density function adapt much faster for the embedded case, so we actually obtain slightly better results by using the embedded coder.)

Each quantizer $\mathcal{Q}_k$ applied to subband j has associated with it a cost $C_j(k)$ of storing the entropy coded quantized values and a distortions $D_j(k)$. We use the total squared error as a distortion measure, but the algorithms described in this paper will function equally well with other additive metrics such as the perceptually weighted metric described in [14].

For an image decomposed into n subbands, our goal is to find a vector $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ of quantizer indices so that the total distortion $D_{total}(\mathbf{q}) = \sum_{j=0}^{n} D_j(q_j)$ is minimized subject to the constraint that the total cost in bits, $C_{total}(\mathbf{q}) = \sum_{j=0}^{n} C_j(q_j)$ is less than or equal to some given bit budget C_{max}. Thus we seek a minimization over $\mathbf{q} \in Q$ where Q is a given set of valid vectors of quantizer indices.

We use a discrete Lagrange multipliers scheme to minimize D_{total} subject to the constraint. We minimize the sum $D_{total} + \lambda C_{total}$ where an appropriate λ is found using a binary search as described in Shoham and Gersho[20]. Because not all cost constraints are achievable using this Lagrangian scheme, we allocate any remaining bits using marginal analysis.

3.2 Bit Allocation for Joint Source/Channel Coding

We can optimize the distribution of channel codes for subbands using a simple extension of the above allocation procedure. Each subband in an N-packet transmission is protected by an (N, k) code, where k ranges from 0 to N. We denote the level of error correction for subband j by m_j, where m_j indicates that an $(N, N - m_j)$ code is used for subband j.

We replace the cost and distortion functions $C_j(q_j)$ and $D_j(q_j)$ with the functions $\hat{C}_j(q_j, m_j)$ and $\hat{D}_j(q_j, m_j)$ that incorporate the cost of the channel codes and the resulting expected distortion incurred in transmission. The new cost function $\hat{C}_j(q_j, m_j)$ will equal the old $C_j(q_j)$ plus the number of bits used for the channel coding. The new distortion function $\hat{D}_j(q_j, m_j)$ is obtained using the expected distortion given the channel code $(N, N - m_j)$. We again use a discrete Lagrange multipliers scheme followed by marginal analysis for the allocation.

The results of this procedure are illustrated in figure 1. The increased flexibility in channel code allocation yields reductions in the expected squared error of up to 1.2 dB at low rates. Moreover, as figure 1 shows, the amount of channel coding overhead required to achieve these low errors is 15 to 25 % less than in the uniformly distributed case. The effects of our joint source/channel optimization are even more pronounced for bursty channels and when the code lengths are restricted. Figure 2 illustrates the results of our optimization for restricted code lengths and for a bursty channel. Losses in the bursty channel follow a Markov model: the probability of an erasure is 60% if the previous packet was erased and 10% if not, yielding an overall erasure rate of 20%.

The data in these figures has been computed analytically from the optimizer output. Actual implementation will involve some small amounts of additional overhead to describe quantizer parameters, etc. Achieving the prescribed channel coding levels per subband is possible by partitioning each network packet into subsets protected by channel codes of different strengths. Such a partitioning is illustrated in figure 3.

3.3 Bitplane-Level Allocation

Even at the subband level, not all bits in our compressed image contribute equally to reconstructed image fidelity. In particular, the coarsest scale layers of our nested quantization scheme contribute much more than do the fine scale layers. In fact, we can convert *any* scalar quantization scheme to such a nested scheme by grouping adjacent quantizer bins and using an entropy coding scheme with conditioning. We can obtain further reductions in expected distortion by allocating our channel codes at the "bitplane" level.

Allocation of channel codes at the bitplane level is complicated by the fact that successive bitplanes are coupled. Entropy coding of fine scale bitplanes is conditioned on coarser scale bitplanes, so the loss of a coarse scale bitplane results in the loss of all finer scale bitplanes. Algorithms have been developed for handling such dependent quantization problems for MPEG[18]. As we show below, however, in this particular case the coupling

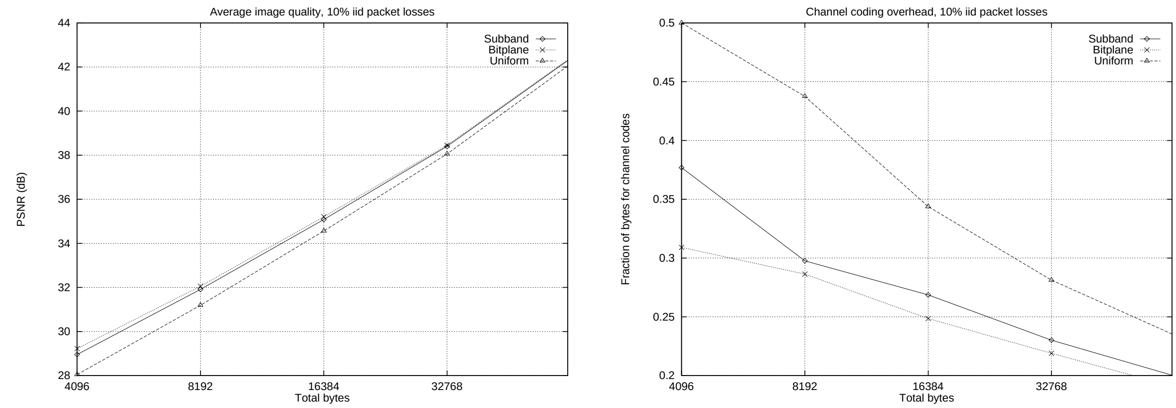


Figure 1: Image quality and fraction of total bytes devoted to channel coding as a function of the total byte budget. Our image quality metric is expected squared distortion, and we have expressed this as a peak signal to noise ratio. The test image is the 512×512 Lena image. Packets are erased independently with a 10% probability of erasure. Results are shown for (1) optimal uniformly distributed channel codes, (2) channel codes optimized per subband, and (3) channel codes optimized per bitplane.

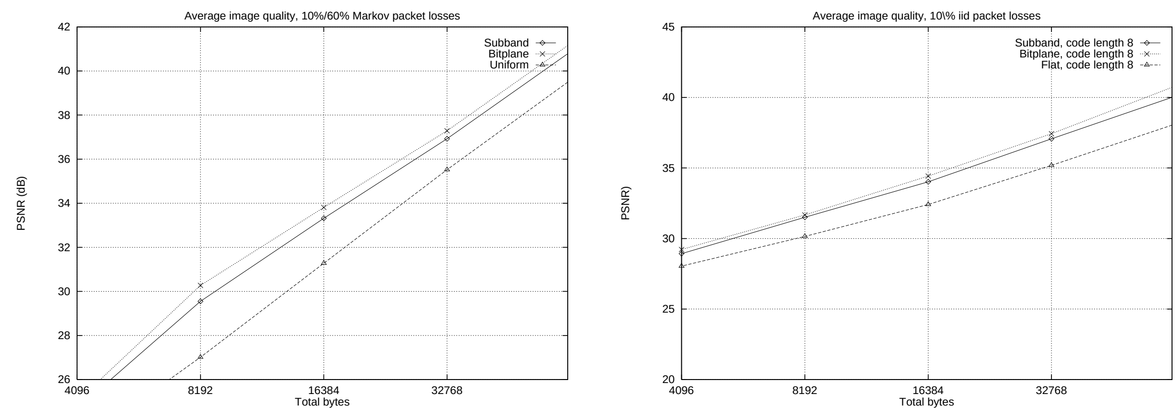


Figure 2: Image quality as a function of the total byte budget. For the plot on the left, packets are erased in bursts following a Markov model: the probability of an erasure is 60% if the previous packet was erased and 10% if not. The benefits of joint source/channel coding are even more pronounced for such a channel than for figure 1. The channel on the right is as in figure 1, but all channel codes are limited to length 8. Again the benefits of joint source/channel coding are enhanced.

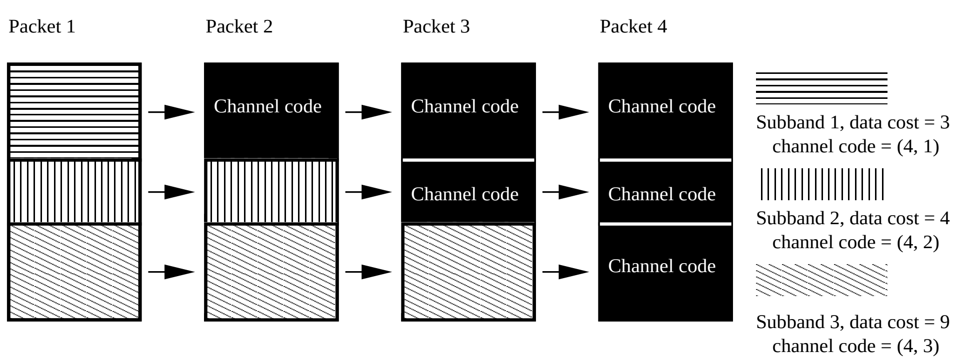


Figure 3: Distribution of data and channel codes into N network packets. A series of (N, k) block codes are formed from byte j of each packet. By varying k for each byte we can create packets in which data has varying levels of protection.

is not problematic.

We require an assumption for our allocation procedure: that the operational rate-distortion curve produced by our quantization procedure is convex. Let $C_{j,k}$ and $D_{j,k}$ be the cost and the reduction in distortion obtained from quantizing bitplane k of subband j. We assume that $\frac{D_{j,k}}{C_{j,k}} \leq \frac{D_{j+1,k}}{C_{j+1,k}}$. This has been verified empirically for our test images. Moreover, analysis of Gish and Pierce[10] show that it holds at high rates for entropy coded, uniformly quantized i.i.d. sources.

If there were no coupling between the bitplanes, we could find an optimal allocation using Lagrange multipliers. We find for each subband j and bitplane k

$$\max_{m_{j,k}} \left\{ D_{j,k} P(m_{j,k}) - \lambda C_{j,k} \left(\frac{N}{N - m_{j,k}} \right) \right\}. \tag{9}$$

Here $m_{j,k}$ denotes coding with an $(N, N - m_{j,k})$ coder and $P(m_{j,k})$ is the probability that $m_{j,k}$ or fewer errors occur during transmission of the N packets. If the maximum of this difference is positive, then we use $(N, N - m_{j,k})$ codes for bitplane k of subband j; otherwise we discard the bitplane.

Now because of our assumed convexity we must have $m_{j,k} \geq m_{j,k+1}$, i.e. coarser bitplanes are more heavily protected than finer bitplanes. The coupling between bitplanes leaves this inequality unchanged, since the coupling increases the channel coding requirements of coarse bitplanes relative to fine bitplanes. In fact, the monotonicity of the levels serves to decouple the allocation problem. The monotonicity means that fine scale bitplanes will always be lost before coarse quantizer bitplanes. Fine scale bitplanes are never successfully transmitted without the coarser bitplanes upon which they depend, so the above maximization of (9) in fact yields the optimal coupled solution.

In figure 1 we see that this bitplane-level optimization yields a modest additional reduction in expected distortion as well as further reductions in channel coding overhead over the subband-level optimization.

3.3.1 Geometric Interpretation

The above bitplane-level bit allocation procedure ties in with work on vector quantization for noisy channels. Zeger[26] has shown that one can obtain better results by transmitting information over a noisy channel using a high resolution vector quantizer with no explicit error control than by using a lower resolution quantizer with forward error correction. The reason is that error coding entails the explicit assignment of collections of received vectors to a single code value. With straight vector quantization, on the other hand, received vectors can be arbitrarily assigned to decoded values. This additional freedom can be used to improve the performance of the coder. Steps towards constructing such channel-optimized vector quantizers have been described in [27] and [7]. In each of these works, quantizer indices are assigned so that Hamming distances between codewords correspond to the Euclidean distances between their decoded values. Our construction embodies similar ideas, but entails the use of much lower complexity scalar quantizers.

In our case, we would like to code our data in such a way that small numbers of erasures correspond to small Euclidean distances between received codewords. In our present framework, we distribute n quantized samples over N packets, and the loss of a single packet prevents the decoding of the entire set of samples. Hence any nonzero number of erasures corresponds to a large Euclidean distance. The addition of error correcting codes increases the correspondence between the numbers of erasures and Euclidean distance: small numbers of erasures correspond to a Euclidean distance of 0, and larger numbers of erasures correspond to a large Euclidean distance. The above bitplane-level assignment of forward error correction takes this process one step further–quantizer reconstructions degrade gracefully as the number of erasures increases. The result is finer control over the assignment of received vectors to decoded values than is attainable with a single level of correction for all bitplanes.

3.4 Error-Resilient Source Coding

We can further reduce the expected distortion by carrying the above analogy with vector quantization one step further. Consider the problem of decoding a vector distributed over N packets when the last packet has been erased. The decoded value that minimizes the error satisfies a centroid condition: it is the average of the decoded values for all vectors that whose first $(N-1)b$ bits agree with the $(N-1)b$ received bits, weighted by the probability of generating each vector. The analogous process for a fixed-rate scalar quantizer is to decode erased variables to their means. This is a straightforward process in our implementation, since high-pass subband coefficients are known *a priori* to have means of roughly 0.

We can get an idea of how the use of this fragmentary data affects the optimal asymptotic tradeoff between source and channel coding by modifying our vector quantization from section 2.1. Rather than using a vector quantizer whose output spans all N packets, we reduce the dimension so that the output of the quantizer spans a single packet. With this modification, erased packets affect only a small number of data values. For large N the role of this fragmentary data in the tradeoff is negligible, since the resulting modifications of $P(r)$ introduce only a polynomial modification of a function whose behavior is dominated by an exponential. However, we once again find that this fragmentary data plays a useful role for the values of N in which we are interested.

Making use of fragmentary information in variable length scalar quantizers is more difficult than for fixed rate VQ's, because we must determine which particular coefficients have been erased. One way to accomplish this is by adding synchronization data to packets. For example, for each level of channel coding in each packet we can add a tag indicating the index of the first coefficient at that level. (Making use of fragmentary information at the bitplane level is further complicated by the interdependencies between bitplanes; methods of efficiently making use of this fragmentary information are currently under investigation.) For subband-level channel optimization, the side synchronization information is small relative to the total packet size (Internet packets can be as large as 576 bytes without being fragmented by routers) and is also strongly dependent on the method of implementation. Although our experiments do not incorporate the cost of this side information, they should give a good approximation of what is attainable by using fragmentary data in the subband-level case.

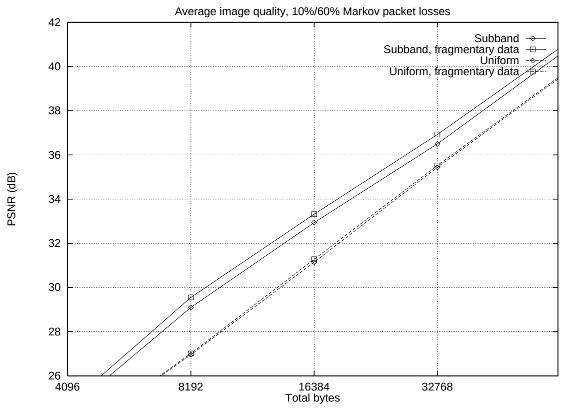
As we see from figure 4 we obtain only a small additional reduction in the expected squared error from using this additional fragmentary data. As we discuss in the next section, the minimum squared error metric results in a very conservative allocation with heavy use of channel coding. Losses occur insufficiently frequently for the use of fragmentary data to have much effect. We discuss alternative error metrics below.

4 DISTORTION MEASURES

The expected squared error distortion measure is widely used due to its simple mathematical properties. It is a natural measure for us to use given that our development has been motivated by work in a more abstract information theoretic setting. We now examine in more detail the suitability of this metric to our purposes. Our goal is to speed the transmission of images sent over the Internet. We are concerned primarily with the most frequently observed image qualities. Catastrophic events that occur extremely rarely are not of great importance.

We can get an better idea of the average user-experienced quality of our error-corrected images by ignoring levels of losses that occur infrequently, say less than 1% of the time. The plot on the right of figure 4 shows the result of our optimization when we compute our expectations over frequent levels of loss, i.e. we exclude the 99^{th} percentile high loss events. We find that the gains in expected squared error from joint source/channel coding are significantly reduced under this restricted expectation.

The penalty for a lost packet under our squared error metric is quite severe. The large size of these penalties means that rare events, far out in the tail of the error cdf, can produce large changes in the overall average



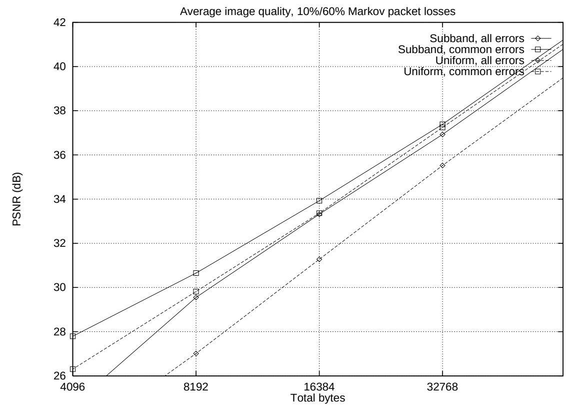


Figure 4: Image quality as a function of the total byte budget. For both plots, packets are erased in bursts following a Markov model: the probability of an erasure is 60% if the previous packet was erased and 10% if not. The plot on the left shows that being able to make use of fragments of block coded data does not greatly reduce the expected error. The plot on the right compares expected errors computed over the set of all possible errors to expected errors computed over the set of errors that occur 99% of the time; rare catastrophic errors have been excluded. The comparison illustrates a problem with the expected squared error metric: it is strongly influenced by rare, catastrophic events.

distortion. Moreover, even in the error-limited regime, bit allocation is done in a very conservative fashion–the optimizer sets the probability of uncorrectable errors to be nearly 0 for all subbands and bitplanes. Table 5 shows that channel code assignments for the L^2 subband-level optimized allocation are very similar to the optimal uniform channel code assignments.

We can achieve less conservative allocations by changing our norm. The L^1 norm imposes much smaller penalties for packet drops than does L^2. As a result, our L^1-optimized distributions of channel codes are much less flat. They significantly reduce the channel coding overhead by permitting relatively rare losses of high frequency data. Moreover, DeVore et al[6] make the case, based on arguments about the frequency sensitivity of the human visual system, that L^1 is a more appropriate distortion measure for images than L^2. Experiments by Healy et al[24] also suggest that the L^p norms for small values of p correspond more closely to perceived distortion than does L^2.

We see from figure 5 that the L^1 optimization results in a gradual reduction in the amount of channel coding protection of subbands. In the L^1 optimized subband-level allocation, 14% of all bytes are used for channel coding, versus 20% for the L^2 optimized subband-level allocation and 22% for the L^2 optimized uniform allocation. With the L^1 norm we are able to use a higher rate coder, giving us a higher quality image for much of the time and a gradual degradation under increasing error loads.

5 CONCLUSION

Joint source/channel coding, although asymptotically of little benefit, provides significant gains in expected image quality for regimes of practical interest. These gains are particularly pronounced for bursty channels and when channel code block lengths are restricted. We must be careful in our evaluation of expected squared error results, however, since this metric is strongly influenced by rare catastrophic events and as a result may overstate performance. Alternative distortion measures such as expected absolute error are currently under investigation.

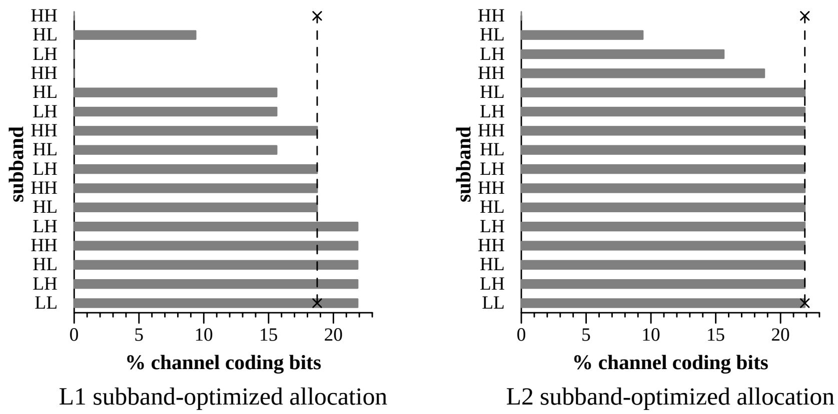


Figure 5: Channel code allocations for L^1 and L^2 optimization. Subbands are from a 5-stage wavelet decomposition and are ordered from coarsest (bottom) to finest (top). The vertical dashed lines indicate optimal uniform channel code levels.

6 ACKNOWLEDGMENTS

The authors would like to thank Kannan Ramchandran for valuable discussions and feedback. This research was supported in part by a National Science Foundation Mathematical Sciences Postdoctoral Research Fellowship, DMS-9627391, and ARPA as administered by the AFOSR under contracts AFOSR-90-0292 and DOD F4960-93-1-0567.

7 REFERENCES

[1] A. Albanese, J. Bloemer, J. Edmonds, M. Luby, and M. Sudan. Priority encoding transmission. In *Proc. Foundations of Computer Sciences*, pages 604–612, Santa Fe, NM, 1994.

[2] M. Antonini, M. Barlaud, and P. Mathieu. Image Coding Using Wavelet Transform. *IEEE Trans. Image Proc.*, 1(2):205–220, Apr. 1992.

[3] E. Biersack. Performance evalutaion of forward error correction in ATM networks. In *Proc. SIGCOMM*, pages 248–257, Baltimore, 1992.

[4] T. Cover and J. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., New York, 1991.

[5] C. R. Cunha, A. Bestavros, and M. E. Crovella. Characteristics of WWW client-based traces. Technical Report TR-95-010, Boston University Computer Science Department, 1995.

[6] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image Compression through Wavelet Transform Coding. *IEEE Trans. Info. Theory*, 38(2):719–746, Mar. 1992.

[7] N. Farvardin. A study of vector quantization for noisy channels. *IEEE Transactions on Information Theory*, 36(4):799–809, July 1990.

[8] M. W. Garrett and M. Vetterli. Joint source/channel coding of statistically multiplexed real-time services on packet networks. *IEEE Transactions on Networking*, 1(1):71–80, Feb. 1993.

[9] A. Gersho. Asymptotically optimal block quantization. *IEEE Transactions on Information Theory*, IT-25(4):373–380, July 1979.

[10] H. Gish and J. N. Pierce. Asymptotically efficient quantizing. *IEEE Transactions on Information Theory*, IT-14(5):676–683, Sept. 1968.

[11] B. Hochwald and K. Zeger. Tradeoff between source and channel coding. *preprint*, 1996.

[12] G. Karlsson and M. Vetterli. Subband coding of video for packet networks. *Optical Engineering*, 27(7):574–586, 1988.

[13] C. Leicher. Hierarchical encoding of mpeg sequences using priority encoding transmission (PET). Technical Report TR-94-058, ICSI, Berkeley, CA, Nov. 1994.

[14] A. S. Lewis and G. Knowles. Image compression using the 2-D wavelet transform. *IEEE Trans. on Image Processing*, 1(2):244–250, Apr. 1992.

[15] E. Posnak, S. Gallindo, A. Stephens, and H. Vin. Techniques for resilient transmission of jpeg video streams. In *Proceedings of Multimedia Computing and Networking, San Jose*, volume 2417, pages 243–252, Feb. 1995.

[16] F. P. Preparata. Holographic dispersal and recovery of information. *IEEE Transactions on Information Theory*, 35(5):1123–1124, Sept. 1989.

[17] M. O. Rabin. Efficient dispersal of information for security, load balancing, and fault tolerance. *Journal of the ACM*, 36(2):335–348, Apr. 1989.

[18] K. Ramchandran, A. Ortega, and M. Vetterli. Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders. *IEEE Transactions on Image Processing*, 3(5):533–545, Sept. 1994.

[19] J. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, Dec. 1993.

[20] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 36(9):1445–1453, Sept. 1988.

[21] N. Tanabe and N. Farvardin. Subband image coding using entropy-coded quantization over noisy channels. *IEEE Journal on Selected Areas in Communications*, 10(5):926–943, 1992.

[22] D. Taubman and A. Zakhor. Multirate 3-D subband coding of video. *IEEE Trans. Image Proc.*, 3(5), Sept. 1994.

[23] C. J. Turner and L. L. Peterson. Image transfer: and end-to-end design. In *Proc. SIGCOMM*, pages 258–268, 1992.

[24] J. B. Weaver, J. Dennis M. Healy, D. W. Warner, S. Chawla, and J. Lu. Mr imaging with a reduced number of encoding steps. In *Proc. SPIE: Medical Imaging 2710-79, Newport Beach*, Feb. 1996.

[25] P. L. Zador. Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Transactions on Information Theory*, IT-28(2):139–149, Mar. 1982.

[26] K. Zeger. *Source and channel coding with vector quantization.* PhD thesis, U.C. Santa Barbara, 1990.

[27] K. Zeger and A. Gersho. Pseudo-Gray coding. *IEEE Transactions on Communications*, 38(12):2147–2158, Dec. 1990.