Multiple Description Transform Coding of Images

Vivek K Goyal
U. of California, Berkeley
Berkeley, CA 94720-1772
v.goyal@ieee.org

Jelena Kovačević
Bell Laboratories
Murray Hill, NJ 07974
jelena@bell-labs.com

Ramon Arean
École Poly. Féd. de Lausanne
CH-1015 Lausanne Switzerland
arean@eurecom.fr

Martin Vetterli*
École Poly. Féd. de Lausanne
CH-1015 Lausanne Switzerland
Martin.Vetterli@de.epfl.ch

Abstract

Generalized multiple description coding (GMDC) is source coding for multiple channels such that a decoder which receives an arbitrary subset of the channels may produce a useful reconstruction. This paper reports on applications of two recently proposed methods for GMDC to image coding. The first produces statistically correlated streams such that lost streams can be estimated from the received data. The second uses quantized frame expansions and hence is conceptually similar to block channel coding, except it is done prior to quantization.

1 Introduction

Recently the problem of transmitting data over heterogeneous networks has received considerable attention. A typical scenario might require data to move from a fiber link to a wireless link, which necessitates dropping packets to accommodate the lower capacity of the latter. If the network is able to provide preferential treatment to some packets, then the use of a multiresolution or layered source coding system is the obvious solution. But what if the network will not look inside packets and discriminate? Then packets will be dropped at random, and it is not clear how the source (or source-channel) coding should be designed. If packet retransmission is not an option (*e.g.*, due to a delay constraint or lack of a feedback channel), one has to devise a way of getting meaningful information to the recipient despite the loss. The situation is similar if packets are lost due to transmission errors or congestion.

This problem is a generalization of the "multiple description" (MD) problem. In the MD problem, a source is described by two descriptions at rates R_1 and R_2. These two descriptions individually lead to reconstructions with distortions D_1 and D_2, respectively; and the two descriptions together yield a reconstruction with distortion D_0. The original problem was to characterize the achievable quintuples $(R_1, R_2, D_0, D_1, D_2)$. The first design algorithm for practical MD coding was given by Vaishampayan [1]. Vaishampayan's approach was based on scalar quantization with the MD property. Though MD scalar quantizers can be used in conjunction with transforms [2], the first successful attempts to use transforms to obtain the MD property were reported in [3, 4]. For background on the MD problem, see [1] and the references therein.

The aforementioned techniques are specific to the two-channel MD problem where all, half, or none of the bits make it to the decoder. When applied to packet communication with more than two packets, they fail to take full advantage of the "finer granularity." This provided motivation for the techniques for GMDC presented in [5, 6].

Our specific interest in this work is the communication of still images. The most common way to communicate an image over the internet is to use a progressive encoding system and to transmit the coded image as a sequence of packets over a TCP connection. When there are no packet losses, the receiver can reconstruct the image as the packets arrive; but when there is a packet loss, there is a large period of latency while the transmitter determines that the packet must be retransmitted and then retransmits the packet. The latency is due to the fact that the application at the receiving end uses the packets only after they have been put in the proper sequence. Changing to UDP does not solve the problem: because of the progressive nature of the encoding, the packets are useful only in the proper sequence. (The problem is more acute if there are stringent delay requirements, *e.g.*, for fast browsing or for streaming video. In this case retransmission is not just undesirable but impossible.) To combat this latency problem, it is desirable to have a communication system that is robust to arbitrarily placed packet erasures and that can reconstruct an image progressively from packets received in any order.

We approach the problem through the generalized multiple description framework. This paper reports preliminary image communication experiments using the methods of [5, 6]. The following two sections describe each technique and give simulation results.

2 Square Correlating Transforms

In this first method (see [5]), a block of n independent, zero-mean variables with different variances are transformed to a block of n transform coefficients in order to create a known statistical correlation between transform co-

*M. Vetterli is also with the Univ. of California, Berkeley.

0-8186-8821-1/98 $10.00

efficients. The transform coefficients from one block are distributed to different packets so in the case of a packet loss, the lost coefficients can be estimated from the received coefficients. The redundancy comes from the relative inefficiency of scalar entropy coding on correlated variables. This method is a generalization of the technique proposed in [3, 4] for two channels.

The coding of a source vector x proceeds as follows:

1. x is quantized with a uniform scalar quantizer with step size Δ: $x_{q_i} = [x_i]_\Delta$, where $[\cdot]_\Delta$ denotes rounding to the nearest multiple of Δ.
2. The vector $x_q = [x_{q_1}, x_{q_2}, \ldots x_{q_n}]^T$ is transformed with an invertible, discrete transform $\hat{T} : \Delta\mathbb{Z}^n \rightarrow \Delta\mathbb{Z}^n$, $y = \hat{T}(x_q)$.
3. The components of y are independently entropy coded.

The discrete transform $\hat{T}$ is related to a continuous transform T through "lifting." Starting with a linear transform T with determinant one, the first step in deriving a discrete version $\hat{T}$ is to factor T into a product of upper and lower triangular matrices with unit diagonals $T = T_1 T_2 \cdots T_k$. The discrete version of the transform is then given by

$$\hat{T}(x_q) = \left[T_1 \left[T_2 \ldots [T_k x_q]_\Delta\right]_\Delta\right]_\Delta . \qquad (1)$$

The lifting structure ensures that the inverse of $\hat{T}$ can be implemented by reversing the calculations in (1):

$$\hat{T}^{-1}(y) = \left[T_k^{-1} \ldots \left[T_2^{-1} \left[T_1^{-1} y\right]_\Delta\right]_\Delta\right]_\Delta .$$

When all the components of y are received, the reconstruction process is to (exactly) invert the transform $\hat{T}$ to get $\hat{x} = x_q$. The distortion is precisely the quantization error from Step 1. If some components of y are lost, they are estimated from the received components using the statistical correlation introduced by the transform $\hat{T}$. The estimate $\hat{x}$ is then generated by inverting T. The reader is referred to [5, 7] for the algebraic details.

The optimal design of the transform $\hat{T}$ for Gaussian sources, where arbitrary (unequal, dependent) packet loss probabilities are allowed, is discussed in [5]. Here we consider the simpler case where packet losses are i.i.d. and the transform is implemented as parallel and/or cascade combinations of 2-by-2 transforms. It is shown in [5] that for coding a two-tuple source over two channels, where each is equally like to fail, it is sufficient to consider transforms of the form

$$T_a \triangleq \begin{bmatrix} a & 1/(2a) \\ -a & 1/(2a) \end{bmatrix} . \qquad (2)$$

We use this as a building block to form larger transforms; *e.g.*, as shown in Fig. 1. The cascade structure simplifies the encoding, decoding, and design when compared to using a general $n \times n$ transform. Empirical evidence suggests that for $n = 4$ and considering up to one component erasure, there is no performance penalty in restricting consideration to cascade structures.

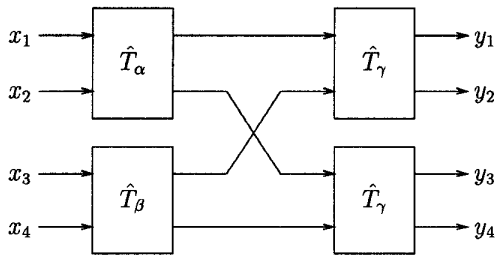

Figure 1: Cascade structure allows simple and efficient GMDC for more than two channels.

2.1 Application to images

To demonstrate the efficacy of the correlating transform method for image coding, we consider the case of coding for four channels. This method is designed to operate on source vectors with uncorrelated components. We (approximately) obtain such a condition by forming vectors from DCT coefficients separated both in frequency and in space. A straightforward application proceeds in the following steps:

1. An 8-by-8 block DCT of the image is computed.
2. The DCT coefficients are uniformly quantized.
3. Vectors of length 4 are formed from DCT coefficients separated in frequency and space. The spatial separation is maximized, *i.e.*, for 512×512 images, the samples that are grouped together are spaced by 256 pixels horizontally and/or vertically.
4. Correlating transforms are applied to each 4-tuple.
5. Entropy coding akin to that of JPEG is applied.

The system design is completed by determining which frequencies are to be grouped together and designing a transform (an (α, β, γ)-tuple for use as in Fig. 1) for each group. This can be done based on training data. Even with, say, a Gaussian model for the source data, the transform parameters must be numerically optimized.[1]

We have simulated an abstraction of this system. If we were to use precisely the strategy outlined above, the importance of the DC coefficient would dictate allocating most of the redundancy to the group containing the DC coefficient.

[1] In the case of pairing transforms as in [4], the optimal pairing and allocation of redundancy between the pairs can be found analytically [7].

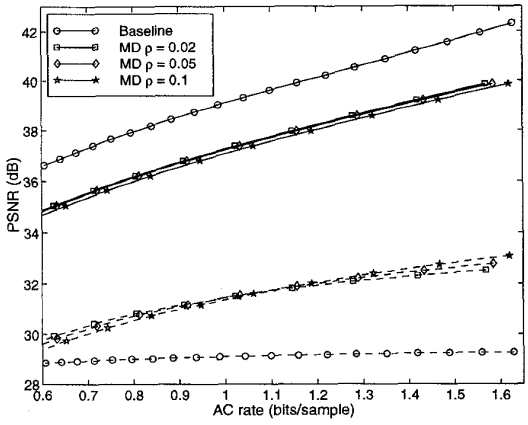

Figure 2: Rate-distortion results for the correlating transform method. Top (solid) and bottom (dashed) curves are for when all and three-fourths of the transmitted data arrives at the decoder, respectively.

Thus for simplicity we assume that the quantized DC coefficient is communicated reliably through some other means. We separate the remaining coefficients into those that are placed in groups of four and those that are sent by one of the four channels only. The optimal allocation of redundancy between groups is difficult, so we allocate approximately the same redundancy to each group. For comparison we consider a baseline system that also communicates the DC coefficient reliably. The AC coefficients for each block are sent over one of the four channels. The rate is estimated by sample scalar entropies.

Simulation results for the standard 512×512 'Lena' image are given in Fig. 2. The curve type indicates whether all (solid) or three-fourths (dashed) of the transmitted data arrives at the decoder. The marker type differentiates the MD and baseline systems. For the MD systems, the objective redundancy in bits per pixel (ρ from [4, 5]) is also given. As desired, the MD system gives a higher quality image when one of four packets is lost at the expense of worse rate-distortion performance when there are no packet losses. Size 128×128 sections of sample images are given in Fig. 3.

2.2 Comments

The results presented here are only preliminary because we have applied the techniques of [5] without much regard for the the structure and properties of images. The transform design is based on high-rate entropy estimates for uniformly quantized Gaussian random variables. Effects of coarse quantization, dead zone, divergence from Gaussianity, run length coding, and Huffman coding are neglected. Incorporating these will require a refinement of the theory

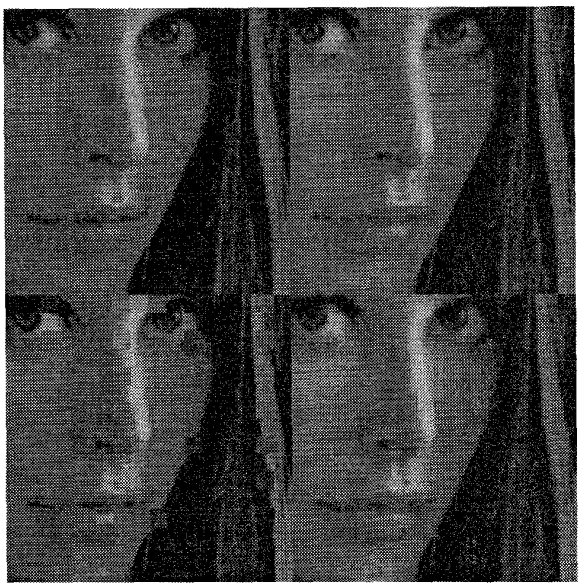

Figure 3: Results for correlating transform method at 1 bpp. Top row: no packet losses; bottom row: one packet lost. Left column: baseline system; right column: MD system.

and/or an expansive numerical optimization. Aside from transform optimization, this coder could be improved by using a perceptually tuned quantization matrix as suggested by the JPEG standard. Here we have used a constant quantization matrix for simplicity. With this type of tuning it should be possible to design a system which, say, performs precisely as well as the system in [3] when two or four of four packets arrive, but which performs better when one or three packets arrive.

A full image communication system would probably require packetization. We have not explicitly considered this, so we do not produce four streams with precisely the same number of bits. The expected number of bits for each stream is equal because of the form of (2). In contrast, with the transforms used in [4] one must multiplex the streams to produce packets of approximately the same size.

3 Overcomplete Frame Expansions

Robustness to lost packets comes from redundancy in the source representation. In the previous technique, the redundancy is *statistical*: the distribution of one part of the representation is reduced in variance by conditioning on another part. The second method that we consider (see [6]) uses a *deterministic* redundancy between descriptions.

Consider a discrete block code which represents k input symbols through a set of n output symbols such that any k of the n can be used to recover the original k. (For con-

creteness, this may be a systematic (n, k) Reed-Solomon code over $GF(2^m)$ with $n = 2^m - 1$ [8].) If the k input symbols are quantized transform coefficients, this may be a good way to communicate a k-dimensional source over an erasure channel that erases symbols with probability less than $(n - k)/n$. A problem with this approach is that except in the case that exactly k of n transmitted symbols are received, the channel has not been used efficiently. When more than k symbols are received, those in excess of k provide no information about the source vector; and when less than k symbols are received, it is computationally difficult to use more than just the systematic part of the code.

An alternative to (discrete) block coding was proposed in [6]. A linear transform from $\mathbb{R}^k$ to $\mathbb{R}^n$, followed by scalar quantization, is used to generate n descriptions of a k-dimensional source. These n descriptions are such that a good reconstruction can be computed from any k descriptions, but also descriptions beyond the kth are useful and reconstructions from less than k descriptions are easy to compute.

Assume that we have a tight frame $\Phi = \{\varphi_m\}_{k=1}^n \subset \mathbb{R}^k$ with $\|\varphi_m\| = 1$ for all m and that $y = Fx$, where F is the frame operator associated with Φ (see [9] and references therein for details on frames). This vector passes through the scalar quantizer Q: $\hat{y} = Q(y)$. The entropy-coded components of $\hat{y}$ can each be considered a description of x.

For simplicity, let us assume that Q is a uniform quantizer with step size Δ and that $n < 2k$. If $m \geq k$ of the components of $\hat{y}$ are known to the decoder, then x can be specified to within a cell with diameter approximately equal to Δ and thus is well approximated. Since the constraints on x provided by each description are independent, on average, the diameter is a nonincreasing function of m.

When $m < k$ components of $\hat{y}$ are received, $\mathbb{R}^k$ can be partioned into an m-dimensional subspace and a $(k - m)$-dimensional orthogonal subspace such that the component of x in the first subspace is well specified. With a mild zero-mean condition on the component in the latter space, a reasonable estimate of x is easily computed. For any m, estimating x can be posed as a simple least-squares problem.[2] An outline for an analysis using an additive white noise model for the quantization error is given in [6].

3.1 Application to images

As an example, we consider a frame alternative to a $(10, 8)$ block code. For the 10×8 frame operator F we use a matrix corresponding to a length-10 real Discrete Fourier Transform of a length-8 sequence [9]. This can be constructed as $F = [F^{(1)}\ F^{(2)}]$, where

$$F_{ij}^{(1)} = \frac{1}{2} \cos \frac{\pi(i-1)(2j-1)}{10} \quad \text{and}$$

$$F_{ij}^{(2)} = \frac{1}{2} \sin \frac{\pi(i-1)(2j-1)}{10}, \ 1 \leq i \leq 10, \ 1 \leq j \leq 4.$$

In order to profit from psychovisual tuning, we apply this technique to DCT coefficients and use quantization step sizes as in a typical JPEG coder. The coding proceeds as follows:

1. An 8-by-8 block DCT of the image is computed.
2. Vectors of length 8 are formed from DCT coefficients of like frequency, separated in space.
3. Each length 8 vector is expanded by left-multiplication with F.
4. Each length 10 vector is uniformly quantized with a step size depending on the frequency.

The baseline system against which we compare uses the same quantization step sizes, but quantizes the DCT coefficients directly and then applies a systematic $(10, 8)$ block code which can correct any two erasures. We assume that if there are more than two erasures, only the systematic part of the received data is used. (Maximum likelihood decoding would perform somewhat better, but is complex. In practice, one often discards the entire codeword if there are too many erasures.)

We have simulated the two systems with quantization step sizes conforming to a *quality* setting of 75 in the Independent JPEG Group's software.[3] For the 'Lena' image, this corresponds to a rate of about 0.98 bpp plus 25% channel coding. In order to avoid issues related to the propagation of errors in variable length codes, we consider an abstraction in which sets of coefficients are lost. The alternative would require explicitly forming ten entropy coded packets. The reconstruction for the frame method follows a least-squares strategy. For the baseline system, when eight or more of the ten descriptions arrive, the block code insures that the image is received at full fidelity. The effect of having less than eight packets received is simulated using the following combinatorial result: With $e > n - k$ erasures distributed uniformly in a systematic (n, k) code, the probability that m data symbols are erased is

$$\binom{n}{e}^{-1} \binom{k}{m} \binom{n-k}{e-m} \text{ for } e-(n-k) \leq m \leq \min(e, k).$$

Numerical results are shown in Fig. 4 for one through ten received packets. As expected, the frame system has better performance when less than eight packets are received. It is disappointing to see that the frame system did not have better performance when all ten packets are received, as was expected. Sample images are given in Fig. 5. (From the

[2] For $m \geq k$, a better estimate can be found by exploiting the boundedness of the quantization error [9].

[3] Version 6b of `cjpeg`. The current version is available at `ftp://ftp.uu.net/graphics/jpeg/`.

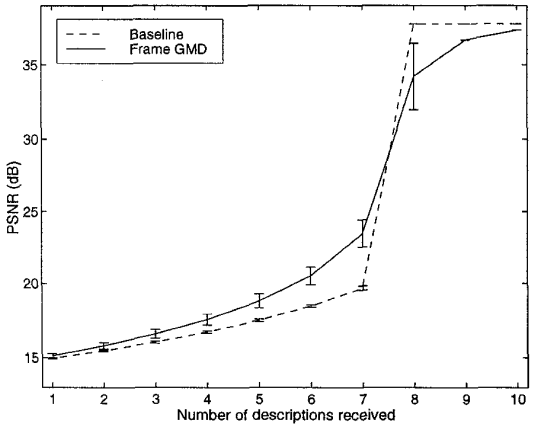

Figure 4: Numerical results for the overcomplete frame method at about 1 bpp. The mean performance is shown for each number of packets received along with the standard deviation.

512×512 images, 64×64 pixel detail images are shown.) Numerically and visually it is apparent that the performance of the MD system degrades gracefully as the number of lost packets increases.

References

[1] V. A. Vaishampayan. Design of multiple description scalar quantizers. *IEEE Trans. Inform. Th.*, 39(3):821–834, 1993.

[2] J.-C. Batllo and V. A. Vaishampayan. Asymptotic performance of multiple description transform codes. *IEEE Trans. Inform. Th.*, 43(2):703–707, 1997.

[3] Y. Wang, M. T. Orchard, and A. R. Reibman. Multiple description image coding for noisy channels by pairing transform coefficients. In *Proc. IEEE Workshop on Multimedia Sig. Proc.*, pp. 419–424, June 1997.

[4] M. T. Orchard, Y. Wang, V. Vaishampayan, and A. R. Reibman. Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms. In *Proc. IEEE Int. Conf. Image Proc.*, vol. I, pp. 608–611, October 1997.

[5] V. K Goyal and J. Kovačević. Optimal multiple description transform coding of Gaussian vectors. In *Proc. IEEE Data Compression Conf.*, pp. 388–397, March 1998.

[6] V. K Goyal, J. Kovačević, and M. Vetterli. Multiple description transform coding: Robustness to erasures using tight frame expansions. In *Proc. IEEE Int. Symp. Inform. Th.*, August 1998. To appear.

[7] V. K Goyal and J. Kovačević. Multiple description source-matched channel coding of a Gaussian source. *IEEE Trans. Inform. Th.*, 1998. In preparation.

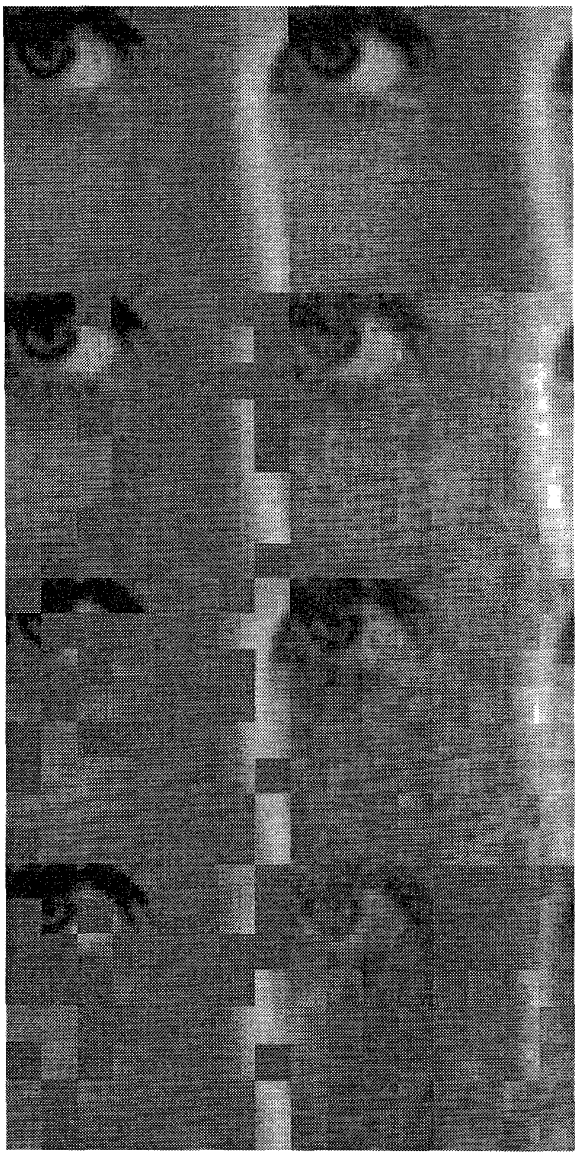

Figure 5: Results for frame method at 1 bpp. Left column: baseline system; right column: MD system. From top to bottom, number of packets received is 8, 7, 6, and 5.

[8] S. Lin and D. J. Costello. *Error Control Coding: Fundamentals and Applications.* Prentice-Hall, 1983.

[9] V. K Goyal, M. Vetterli, and N. T. Thao. Quantized overcomplete expansions in $\mathbb{R}^N$: Analysis, synthesis, and algorithms. *IEEE Trans. Inform. Th.*, 44(1):16–31, 1998.