Optimizing multiple description video coders in a packet loss environment

Amy Reibman
AT&T Labs –Research
Florham Park, NJ 07928
amy@research.att.com

March 24, 2002

Abstract

We consider the optimization of multiple description (MD) video coders for operation in a packet loss environment. Our primary goal in this paper is to address the question: in a packet loss environment, how much advantage can be gained from using multiple description coding instead of using an optimized one-layer encoding?

Zhang, Regunathan and Rose introduced the ROPE method to optimally choose the placement of intra-blocks in a one-layer encoder. We extend ROPE to MD video coders. MD-ROPE optimizes the placement of intra-blocks within the MD coder, in addition to optimally allocating MD redundancy to each block. As in ONE-ROPE, the distortion at the decoder is computed for every pixel and takes into account the effects of error propagation.

Applying MD-ROPE to a simple MD video coder, we show performance improvements at 10% packet loss compared to ONE-ROPE (ROPE applied to a one-layer coder). Further, the MD video coder has better performance across a range of packet loss rates as compared to the one-layer coder.

1 Introduction

Compressed video, which uses predictive coding algorithms and variable length coding, is sensitive to packet losses. Substantial degradation can occur if no action is taken to stop or limit the amount of error propagation. Motion compensation allows the error to propagate both temporally and spatially.

Recently, it has been recognized that the impact of errors in a one-layer coder can be probabilistically characterized at the encoder. This has lead to a number of schemes to select the placement of intra-blocks in a one-layer coder [1, 2, 3, 4]. However, one-layer coders typically only have very few source coding options available for trading off coding efficiency for greater robustness to errors. The primary parameters that can be changed in a one-layer coder are the intra-block rate and the slice (or Group of Block (GOB)) length. Multiple description coders however, have more flexibility in providing coding alternatives that increase robustness without significant increases in bit-rate.

The multiple description (MD) problem was introduced as an abstraction to an erasure (packet-loss) channel. MD coding is an encoder-based approach to partition a source into two (or more) bitstreams for transmission across different unreliable channels. The MD channel environment consists of two (or more) independent channels, such that for each channel data is either completely lost or is received intact. If both partitions (or descriptions) are received, a high-quality reconstruction can be decoded from both bitstreams, while a lower but still acceptable quality reconstruction can be decoded if either of the two bitstreams is lost. This is possible by introducing a structured correlation between the two descriptions, which inherently reduces compression efficiency. The fundamental trade-off of MD coding is between the increased redundancy (extra bit rate) and the improved single-description reconstruction quality.

There has been extensive work in the information theory community, including [5], on finding achievable regions for multiple descriptions. In addition, there has been a recent revival in the signal processing community on building compression systems that can approach these bounds (see [6, 7, 8] and the references in the extensive review [9]). MD video coders have also been presented in [10, 11, 12, 13, 14, 15, 16, 17]. In some papers, the MD coders are explicitly designed to be optimal in an ideal MD channel environment. In others, the term "multiple descriptions" is applied loosely to mean the addition of redundancy to help reconstruction given missing information.

A few papers consider MD coders optimized for packet loss environments. Vaishampayan [18] assign the rates applied to each image transform coefficient by an MD scalar quantizer, assuming the packet loss probability is known. Jiang and Ortega [19] design an MD image coder based on polyphase transform and selective quantization, where redundancy is optimally allocated based on the expected channel loss probability. Comas et. al. [16] extend this idea to an unbalanced MD video coder, where rate is allocated to low-resolution and high-resolution descriptions in an unbalanced way; however, error propagation is not considered.

In general, source adaptivity is the key to achieving such an optimization. The system must allow a sufficient number of options so that the encoding can be tailored to the specific nature of the source, both spatially and temporally. Ideally, it should not cost a lot of bits to signal the different coding options. The unbalanced MD video coder of Comas et. al. [16] achieves this objective. We presented a balanced MD coder based on similar principles in [17]. Both coders allow redundancy to be allocated optimally both spatially and temporally, without any additional rate to achieve the source adaptivity.

In this paper, we extend the one-layer ROPE (ONE-ROPE) framework [2] to the MD problem. In ROPE, the overall distortion of the decoder reconstruction due to quantization, error propagation, and error concealment is estimated for a one-layer video coder subject to packet losses. This overall distortion is computed recursively at the encoder using pixel-level precision, and accurately accounts for both spatial and temporal error propagation. Using the overall distortion, the best location for intra-blocks is found using a Rate-Distortion (RD) optimization. This method demonstrated significant gains over existing techniques at that time [2].

In Section 2, we review the ONE-ROPE optimization strategy presented in [2]. In Section 3, we extend this to the optimization of a generic MD video coder in a packet loss environment. In Section 4, we apply the optimization strategy of Section 3 specifically to our MD-split video coding algorithm. In Section 5 we examine the performance of MD-ROPE with the MD-split video coding algorithm to ONE-ROPE, in random packet loss environments. Our goal is to address the fundamental question: when can MD coding principles outperform one-layer techniques for video in a packet loss environment?

2 Review of one-layer ROPE

In this section, we review the ROPE method for optimizing the performance of a one-layer video coder in a packet loss environment [2]. ROPE considers optimal placement of intra-block (I-blocks) to optimize the rate-distortion performance while considering the impact of error propagation on the overall distortion.

We note that [2] assumed that every slice (or Group of Blocks (GOBs)) is sent in a separate packet. However, identical equations apply if instead we send two packets per frame, where each packet contains either the even or odd GOBs. Since two packets per frame provides a more efficient trade-off between packetization overhead and packet-loss performance [1], we consider the latter here.

The basic premise of ONE-ROPE is to choose the coding mode (ie, either I-block or P-block) on a macroblock basis to minimize the cost

$$\min_{\text{mode}} J_{MB} = \min_{\text{mode}} (D + \lambda R). \tag{1}$$

The distortion for a one-layer coder can be expressed as

$$D = (1-p)D_0 + p(1-p)D_1 + p^2 D_2 \tag{2}$$

where D_m is the distortion when m packets are lost. When computing the distortion within the context of determining which mode is best, the D_1 and D_2 terms may be dropped because they are the same for both modes. Also, the $(1-p)$ multiplier on the D_0 term may be ignored (with an appropriate adjustment to λ). It is therefore sufficient to consider only

$$D_{0,I} = \left(f_n^i - \hat{f}_n^i\right)^2 \tag{3}$$

and

$$D_{0,P} = \left(f_n^i - \hat{e}_n^i\right)^2 - 2\left(f_n^i - \hat{e}_n^i\right) E\left\{\tilde{f}_{n-1}^j\right\} + E\left\{\left(\tilde{f}_{n-1}^j\right)^2\right\}. \tag{4}$$

Here, j is the pixel corresponding to the actual motion vector, k is the pixel corresponding to the estimated concealment motion vector, and i is the pixel corresponding to the zero motion vector. Also, f_n^i is the original value of pixel i in frame n, and $\hat{f}_n^i$ is its encoder reconstruction. The decoder's reconstruction including all effects of packet loss, $\tilde{f}_n^i$, is a random variable. Finally, $\hat{e}_n^i$ is the encoder's reconstruction of the prediction error signal.

Computing $D_{0,I}$ is straightforward, but to compute $D_{0,P}$ the encoder needs to recursively compute the first and second moment of the decoder's reconstruction, $\tilde{f}_n^i$. If the block is compressed as an I-block, then these can be computed as

$$E\left\{\tilde{f}_n^i\right\} = (1-p)\hat{f}_n^i + p(1-p)E\left\{\tilde{f}_{n-1}^k\right\} + p^2 E\left\{\tilde{f}_{n-1}^i\right\} \tag{5}$$

and

$$E\left\{\left(\tilde{f}_n^i\right)^2\right\} = (1-p)\left(\hat{f}_n^i\right)^2 + p(1-p)E\left\{\left(\tilde{f}_{n-1}^k\right)^2\right\} + p^2 E\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\}. \tag{6}$$

It is important to consider clipping carefully; (5) is clipped between 0 and 255, and (6) is clipped to between 0 and 255^2.

If the block is compressed as a P-block, the first and second moments can be computed as [2]

$$E\left\{\tilde{f}_n^i\right\} = (1-p)\left(\hat{e}_n^i + E\left\{\tilde{f}_{n-1}^j\right\}\right) + p(1-p)E\left\{\tilde{f}_{n-1}^k\right\} + p^2 E\left\{\tilde{f}_{n-1}^i\right\} \tag{7}$$

and

$$E\left\{\left(\tilde{f}_n^i\right)^2\right\} = \tag{8}$$
$$(1-p)\left(\left(\hat{e}_n^i\right)^2 + 2\hat{e}_n^i E\left\{\tilde{f}_{n-1}^j\right\} + E\left\{\left(\tilde{f}_{n-1}^j\right)^2\right\}\right) + p(1-p)E\left\{\left(\tilde{f}_{n-1}^k\right)^2\right\} + p^2 E\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\}.$$

Again, clipping must be performed carefully, on the first term of (7) and the intermediate terms of (8). The starting point for the iterations are $E\left\{\tilde{f}_n^i\right\} = 128$ and $E\left\{\left(\tilde{f}_n^i\right)^2\right\} = 128^2$.

3 Extension of ROPE to Multiple Description

In a one-layer coder, there is little that can be done to improve the distortion at the decoder given packet losses. ROPE chooses optimally whether to use a P-block or an I-block, the choice of which affects only the D_0 term in (2) as discussed above. Error concealment techniques can improve the D_1 term in (2), but these are not at the discretion of the encoder and have only limited effect.

One approach to improve the quality when only one packet is received (beyond what is possible with error concealment), is to send some extra motion information in the adjacent packet to help improve the concealment. This, as a general approach, can be considered to be an unbalanced MD coder. In general, we consider in this section a broad definition of multiple description video coders that includes this option. We consider the case of two packets per frame, with extra rate applied to provide supplemental motion information and/or supplemental coefficient information. These will reduce the D_1 term, but at the cost of increased bit-rate. Therefore, the selection of the supplemental information should be factored into the optimization of (1).

Therefore, in this section we extend the method of ROPE [2] to the case of a generic multiple description video coder.[1] First, we present the computation of the distortion in the MD coder. Next, we present the recursion of the first and second moments of the decoder's reconstruction, and finally we present the optimization.

3.1 Distortion

We begin by further decomposing (2) into

$$D = (1-p)^2 D_0 + p(1-p) D_{1,2} + p(1-p) D_{1,1} + p^2 D_2, \tag{9}$$

where $D_{1,m}$ is the distortion when only one packet, packet m, is lost. (Note that this can easily be extended to the case where the packets from each description are subjected to different loss probabilities. However to simplify notation, we will not explicitly consider that situation here.)

D_0 is nominally the same for either a one-layer or MD coder, so equations (3) and (4) are applicable here as well. In the case of a one-layer coder, D_0 was all that was necessary to compute because the D_1 and D_2 were the same for both I- and P- blocks. However, for the MD case, D_1 differs for I- and P-blocks although D_2 can still be ignored for MD. Therefore, for optimization it is sufficient to consider only

$$D = (1-p)^2 D_0 + p(1-p) D_{1,2} + p(1-p) D_{1,1}.$$

For MD, we can compute $D_{1,m}$ for a I-block by

$$D_{1,m} = \left(f_n^i - \left(\hat{f}_n^i\right)_m\right)^2 \tag{10}$$

where $\left(\hat{f}_n^i\right)_m$ is the reconstruction from the m-th description only of the I-block. For a P-block, it is

$$D_{1,m} = \left(f_n^i - \left(\hat{e}_n^i\right)_m\right)^2 - 2\left(f_n^i - \left(\hat{e}_n^i\right)_m\right) E\left\{\tilde{f}_{n-1}^{j_m}\right\} + E\left\{\left(\tilde{f}_{n-1}^{j_m}\right)^2\right\} \tag{11}$$

---

[1] We currently assume the extra motion information is only sent for P-blocks. The I-blocks only have the possibility of sending supplemental coefficient information. An extension would allow motion information to be sent with I-blocks, to allow concealment to help supplement the MD coefficient information.

where $\left(\hat{e}_n^i\right)_m$ is the reconstruction from the m-th description only of the residual sent for the P-block, and j_m is the pixel corresponding to the reconstruction of the motion vector sent in description m. With this notation, we do not restrict ourselves to any particular MD technique applied to either motion vectors or coefficients. Note that (11) has the same basic functional form as (4).

This formulation allows a variety of methods to be used for sending motion vector information. For example, if we were to duplicate motion vectors, then $j_m = j$. Alternatively, we could apply existing MD techniques to a motion vector to generate two descriptions for a motion vector [15]. Finally, it would be possible to send the desired motion in one description, and for the other description send motion information that would help with the concealment process only.

3.2 Decoder reconstruction

We now consider the recursion to compute the first and second moments of the decoder's reconstruction.

For I-blocks, we assume here that the encoder does not send additional information (for example, motion vectors) that would help for concealment when only one packet is received. Therefore, for an I-block, we have

$$E\left\{\tilde{f}_n^i\right\} = (1-p)^2\hat{f}_n^i + p(1-p)\left(\left(\hat{f}_n^i\right)_1 + \left(\hat{f}_n^i\right)_2\right) + p^2E\left\{\tilde{f}_{n-1}^i\right\} \tag{12}$$

and

$$E\left\{\left(\tilde{f}_n^i\right)^2\right\} = (1-p)^2\left(\hat{f}_n^i\right)^2 + p(1-p)\left(\left(\hat{f}_n^i\right)_1^2 + \left(\hat{f}_n^i\right)_2^2\right) + p^2E\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\}. \tag{13}$$

For a P-block, we have

$$E\left\{\tilde{f}_n^i\right\} = (1-p)^2\left(\hat{e}_n^i + E\left\{\tilde{f}_{n-1}^j\right\}\right) + p(1-p)\left(\left(\hat{e}_n^i\right)_1 + E\left\{\tilde{f}_{n-1}^{j_1}\right\} + \left(\hat{e}_n^i\right)_2 + E\left\{\tilde{f}_{n-1}^{j_2}\right\}\right) + p^2E\left\{\tilde{f}_{n-1}^i\right\} \tag{14}$$

and

$$\begin{aligned} E\left\{\left(\tilde{f}_n^i\right)^2\right\} &= (1-p)^2\left(\left(\hat{e}_n^i\right)^2 + 2\hat{e}_n^iE\left\{\tilde{f}_{n-1}^j\right\} + E\left\{\left(\tilde{f}_{n-1}^j\right)^2\right\}\right) \\ &\quad + p(1-p)\left(\left(\hat{e}_n^i\right)_1^2 + 2\left(\hat{e}_n^i\right)_1E\left\{\tilde{f}_{n-1}^{j_1}\right\} + E\left\{\left(\tilde{f}_{n-1}^{j_1}\right)^2\right\}\right. \\ &\quad \left. + \left(\hat{e}_n^i\right)_2^2 + 2\left(\hat{e}_n^i\right)_2E\left\{\tilde{f}_{n-1}^{j_2}\right\} + E\left\{\left(\tilde{f}_{n-1}^{j_2}\right)^2\right\}\right) \\ &\quad + p^2E\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\} \end{aligned} \tag{15}$$

Again, clipping must be performed carefully.

It can be seen that the first and last terms of the recursions are the same (except for the initial multiplier) for both ONE-ROPE [2] and MD-ROPE. The middle term, which describes what happens when only one packet is lost, is different.

3.3 Optimization

Next, we extend the optimization in equation (2) to the case of a multiple description video coder. For MD-ROPE, we choose the mode and the MD method to minimize the cost,

$$\min_{\text{mode},\text{MD}_{\text{options}}} J_{MB} = \min_{\text{mode},\text{MD}_{\text{options}}} (D + \lambda R). \tag{16}$$

This can be broken down into

$$\min_{\text{mode}} \left[(1-p)D_0 + \lambda(R_{mv} + R_{\text{coeff}}) + \min_{\text{MD}_{\text{options}}} [(pD_{1,2} + pD_{1,1}) + \lambda(\rho_{mv} + \rho_{\text{coeff}})] \right], \tag{17}$$

where we have separated the MD-specific optimization from the mode-specific optimization. Note that the inner optimization is similar to the RRD optimization in [17], except that the distortions are computed here based on the expected decoder reconstructions instead of in [17] based on the side decoder reconstructions. Here, λ is the λ of [17] divided by $(1-p)$. Examples of MD_{options} are the number of coefficients in a low-resolution description (as in [16]), the possible quantizers in [19], or the threshold at which to switch from duplicating or alternating coefficients in [17].

4 MD-split video coder

In [17], we introduced MD-split, a video coder which creates MD bitstreams by splitting the coefficients output by a standard codec (ie, H.263). We determined the threshold at which coefficients should be duplicated or alternated based on the criterion of optimizing the redundancy-rate-distortion (RRD) cost function. The resulting two descriptions were individually compatible with the original standard. In this paper, we choose to use the same method to create multiple description bitstreams, but instead we use the criterion outlined in Section 3, to maximize performance in a packet loss environment.

With one exception [17], in MD-split we allow only simple alternation and duplication of the non-zero DCT coefficients produced by a traditional one-layer encoder. Larger, more important coefficients are worth the cost of duplicating, while smaller ones have less effect if lost and should be alternated. MD-split has the advantage that it allows the coding strategy to be adapted to the local signal statistics; more redundancy can be allocated to those parts of the signal that require it. Also, MD-split does not require the encoder to specifically signal to the decoder information regarding local adaptivity. We use the same syntax and Variable Length Code (VLC) tables as H.263.

The basic scheme is as follows. For each block in the frame, we consider different values of a threshold, such that non-zero coefficients larger than the threshold are duplicated into each description and coefficients smaller than the threshold will be alternated between the descriptions according to their magnitude. The choice of the threshold is made based on the cost in (17), where MD_{options} is the set of possible thresholds T_{split} at which to split. The threshold that minimizes (17) is chosen to create the final MD strategy for that block. Note that the best threshold is not needed by the decoder. Coefficients are sent with the H.263 zigzag scan order.

At the decoder, if both descriptions are received, it is a simple matter to merge the coefficients from the two bitstreams either into a single block of coefficients, or into a single bitstream that can be decoded by a standard H.263 decoder. In either case, the exact single-description video can be produced. If only one description is received, that bitstream can simply be decoded using the same method as a standard H.263 decoder.

An unbalanced MD video coder with very similar properties was also presented in [16]. There, each description was also compatible with H.263, where one description consisted of a select number of low-frequency coefficients while the other consisted of the complete set of coefficients.

5 Packet loss performance

We now compare the performance of our MD-split coder with MD-ROPE to that of a single-description (SD) H.263 video coder with ONE-ROPE in a packet loss scenario. We packetize each frame into two packets. In the case of MD, each packet corresponds to one description. In the case of SD, each packet contains either the odd or the even Group-of-Blocks of the frame, while both packets contain the picture header. For SD, if only one packet from a frame is received, we use motion-compensated concealment (using motion vectors from the above GOB) to fill in the missing GOBs. We consider random packet losses, where identical losses are injected into the MD and one-layer streams.

We adjust the quantizer step-size to keep the overall bit-rate fixed, using the typical TMN8 rate control. For MD, the sum of the bit-rates produced in each channel is used for the rate control algorithm. In addition, we increase the target buffer fullness value by 10%. The resulting output bit-rate summed across both channels typically is within about 1% of the target bit-rate. We use the 10-second CIF sequences *Silent Voice* and *Foreman* at 15 fps and 7.5 fps respectively coded at 384 kbps, and inject 25 different sequences of packet losses. Performance is presented as the average of these 25.

Figure 1 shows the performance for Silent at 384 kbps. and Figure 2 shows the performance for Foreman at the same rate. Shown are the best performances for MD-ROPE and for ONE-ROPE across the range of designed packet loss probabilities. For both sequences, as the packet loss rate increases beyond approximately 2%, the MD-ROPE encoding performs better than the ONE-ROPE encoding. At the packet loss rate of 10%, MD-ROPE performs 1 dB better than ONE-ROPE for Silent and more than 2 dB better than ONE-ROPE for Foreman.

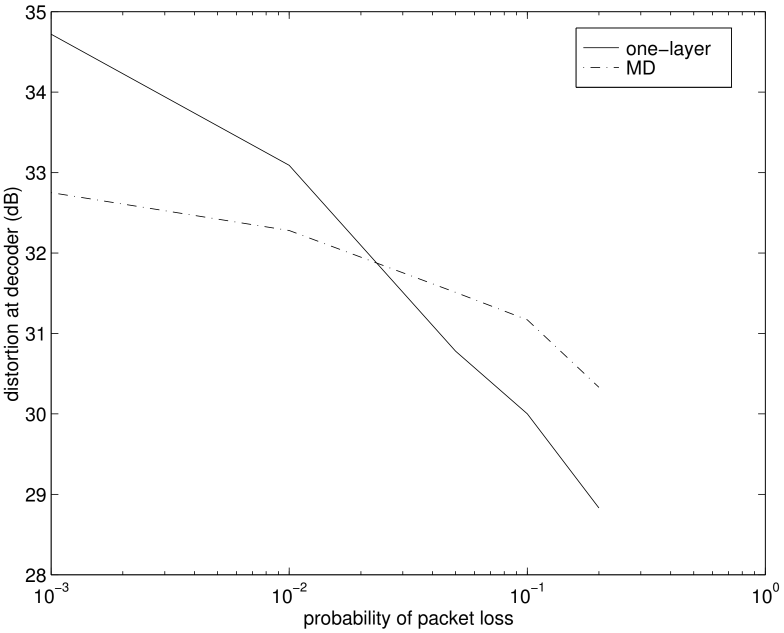

Figure 1: MD-ROPE vs. ONE-ROPE, Silent.

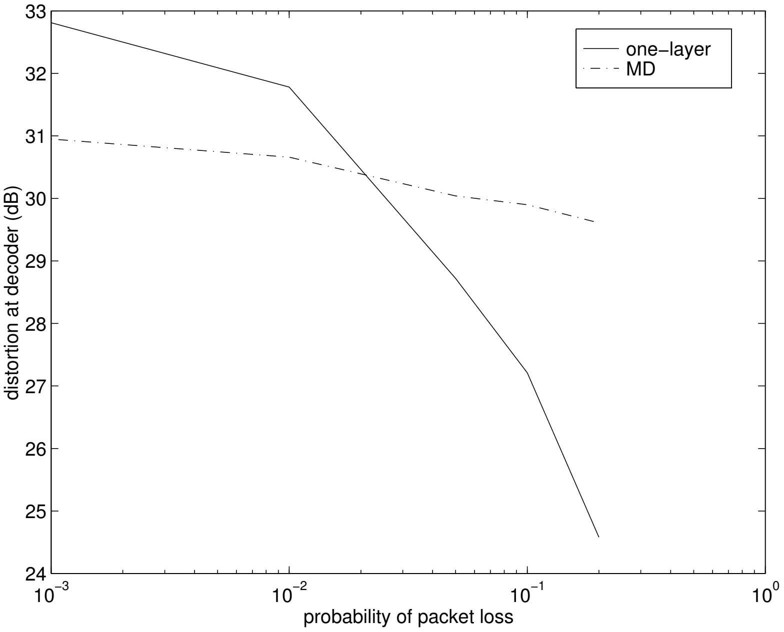

Figure 2: MD-ROPE vs. ONE-ROPE, Foreman.

References

[1] S. Wenger and G. Côté, "Using RFC2429 and H.263+ at low to medium bit-rates for low-latency applications", in *Packet Video Workshop '99*, New York, NY, April 1999.

[2] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal Inter/Intra-mode switching for packet loss resilience", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966-976, June 2000.

[3] D. Wu, Y. T. Hou, B. Li, W. Zhu, Y.-Q. Zhang, and H. J. Chao, "An end-to-end approach for optimal mode selection in Internet video communication: Theory and application", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 977-995, June 2000.

[4] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 952-965, June 2000.

[5] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.

[6] V. A. Vaishampayan, "Design of multiple description scalar quantizer," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.

[7] M. Orchard et al. "Redundancy Rate Distortion Analysis of Multiple Description Image Coding Using Pairwise Correlating Transforms," in *Proc. ICIP97,* (Santa Barbara, CA), Oct, 1997.

[8] V. K. Goyal and J. Kovacevic, "Optimal multiple description transform coding of Gaussian Vectors," *Proc. Data Compress. Conf.*, pp. 388–397, Mar. 1998.

[9] V. K. Goyal, "Multiple description coding: Compression meets the network", *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74-93, September 2001.

[10] W. S. Lee, M. R. Pickering, M. R. Frater, and J. F. Arnold, "A robust codec for transmission of very low bit-rate video over channels with bursty errors " *Circuits and Systems for Video Technology, IEEE Transactions on* Vol. 10, no. 8, pp. 1403-1412, Dec. 2000

[11] A. Reibman et al. "Multiple description coding for video using motion compensated prediction", in *Proc. IEEE Int. Conf. Image Proc,* (Kobe, Japan), October 1999.

[12] V. Vaishampayan and S. John, "Interframe Balanced Multiple Description Video Compression," in *Proc. Packet Video 1999*, (New York, NY), April 1999.

[13] V. Vaishampayan and S. John, "Balanced Interframe Multiple Description Video Compression," in *Proc. IEEE Int. Conf. Image Proc,* (Kobe, Japan), Oct. 1999.

[14] J. G. Apostolopoulos, "Error-resilient video compression through the use of multiple states", in *Proc. IEEE Int. Conf. Image Proc,* (Vancouver, CA), Sept. 2000.

[15] C.-S. Kim and S.-U. Lee, "Multiple description motion coding algorithm for robust video transmission", in *IEEE Int. Symp. on Circuits and Syst.*, (Geneva, Switzerland), May 2000.

[16] D. Comas, R. Singh, and A. Ortega, "Rate-distortion optimization in a robust video transmission based on unbalanced multiple description coding", IEEE Workshop on Multimedia Signal Processing, October 2001.

[17] A. R. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, "Multiple description video using rate-distortion splitting", *ICIP* 2001, Thessaloniki, Greece, October 2001.

[18] V. Vaishampayan, “Application of multiple description codes to image and video transmission over lossy networks”, *Packet Video Workshop*, Brisbane, Australia, March 1996.

[19] W. Jiang and A. Ortega, “Multiple description coding via polyphase transform and selective quantization”, *Proc. of VCIP*,San Jose, CA, January, 1999.