EFFICIENT PREDICTION IN MULTIPLE DESCRIPTION VIDEO CODING

Shankar L. Regunathan and Kenneth Rose

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

Abstract

This work is concerned with the design of multiple description (MD) compression systems with emphasis on video coding. Although inter-frame prediction is critical to the performance of video coders, the problem of efficient prediction in MD systems has not been satisfactorily resolved. We focus on quantizer based MD coding and propose an estimation theoretic (ET) approach to prediction and reconstruction. The advantage of ET prediction is two fold: (1) It takes into account all the information available at each decoder for an optimal estimate. (2) It mitigates the degradation due to quantization in the prediction feedback loop. The ET approach is first shown to achieve significant gains in the simpler setting of predictive MD coding of synthetic Gauss-Markov and Laplace-Markov processes. We then present performance results for MD compression of video sequences and demonstrate consistent and substantial PSNR gains over conventional techniques.

1. INTRODUCTION

Multiple description (MD) coding has been proposed for robust transmission over packet-switched networks. The basic idea in MD coding is to encode the source into two (or possibly more) bit-streams that are transmitted in separate packets. The packets are viewed as channels with independent probability of failure. An acceptable but coarse quality of reconstruction is obtained at the "side" decoders where only one of the channels (bit-streams) is available, while better reconstruction is achieved at the "central" decoder that receives both bit-streams. MD coding achieves packet loss resilience by trading increased central distortion for lower side distortion. Multiple description coding methods were developed in [1], [2], [3], and the framework was extended to video compression in [4], [5].

The main challenge in developing MD video coders is associated with the use of prediction [4], [5]. Standard single description video coders use motion compensated prediction to exploit inter-frame redundancy efficiently at low delay and implementation complexity. However, the incorporation of prediction within the MD framework is complicated by the existence of multiple candidate predictors. For the central decoder, the best predictor is formed from information available from both the bit-streams. However, only partial information is available at the side decoders which can form two different predictors from information received via their respective channels. While a good predictor should aim to exploit information available from both bit-streams, prediction based on information unavailable at either of the side decoders may cause "drift" and compromise compression efficiency.

Note that a similar "prediction difficulty" is present in scalable video coding (a special case of multiple description). An estimation theoretic approach to scalable video coding was proposed in [6] [7], and was shown to achieve significant improvements in compression performance. In this paper, the ET prediction framework is extended to MD coding. While our framework is applicable to any form of MD coding, we focus on the method developed in [4] where multiple description quantizers are used to encode the prediction error. We derive an estimation theoretic approach (ET) that uses all the information available to each decoder in producing its prediction. We also present simulation results for MD coding of first-order Markov process and video sequences. They demonstrate that ET prediction can achieve significant performance gains for MD video coding.

The organization of the paper is as follows: Section 2 describes the MD quantizer approach to video coding and covers the limitations of conventional prediction. In Section 3, we derive the estimation theoretic (ET) approach to prediction in multiple description coding. Section 4 presents simulation results for MD coding.

2. PREDICTION IN MD VIDEO CODING

We present the conventional approach to prediction in MD coding in this section (see [4] for more detail). Let x_n be the current source sample. Let its reconstruction by the central

---

This work is supported in part by the NSF under grant MIP-9707764, the University of California MICRO Program, Conexant Systems, Inc., Fujitsu Laboratories of America, Inc., Lernout & Hauspie Speech Products, Lucent Technologies, Inc., and Qualcomm, Inc.

0-7803-6297-7/00/$10.00 © 2000 IEEE

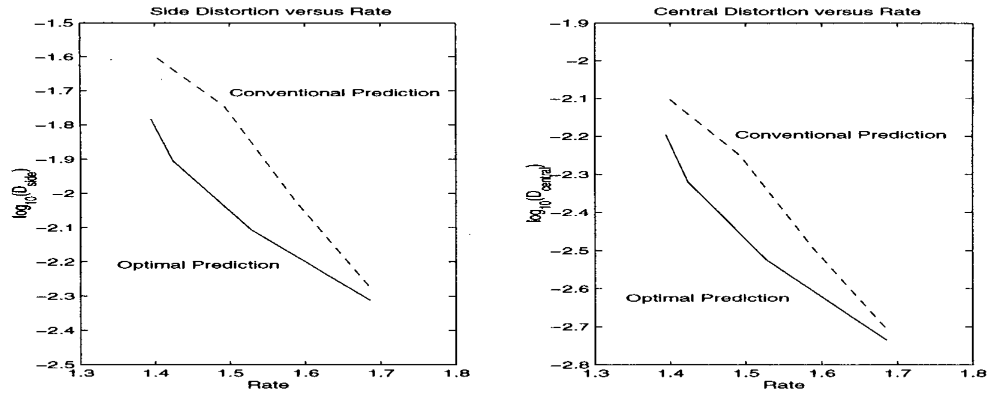

Fig. 1. Performance of multiple description coding for a Gauss-Markov source with $\rho = 0.99$.

decoder be denoted by $\hat{x}_{0,n}$, and let $\hat{x}_{1,n}$ and $\hat{x}_{2,n}$ be the respective reconstruction at the side decoders. We assume that the source is characterized by a first-order Markov process,

$$x_n = \rho x_{n-1} + z_n,$$

where ρ is the correlation coefficient, and z_n is zero-mean, white, stationary, and independent of $x_{n-i}, \forall i > 0$.

Ideally, we would choose the prediction for the side coders as given by $\tilde{x}_{1,n} = \rho\hat{x}_{1,n-1}$ and $\tilde{x}_{2,n} = \rho\hat{x}_{2,n-1}$ respectively. However, this complicates the use of an MD quantizer for encoding the resulting prediction error. Note that using different predictions for the two side channels implies that the encoding intervals of multiple description quantizers will be shifted, and out of alignment with respect to x_n. The misalignment will degrade the performance of the MD quantizer.

This problem was addressed in [4], where additional quantization was included *in the prediction feedback loop*, i.e., the actual prediction at the encoder was a quantized value: $q(\tilde{x}_{i,n}), i = 1, 2$. The feedback quantizers ensure that the predictors for the side channels are often identical. By making the step size of the feedback quantizer sufficiently large, the probability of misalignment can be made arbitrarily small.

The prediction errors (residuals) at the side channels, $r_{i,n} = x_n - q(\tilde{x}_{i,n})$, are encoded by a multiple description quantizer. The index of the first quantizer is sent over channel-1, and the second quantizer index over channel 2. The two indices are individually used by the respective side decoders to obtain the corresponding quantized residual. The central decoder uses the indices jointly to obtain a better reconstruction of the prediction error. The decoded residual is added back to the prediction to reconstruct the source.

This approach to prediction in MD coders has two major drawbacks. First, there is a degradation in the prediction gain of the side coders due to predictor quantization. In fact, this degradation can be very large for the important case of MD video coding. Further, note that the central decoder's reconstruction of the previous sample, $\hat{x}_{0,n-1}$, is in general better than the corresponding reconstruction of the side decoders, $\hat{x}_{1,n-1}$ and $\hat{x}_{2,n-1}$. However, the information contained in $\hat{x}_{0,n-1}$ is not directly used in coding and reconstruction of the current sample. We address these issues in the next section by developing an estimation theoretic approach to prediction in MD coding.

3. ESTIMATION-THEORETIC PREDICTION

We reformulate the prediction problem at each decoder as one of optimal estimation of the current sample given all the information available at that decoder. Let the encoding operation remain unchanged, i.e., form the quantized predictors $q(\tilde{x}_{i,n})$ and the respective residuals $r_{i,n}$ as before, and encode the residuals using an MD quantizer. Let (a_i, b_i) be the quantization intervals associated with the two indices, i.e., $r_{i,n} \in (a_i, b_i), i = 1, 2$.

Clearly, the statement $x_n \in (q(\tilde{x}_{i,n}) + a_i, q(\tilde{x}_{i,n}) + b_i)$ captures *all the information* on x_n provided to decoder i by the received index. Therefore the optimal reconstruction of x_n at side decoder i is given by:

$$\begin{aligned}\hat{x}_{i,n} &= E[x_n | x_n \in (q(\tilde{x}_{i,n}) + a_i, q(\tilde{x}_{i,n}) + b_i), \hat{x}_{i,n-1}] \\ &= \tilde{x}_{i,n} + E[z_n | z_n \in (a_i + q(\tilde{x}_{i,n}) - \tilde{x}_{i,n}, b_i + q(\tilde{x}_{i,n}) - \tilde{x}_{i,n}\end{aligned}$$

This expectation is computed as the centroid of the interval $(a_i + q(\tilde{x}_{i,n}) - \tilde{x}_{i,n}, b_i + q(\tilde{x}_{i,n}) - \tilde{x}_{i,n})$ with respect to the density $p(z_n)$. Note that the estimate takes advantage of the *unquantized* prediction, $\tilde{x}_{i,n}$, available at the decoder, and thus mitigates the effects of quantization in the prediction loop.

Next, we turn our attention to the central decoder. Recall that the previous sample reconstruction at the central decoder is $\hat{x}_{0,n-1}$. The information provided by the two quantizer indices can be summarized as: $x_n \in (c, d)$, where

$$\begin{aligned} c &= \max[q(\tilde{x}_{1,n}) + a_1, q(\tilde{x}_{2,n}) + a_2], \\ d &= \min[q(\tilde{x}_{1,n}) + b_1, q(\tilde{x}_{2,n}) + b_2]. \end{aligned}$$

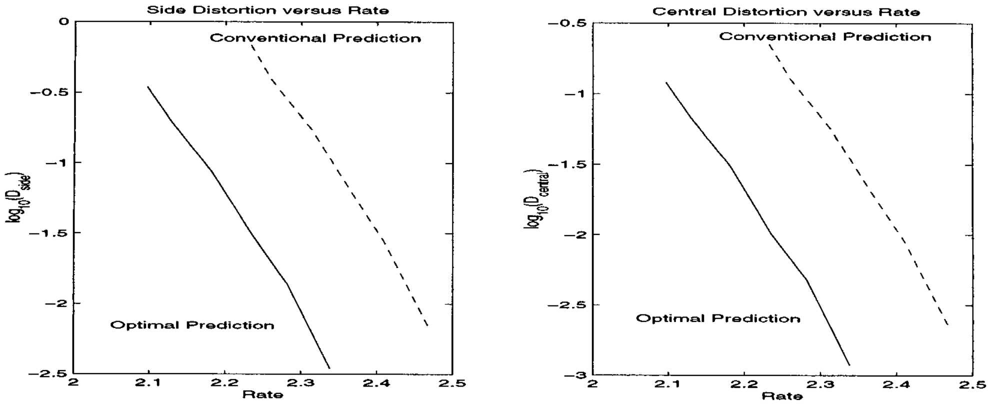

Fig. 2. Performance of multiple description coding for a Laplace-Markov source with $\rho = 0.95$.

Thus the optimal estimate of the current sample at the central decoder is given by:

$$\begin{aligned}\hat{x}_{0,n} &= E[x_n|x_n \in (c,d), \hat{x}_{0,n-1}] \\ &= \rho\hat{x}_{0,n-1} + E[z_n|z_n \in (c-\rho\hat{x}_{0,n-1}, d-\rho\hat{x}_{0,n-1})].\end{aligned}$$

The expectation is calculated as the centroid of the interval $(c-\rho\hat{x}_{0,n-1}, d-\rho\hat{x}_{0,n-1})$ with respect to the density $p(z_n)$. Thus, this estimate takes into account all the information available to the central decoder for its reconstruction of the current sample. In particular, this estimate takes advantage of the fact that the central decoder's reconstruction of the previous sample, $\hat{x}_{0,n-1}$, is better than those of the side decoders, $\hat{x}_{1,n-1}$ and $\hat{x}_{2,n-1}$.

4. RESULTS

We illustrate the power of ET prediction by applying it within the simple setting of multiple description coding for first order Gauss-Markov and Laplace-Markov signals. We then present simulation results for MD coding of video.

4.1. Markov sequences

Figures 1 and 2 compare the performance of the proposed ET prediction with conventional prediction for MD coding of Markov signals. We used the simplest multiple description quantizer which consists of two uniform offset quantizers with the same step size. Note that this is equivalent to a central quantizer with half the step size. The step sizes for the quantizers in prediction feedback are chosen as in [4]. The rate is calculated as the entropy of the quantizer indices. ET prediction achieves significant reduction in "side" distortion as well as "central distortion". Thus ET prediction can improve both the reconstruction of the side decoder by mitigating the effects of prediction quantization, and that of the central decoder by taking into account all the information available. Note that higher gains are achieved for the Laplace-Markov source which has been shown to be a more appropriate model for DCT coefficients in video sequences [7].

4.2. Simulation Results for MD Video Coding

Rate (kbps)	Conv. Pred		ET Pred	
	PSNR central	PSNR side	PSNR central	PSNR side
16	27.86	27.21	29.19	28.39
32	30.81	30.11	31.80	30.69
64	34.28	33.35	35.05	33.62

Table 1. Performance comparison of prediction methods for MD coding of the sequence "Carphone". Reconstructed PSNR (in dB) versus rate (kbps/channel)

Rate (kbps)	Conv. Pred		ET Pred	
	PSNR central	PSNR side	PSNR central	PSNR side
16	30.68	30.33	31.89	30.89
32	33.49	32.89	34.96	33.23
64	37.48	36.67	38.62	36.63

Table 2. Performance comparison of prediction methods for MD coding of the sequence "Salesman". Reconstructed PSNR (in dB) versus rate (kbps/channel)

We developed a test bed for MD video coding by using the publicly available H.263 coder [8]. The macroblocks in each frame could be encoded either in intra-mode or in inter-mode. For simplicity, the intra-mode coding of blocks is exactly as in H.263, and the same information is transmitted in both channels (packets). In inter-mode, the motion estimation/compensation is identical to H.263. The corresponding (after motion compensation) blocks in the previ-

Sequence (kbps)	Conv. Pred		ET Pred	
	PSNR central	PSNR side	PSNR central	PSNR side
Susie	34.03	33.26	34.62	33.61
Grandma	36.24	35.62	37.49	35.91
MTDT	33.79	33.18	34.84	33.48

Table 3. Performance comparison of prediction methods for MD coding of different sequences. at the rate of 32 Kbps per channel.

ous reconstructed frame for each of the decoders is transformed (by DCT) to obtain the predictors.

The predictors (DCT coefficients) are quantized by prediction feedback quantizer, and the prediction error is obtained in the DCT domain. Note that implementation of the prediction in the DCT domain will produce no change in the performance of conventional prediction. For ET prediction, DCT domain is more convenient because the DCT coefficients of the residual are almost uncorrelated. Further, the quantizer interval of each DCT coefficient is readily available. Thus, the ET predictor can be independently implemented for each DCT coefficient with virtually no loss of optimality.

The prediction errors are quantized by the simplest MD quantizer consisting of two offset uniform quantizers with the same step size. The quantizer indices of each block are scanned in a zig-zag manner, and entropy coded using the run-length and Huffman tables of H.263. The motion vector information and macroblock header information is repeated in both channels. The Laplace-Markov process was found to be a good model for the DCT coefficients [7]. The model parameters were estimated from a training set extracted from the *Miss America* sequence.

The average PSNR of the luminance component of the decoder reconstruction is used as the performance metric. Tables 1 and 2 show the performance results for MD coding of sequences *Carphone* and *Salesman* at various bit rates. Table 3 shows the performance results for MD coding of several video sequences at the rate of 32kbps per channel. It is easily seen that proposed ET prediction outperforms all the competing approaches, and achieves substantial gains in reconstructed PSNR at both the side and central decoders.

5. CONCLUSION

This paper presents a new estimation theoretic approach to prediction in MD coding of video. The reconstruction of the current sample at each decoder is optimal given all the information available at that decoder. In particular, the ET predictor was shown to mitigate the effect of prediction quantization at the side decoders, and take advantage of the better reconstruction of the previous sample available at the central decoder. Simulation results show that the proposed technique offers substantial gains in performance over conventional prediction for MD coding of both Markov sources and video sequences. Although ET prediction was applied in conjunction with standard DCT based (H.263) coding systems, it is easily extendible to sub-band based, and pixel-domain coders.

6. REFERENCES

[1] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. on Information Theory*, vol. 39, pp. 821-834, May 1993.

[2] Y. Wang, M. Orchard and A. R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *Proc. IEEE MMSP*, June 1997.

[3] V. K. Goyal, J. Kovacevic and M. Vetterli, "Multiple description transform: Robustness to erasures using tight frame expansions," in *Proc. IEEE ISIT*, p. 408, August 1998.

[4] V. A. Vaishampayan and S. John,"Balanced interframe multiple description video compression," in *Proc. IEEE ICIP*, Oct. 1999.

[5] A. Reibman, H. Jafarkhani,Y. Wang, M. T. Orchard and R. Puri, "Multiple description coding for video using motion compensated prediction," in *Proc. IEEE ICIP*, Oct. 1999.

[6] K. Rose and S. L. Regunathan, "Towards Optimal Scalability in Predictive video coding," *Proc. ICIP*, Oct. 1998.

[7] K. Rose and S. L. Regunathan, "Towards Optimality in Scalable Predictive Coding," Submitted to *IEEE Trans. on Image Processing*.

[8] TMN (H.263) coder version 2.0, Telenor R&D Norway, 1996.