WYZE-PMD BASED MULTIPLE DESCRIPTION VIDEO CODEC

A. Jagmohan, A. Sehgal, N. Ahuja

University of Illinois, Urbana-Champaign
{jagmohan, asehgal, n-ahuja}@uiuc.edu

ABSTRACT

The main hindrance to the development of efficient low-latency multiple description (MD) video coders is the problem of predictive mismatch. In this paper, we present a two-channel predictive MD video codec architecture based on the recently proposed WYZE-PMD framework. The proposed codec transmits coset information to curtail error-propagation caused by predictive mismatch, without requiring high latency or restrictive channel assumptions. MD scalar quantizers are used to generate multiple descriptions, low-density parity check (LDPC) codes are used to generate coset information, and the H.263 video coding standard is used for efficient motion compensation. The proposed codec is used to code descriptions of CIF video for communication over two erasure channels with independent failure probabilities. Results indicate that the proposed codec provides efficient, drift-free predictive MD coding.

1. INTRODUCTION

Multiple Description (MD) coding provides a robust communication methodology for communication applications where the alternative error-protection techniques of forward-error correction (FEC) and selective retransmission are ill-suited, due to low-delay constraints and the absence of feedback. MD coding of predictively encoded streams (termed predictive MD coding) is of practical interest in low-latency multimedia applications that involve communication of compressed video/audio data over error-prone channels. Such applications include real-time video streaming, broadcast of multimedia data and video-conferencing, over packet networks, multimedia communication over frequency-hopping wireless systems, and robust, distributed storage of multimedia data [1].

The key problem encountered in predictive MD coding systems is that of predictive mismatch. Predictive mismatch refers to a scenario in which there is a mismatch between the predictor symbols at the encoder and the decoder. In the context of MD coding, this mismatch arises because the subset of predictor symbol descriptions received at the decoder is unknown at the encoder. Previous approaches for predictive MD coding [2, 3, 4] impose strong conditions on the channel failure probabilities to eliminate predictive mismatch. Specifically, it is assumed that the subset of channels received over time remains fixed, which is rarely satisfied in multimedia communication applications. Other techniques avoid these assumptions at the expense of sacrificing low-latency ([2, 5]).

In this paper, we present a two-channel predictive MD video codec design, in which predictive mismatch is avoided without requiring restrictive assumptions or high latency. The presented video codec is based on the recently proposed WYZE-PMD framework [6], which models the predictive MD coding problem as a variant of the Wyner-Ziv decoder side-information problem [7], and eliminates predictive mismatch by transmitting appropriate coset information. The proposed predictive MD video codec uses MD scalar quantizers [8] to generate descriptions, low-density parity check (LDPC) codes [9] to generate coset information, and the H.263 video coding standard [10] for efficient motion compensation. To illustrate the efficacy of the proposed video codec, we consider MD coding of CIF video for communication over two erasure channels with independent failure probabilities. The proposed codec curtails error propagation caused by predictive mismatch and outperforms conventional approaches in terms of rate-distortion performance.

2. THE WYZE-PMD FRAMEWORK

In this section, we provide a brief description of the WYZE-PMD framework, for the specific application of two-channel predictive MD coding. We illustrate the principles involved by briefly describing code constructions based on MD scalar quantizers and turbo codes for this application. Further details are available in [6].

2.1. Problem Formulation

Consider the communication of a M-dimensional source with memory, $\{\mathbf{V}_i\}_{i=1}^{\infty}, \mathbf{V}_i \in \mathbb{R}^M$, across a lossy channel using one-step predictive coding. Given the decoder reconstruction of source symbol $\mathbf{V}_{k-1}$ (denoted $\widehat{\mathbf{V}}_{k-1}$) the encoder communicates $\mathbf{V}_k$ by generating the innovation $\mathbf{T}_k = \mathbf{V}_k - E[\mathbf{V}_k|\widehat{\mathbf{V}}_{k-1}]$ which is input to the channel, where $E[\cdot]$ represents the expectation operator. For the case of two-channel MD coding, the decoder reconstruction $\widehat{\mathbf{V}}_{k-1}$ can take one of multiple values, depending on which of the 2^2 possible subsets of descriptions of $\mathbf{V}_{k-1}$ is received i.e. $\widehat{\mathbf{V}}_{k-1} \in \mathcal{R}_{k-1}$ where $\mathcal{R}_i = \{\widehat{\mathbf{V}}_i^j\}$ denotes the reconstruction set for the i^{th} symbol. The number of possible predictors grows exponentially with time and, in general, $|\mathcal{R}_k| = 2^{2k}$.

The problem of predictive MD coding can be formulated as a variant of the WZ decoder side-information problem as follows. The decoder reconstruction of the predictor $\widehat{\mathbf{V}}_{k-1}$ takes values in the reconstruction set $\mathcal{R}_{k-1} = \{\widehat{\mathbf{V}}_{k-1}^j\}$ with a probability mass function determined by the channel failure probabilities, $P(\widehat{\mathbf{V}}_{k-1} = \widehat{\mathbf{V}}_{k-1}^j) = q(j), j \in \{1, \ldots, |\mathcal{R}_{k-1}|\}. \sum_j q(j) = 1$. Thus, the encoder is required to compress $\mathbf{V}_k$ in the presence of the correlated decoder side-information $\widehat{\mathbf{V}}_{k-1}$, when the only information the encoder has about $\widehat{\mathbf{V}}_{k-1}$ are it's statistics, i.e. $\mathcal{R}_{k-1}$ and $q(\cdot)$.

2.2. Code Construction

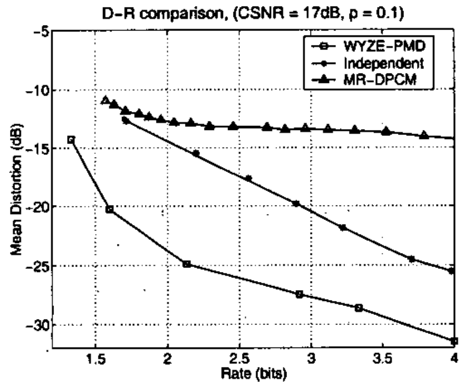

Fig. 1. D-R performance comparison of WYZE-PMD, independent coding, and MR-DPCM for first-order Gauss-Markov source, $p_1 = p_2 = 0.1$.

For two-channel predictive MD coding, the WYZE-PMD encoder consists of two lossy MD encoding functions $\mathbf{m}_i : \mathbb{R}^M \to \mathcal{I}_{2^{mR'_i}}$, $i = 1, 2$ and the respective coset coding functions $\mathbf{c}_i : \mathcal{I}_{2^{mR'_i}} \to \mathcal{I}_{2^{mR_i}}$, where $\mathcal{I}_n = \{1, \dots, n\}$. The encoder transmits the innovation $\mathbf{T}^i_k = \mathbf{c}_i(\mathbf{m}_i(\mathbf{V}_k))$ on the i-th channel. The WYZE-PMD decoder consists of two coset decoding functions, which recover the transmitted MD index based on the side-information and the received coset index, $\widehat{\mathbf{c}}_i : \mathcal{I}_{2^{mR_i}} \times \mathbb{R}^M \to \mathcal{I}_{2^{mR'_i}}$, and four MD decoding functions—the side decoding functions $\mathbf{g}_{i\in\{1,2\}} : \mathcal{I}_{2^{mR'_i}} \times \mathbb{R}^M \to \mathbb{R}^M$, the central decoder $\mathbf{g}_0 : \mathcal{I}_{2^{mR'_1}} \times \mathcal{I}_{2^{mR'_2}} \times \mathbb{R}^M \to \mathbb{R}^M$ and the null decoder $\mathbf{g}_{null} : \mathbb{R}^M \to \mathbb{R}^M$.

For the sake of concreteness, we briefly describe a practical two-channel WYZE-PMD code construction based on MD scalar quantizers and turbo codes (details are available in [6]). For a given source vector $\mathbf{V}_k$, MD scalar quantization is used to generate two index vectors, $\mathbf{m}_i : \mathbb{R}^M \to \mathcal{I}^m_{2^{r_i}}$ (where the i-th side MD quantizer is $\mathbf{q}_i : \mathbb{R} \to \mathcal{I}_{2^{r_i}}$). Each index vector $\mathbf{m}_i(\mathbf{V}_k)$ is encoded using a turbo encoder consisting of two systematic convolutional encoders. The resultant parity bitstream, punctured to a rate of mR_i bits, represents the coset information for the MD index vector, and is transmitted on Channel i. Decoding is performed as follows. For each received channel i, the decoder recovers the MD index vector by iterative decoding on the basis of the received coset information and the decoder predictor reconstruction $\widehat{\mathbf{V}}_{k-1}$ which serves as side-information. The appropriate MD decoding function is invoked to form an MMSE estimate of the transmitted source-vector on the basis of the recovered MD indices and the predictor $\widehat{\mathbf{V}}_{k-1}$.

The key design issue is the choice of the coset transmission rate, which should be high enough such that the probability of turbo decoder failure is negligible. In [6], the above construction was used for transmitting a first-order Gauss-Markov source over two erasure channels with independent failure probabilities. The performance of the construction was compared to that of independent (non-predictive) coding, and the low-latency MR-DPCM approach proposed in [2]. As Fig. 1 shows, the WYZE-PMD construction significantly out-performs other approaches over a wide range of rates. Significantly, the MR-DPCM approach which does not take predictive mismatch into account, performs even worse than the naive independent coding approach, illustrating the performance loss due to mismatch.

3. VIDEO CODEC ARCHITECTURE

In this section we describe the architecture of a two-channel predictive MD video codec based on the WYZE-PMD framework. The video codec is designed to generate descriptions for transmission over two erasure channels with equal and independent failure probabilities. The proposed video codec employs LDPC codes to generate coset information, since LDPC codes offer better compression at high data rates, and LDPC decoders are less likely to have undetected errors, as compared to turbo codes [9].

3.1. Encoder Architecture

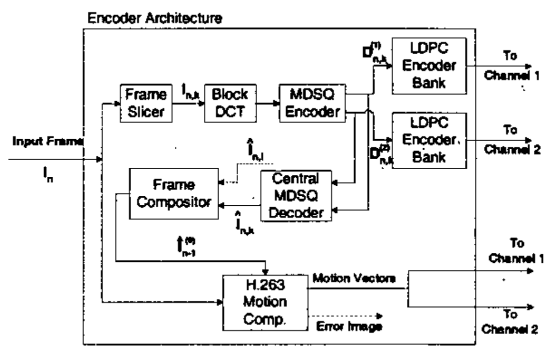

Fig. 2. A simplified representation of the encoder architecture.

Fig. 2 shows a simplified block diagram of the proposed predictive MD encoder. The encoder performs two main functions for each input video frame, namely, (1) it computes the coset information to be transmitted over the two channels, and (2) it computes the motion vectors for the current frame.

Coset information for frame I_n is generated as follows. The frame is segmented into S slices $I_{n,k}$, $k \in \{1, \dots S\}$. A given slice $I_{n,k}$ is transformed using a block-wise DCT transform. The transformed DCT coefficients are encoded using the balanced MD scalar quantizer, with a spread of four, generating two index vectors $\mathbf{D}^{(i)}_{n,k}$, $i = 1, 2$. Each index vector is converted to its corresponding L-bit binary representation and is input to the LDPC encoding bank. As shown in Fig. 3, each LDPC encoding bank consists of L parallel GF(2) systematic LDPC encoders. Prior to coset generation, each binary index vector $\mathbf{D}^{(i)}_{n,k}$ is split into it's L constituent bitplanes, and each bitplane is encoded using one LDPC encoder in the encoder bank. The L resultant parity bitstreams are then concatenated to generate the coset information stream which is transmitted on Channel i. This represents the coset information for slice $I_{n,k}$.

The segmentation of the frame I_n into slices, prior to cosetization, allows the proposed video codec to be used in packet networks with fixed packet sizes. The video codec then transmits a small number of packets ($= S > 1$) over each channel. In such a case, the use of a two-channel MD transform is sub-optimal, and

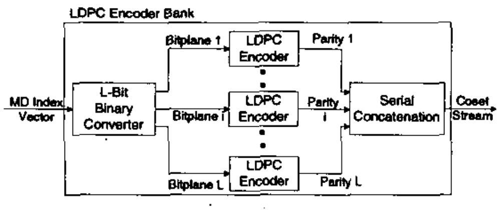

Fig. 3. LDPC encoder bank consisting of L GF(2) systematic LDPC encoders.

performance can be further improved by employing a N-channel MD transform (with N at most $2S$). We defer this for future work.

Motion vectors for the current frame are computed using the motion compensation algorithm specified in the H.263 video coding standard, with the central channel reconstruction $\widehat{I}_{n-1}^{(0)}$ used as the predictor. As the codec is designed to operate over channels with independent failure probabilities, it is essential that the motion vectors be protected using channel-coding techniques, prior to transmission. Since motion vectors typically constitute only $10-15\%$ of the total bitrate, the use of low-block length FEC codes, mandated by the small number of generated packets, does not cause a significant loss in coding efficiency.

3.2. Decoder Architecture

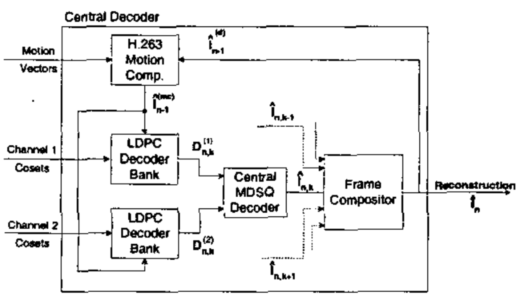

Fig. 4. A simplified representation of the central decoder.

Fig. 4 shows a simplified block diagram of the central predictive MD decoder. The side decoders have similar architectures.

The received motion vectors are used to generate a motion-compensated predictor $\widehat{I}_{n-1}^{(mc)}$ from the decoder reconstruction $\widehat{I}_{n-1}^{(d)}$.[1] For each received channel for a given frame slice, the received coset information and the motion-compensated predictor are used to recover the transmitted MD index vector. This is done by the LDPC decoder bank, shown in greater detail in Fig. 5. The LDPC decoder bank consists of L GF(2) LDPC decoders in series, which sequentially decode the bitplanes extracted from the received coset information. The significance of the serial structure of the LDPC decoder bank can be understood by noting that successful decoding of a given bitplane i yields information about the received description index vector, and can be employed as additional side-information while decoding the subsequent bitplanes $i+1$ through L.

[1]Note that $\widehat{I}_{n-1}^{(d)}$ will not, in general, be the same as the encoder predictor $\widehat{I}_{n-1}$ due to channel failures in the transmission of I_{n-1}.

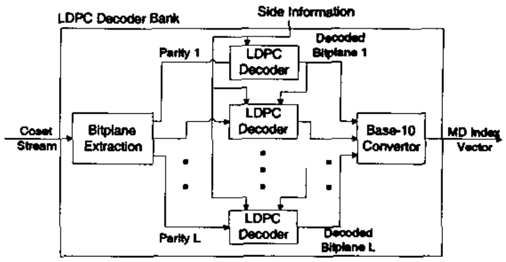

Fig. 5. LDPC decoder bank consisting of L serial LDPC decoders.

The recovered MD index vectors are then used to reconstruct the transmitted slice $\widehat{I}_{n,k}$, and the reconstructed frame slices are composited to reconstruct the frame $\widehat{I}_n$. It should be noted that, provided the LDPC decoder bank can successfully decode each recovered description, the distortion in the reconstructed frame is limited to that caused by channel failures in the communication of frame I_n. Thus error propagation due to predictive mismatch is eliminated.

3.3. Codec Modifications

A number of modifications have been made to improve the performance of the basic codec architecture shown in Fig. 2 and Fig. 4. We briefly describe two modifications that have a significant impact on the rate-distortion performance of the proposed codec.

The efficient use of variable-length codes (VLC) in conjunction with iterative codes is an open problem. As may be noted, the proposed encoder does not use VLC for index vector compression, prior to coset generation. To mitigate the resulting performance loss, the following is done. Two descriptions of the residual error image are obtained by applying the MDSQ transform to the residual obtained from motion compensation. These are compressed as per the H.263 compression standard (which includes the use of VLC), and are transmitted on the respective channels. At the decoder, the residual error reconstructed from the received descriptions is added to the decoder predictor, prior to LDPC decoding, leading to the availability of better side-information at the decoder. This leads to a considerable reduction in the amount of transmitted coset information (which is difficult to compress), and leads to a significant overall gain in rate-distortion performance.

A second modification is the transmission of coset information for only those macroblocks which have non-zero residual error for the current frame. Experiments show that this causes a small amount of error propagation, but significantly improves the performance of the LDPC code[2].

4. RESULTS

We present results for the gray-scale 352×288 (CIF) *cheers* video sequence, with a GOB size of 40 frames (1 I frame followed by 39 P frames). Each frame was segmented into 9 slices, and two packet erasure channels with independent and equal failure probabilities $p_1 = p_2$ were used for transmission. The parity-check matrices of the LDPC codes were designed using the constructions given in [9].

[2]By lowering the effective channel capacity and driving the LDPC performance closer to the bound.

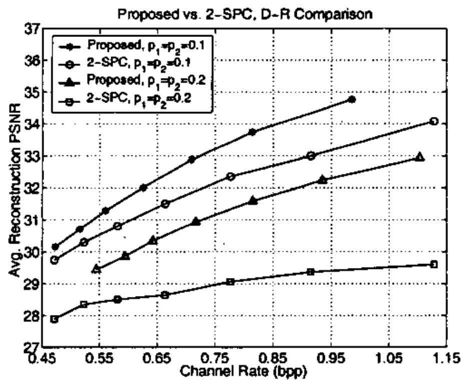

Fig. 6. Performance comparison of proposed approach with 2-SPC for the *cheers* sequence.

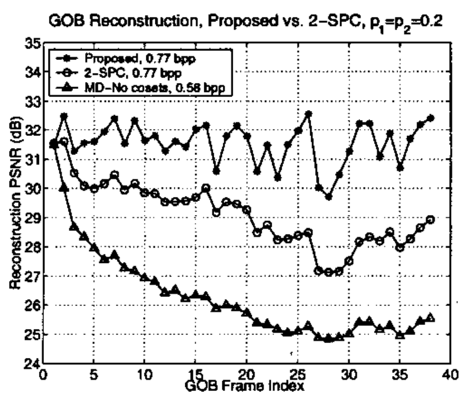

Fig. 7. Frame reconstruction PSNRs for a typical GOB in coded streams generated with proposed approach and with 2-SPC.

The performance of the proposed codec was compared to that of independent MD coding (using H.263 I frames), single-channel H.263 standard predictive coding at the same total bitrate (termed 1-SPC), and two-channel standard predictive coding (termed 2-SPC) in which each channel carried a H.263 standard predictive coded stream. In each case the motion vectors were adequately channel-coded to ensure successful reconstruction. The 2-SPC approach was found to outperform the independent MD coding and 1-SPC approaches—in particular, even with error concealment, the 1-SPC approach incurs significant reconstruction quality loss due to error propagation, for the *cheers* sequence.

Fig. 6 compares the D-R performance of the proposed video codec with that of 2-SPC, for channel failure probabilities of 0.1 and 0.2. The proposed codec outperforms 2-SPC by $0.25 - 1.5$ dB at $p_1 = p_2 = 0.1$, and by $1 - 3$ dB at $p_1 = p_2 = 0.2$. The performance gap increases with higher channel rates and higher channel failure probabilities, reflecting the increased performance loss in standard predictive coding due to error propagation.

Fig. 7 compares the reconstruction quality of the frames in a typical GOB in streams generated by the proposed codec and by 2-SPC, with $p_1 = p_2 = 0.2$. Both streams are coded at approximately the same bitrate—0.770 bpp for the proposed codec, 0.773 bpp for 2-SPC. The 2-SPC reconstruction clearly displays the long-term PSNR decay which characterizes error propagation due to predictive mismatch. In contrast, the reconstruction provided by the proposed codec curtails the effect of predictive mismatch, thereby eliminating error propagation. Also shown, for the sake of comparison, is the performance of the base MD coded stream generated by applying the MDSQ transform to the standard H.263 coded stream, without generation of coset information.

5. CONCLUSIONS

We have presented a two-channel low-latency MD video codec which employs the transmission of coset information to curtail error propagation caused by predictive mismatch. The proposed codec has been shown to yield promising results for MD transmission over packet erasure channels with independent failure probabilities. The performance of the proposed codec can be further enhanced in a number of ways. Simple rate-distortion optimized algorithms can be developed to determine the MD redundancy and the transmission rate for coset information. N-channel MD transforms ($N > 2$) can be used to improve performance for packet networks. Finally, the compression of MD index vectors, prior to coset generation, remains an open problem.

6. REFERENCES

[1] V.K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, pp. 74–93, Sept. 2001.

[2] S. John and V.A. Vaishampayan, "Balanced interframe multiple description video compression," in *IEEE International Conference on Image Processing*, 1999, vol. 3, pp. 812–816.

[3] A.R. Reibman, H. Jafarkhani, Y. Wang, M.T. Orchard, and R. Puri, "Multiple-description video coding using motion-compensated temporal prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 193–204, 2002.

[4] A. Jagmohan and K. Ratakonda, "Multiple description coding of predictively encoded sequences," in *IEEE Data Compression Conference*, 2002, vol. 3, pp. 13–22.

[5] R. Puri and K. Ramchandran, "Multiple description source coding using forward error correction codes," in *Asilomar Conf. Sig. Sys. Computers*, 1999, pp. 342–346.

[6] A. Jagmohan and N. Ahuja, "Wyner-ziv encoded predictive multiple descriptions," To be presented at the Data Compression Conference, Snowbird, Mar. 2003.

[7] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, Jan. 1976.

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

[9] D.J.C. Mackay and R.M. Neal, "Near shannon limit performance of low density parity check codes," *Electronics Letters*, vol. 32, pp. 1645, Aug. 1996.

[10] Draft ITU-T Recommendation H.263, "Video coding for low bitrate communications," May 1996.