A STATE-FREE CAUSAL VIDEO ENCODING PARADIGM

Anshul Sehgal, Ashish Jagmohan, Narendra Ahuja

University of Illinois
asehgal, jagmohan, n-ahuja@uiuc.edu

ABSTRACT

A commonly encountered problem in the communication of predictively encoded video is that of predictive mismatch or drift. The problem of predictive mismatch manifests itself in numerous communication scenarios, including on-demand streaming, real-time streaming and multicast streaming. This paper proposes a state-free video encoding architecture that alleviates this problem. The main benefit of state-free encoding is that there is no need for the encoder and the decoder to maintain the same state, or equivalently, predict using the same predictor. This facilitates robust communication of causally encoded media. The proposed approach is based on the Wyner-Ziv theorem in Information Theory. Consequently, it leverages the superior performance of coset codes for the Wyner-Ziv problem for predictive coding. A video codec, with state-free functionality, based on the H.26L encoding standard is proposed. The performance of the proposed codec is within 1-2.5 dB of the H.26L encoder.

1. INTRODUCTION

The growing popularity of real-time and on-demand streaming of video has spurred much research in the fields of video compression and video transmission. Efficient video compression algorithms such as MPEG-* [1] and H.26* [2] exploit temporal redundancy effectively by predictively encoding the media. Transmission of predictively encoded media, however, is a difficult task owing to the problem of predictive mismatch, or drift. Predictive mismatch refers to the scenario where the reconstruction of the predictor symbol at the decoder is different from the predictor symbol used at the encoder causing an erroneous reconstruction of the predicted symbol at the decoder. This error propagates through the sequence, leading to distorted decoder reconstructions of all subsequent decoded symbols.

The problem of predictive mismatch manifests itself in various scenarios of practical interest such as the multiple description predictive encoding of streams for transmission over error-prone channels, low-delay video conferencing, and scalable predictive coding and multicast communication of predictively encoded media over heterogeneous networks. In each of these scenarios, it is imperative that the encoder knows the state of the decoder at each time step. A mismatch between the state of the encoder and the decoder results in predictive mismatch. Previous approaches aimed at mitigating the effect of predictive mismatch have typically addressed the problem from within the context of the respective applications outlined above.

In this work, we propose the design of a video encoder that does not require the stipulation that the encoder and the decoder maintain the same state at each time step, thereby eliminating the problem of error propagation caused by predictive mismatch. The proposed encoder can thus be used for generating predictively encoded streams for the wide range of scenarios mentioned above.

The formulation of predictive video coding as a Wyner-Ziv side-information problem was first proposed in [3]. Based on this interpretation, it was suggested that the use of powerful coset codes, that approach the capacity of the Wyner-Ziv channel, makes it possible to design W-Z based video encoders that approach the performance of conventional predictive coding. Since the publication of [3], video codecs based on this formulation have been proposed in [4, 5, 6, 7, 8]. We comment on the relative performances of the codecs in Section 7.

In this paper, we leverage this formulation to design a state-free causal video codec which obviates the requirement that the encoder and decoder need to maintain precisely identical states for correct decoding. The proposed video codec employs a *state-free design*[1], and utilizes coset codes for mitigation of predictive mismatch without overly sacrificing compression efficiency. Compression is achieved by using Low-Density Parity Check (LDPC) [9] based coset codes for source coding of video. Further, the added functionality and robustness is achieved without sacrificing compression efficiency.

2. PRELIMINARIES

Consider the one-step predictive encoding of a M-dimensional source with first-order memory, $\{\mathbf{v}_i\}_{i=0}^{\infty}$, $\mathbf{v}_i \in \Re^M$. Given the reconstruction of source symbol $\mathbf{v}_{k-1}$, denoted as $\widehat{\mathbf{v}}_{k-1}$, source symbol $\mathbf{v}_k$ is encoded by generating the innovation $\mathbf{t'}_k = \mathbf{v}_k - E[\mathbf{v}_k|\widehat{\mathbf{v}}_{k-1}]$. We assume that $E[\mathbf{v}_k|\widehat{\mathbf{v}}_{k-1}] = \widehat{\mathbf{v}}_{k-1}$, thus, $\mathbf{t'}_k = \mathbf{v}_k - \widehat{\mathbf{v}}_{k-1}$. Subsequently, $\mathbf{t'}_k$ is quantized to yield $\mathbf{t}_k$. Symbol $\mathbf{t}_k$ is then entropy-coded prior to communication to the decoder. Symbol $\widehat{\mathbf{v}}_k$ is reconstructed as $\widehat{\mathbf{v}}_k = \mathbf{t}_k + \widehat{\mathbf{v}}_{k-1}$ and used as the predictor symbol for symbol $\mathbf{v}_{k+1}$. In this manner, while encoding/decoding symbol $\mathbf{v}_k$, it is imperative that the encoder/decoder knows symbol $\widehat{\mathbf{v}}_{k-1}$. Thus, symbol $\widehat{\mathbf{v}}_{k-1}$ denotes the state of the encoder (and the decoder) at time k [2].

In the context of communication of predictively encoded streams over error-prone channels, it is difficult to meet the stipulation that the encoder and decoder always maintain the same state. This paper presents a video codec design, based on the Wyner-Ziv encoding paradigm, which solves precisely this problem, albeit with a small decrement in compression efficiency as compared to conventional predictive coding.

3. OVERVIEW OF SOLUTION

In [10], Wyner and Ziv considered the problem of communication of a continuous random variable X with correlated side-information Y, *available to the decoder*, but not the encoder. Specifically, they showed that for the case of X and Y jointly-Gaussian, ignorance of Y at the encoder does not entail any loss in the achievable rate-distortion performance of the system over the case where Y was available at the encoder as well as the decoder. For non-Gaussian random variables there is some loss in the achievable rate-distortion performance of the system, but the presence of correlated side-information still allows encoding of the source X at a lower rate.

[1] So called, since the encoder need not know the decoder state while generating the compressed stream.

[2] In general, if an r^{th} order source process is to be predictively encoded, the state at time k is completely represented by the r-dimensional vector $\{\mathbf{v}_{k-1}, \mathbf{v}_{k-2}, \ldots, \mathbf{v}_{k-r}\}$.

0-7803-7750-8/03/$17.00 ©2003 IEEE

Recently, numerous authors have proposed code constructions for the Wyner-Ziv problem [11, 12]. Among the coset codes available in the literature, powerful coset codes based on turbo codes [12] come closest to the Wyner-Ziv rate-distortion function. LDPC codes, like turbo codes, approach the Shannon capacity for conventional channel coding. In addition, LDPCs and turbo codes are both soft-decodable, high-dimensional linear block codes over $GF(2^m)$. Thus, it is reasonable to expect that algorithms based on LDPC codes should have comparable performance to the turbo codes based encoding and decoding algorithms proposed in [12] for the Wyner-Ziv problem. We choose to use LDPC based coset encoding/decoding, since LDPC codes have certain advantages over turbo codes, such as the significantly lower frequency with which undetected errors occur in LDPC decoding.

4. PROPOSED ALGORITHM

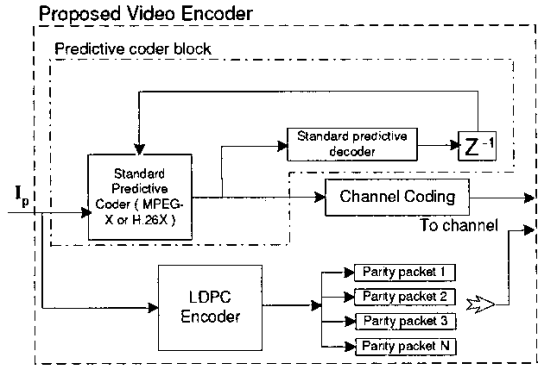

Fig. 1. Block diagram of the proposed encoder.

In this section, we describe the proposed coset-code based video encoding and decoding algorithms. In the ensuing discussion, symbol $\widehat{a}$ denotes the quantized symbol a, and symbol a' denotes the decoder reconstruction of a, which is possibly corrupted due to channel errors. Thus, $\widehat{a}'$ denotes the decoder reconstruction of symbol $\widehat{a}$.

The proposed codec performs coset based encoding/decoding as follows: Consider a constituent source vector, $\mathbf{v}_k$, of the M-dimensional source with first-order memory, $\{\mathbf{v}_i\}_{i=0}^{\infty}$. While encoding, symbol $\mathbf{v}_k$ is first quantized to yield $\widehat{\mathbf{v}}_k$. Adequate coset information of $\widehat{\mathbf{v}}_k$ is then transmitted to the decoder such that the decoder is able to reconstruct $\widehat{\mathbf{v}}_k$ using the transmitted coset information and the side-information $\widehat{\mathbf{v}}_{k-1}$. Even in the event that $\widehat{\mathbf{v}}_{k-1}$ is not decoded correctly due to channel errors, i.e. $\widehat{\mathbf{v}}'_{k-1} \neq \widehat{\mathbf{v}}_{k-1}$, the decoder can still decode $\widehat{\mathbf{v}}_k$ perfectly, so long as adequate coset information is transmitted. Thus the encoder need not know the state of the decoder ($\widehat{\mathbf{v}}'_{k-1}$) while encoding $\mathbf{v}_k$ as long as it knows the statistical correlation between $\widehat{\mathbf{v}}'_{k-1}$ and $\mathbf{v}_k$. This ensures that the proposed approach performs encoding in a state-free fashion.

Fig. 1 depicts a simplified block diagram which shows how coset-based codes can be used for video compression without predictive mismatch. Let $\mathcal{I}_j$ and $\widehat{\mathcal{I}}_j$ denote the j^{th} frame of the original video sequence and the j^{th} frame of the quantized video sequence respectively. Video frame $\widehat{\mathcal{I}}_j$ is obtained from $\mathcal{I}_j$ by applying the H.26L forward transform to each 4×4 block of $\mathcal{I}_j$ and quantizing the transformed image using the H.26L dead-zone quantizer. We represent the k^{th} transform domain vector frequency component of image $\mathcal{I}_j$ as $\mathbf{v}_j^k$, $0 \leq k \leq 15$. The corresponding vectors for $\widehat{\mathcal{I}}_j$ are denoted as $\widehat{\mathbf{v}}_j^k$. Coset packets of frequency vector $\widehat{\mathbf{v}}_j^k$ are generated during encoding. While decoding $\widehat{\mathbf{v}}_j^k$, an appropriate subset of these coset packets are used along with $\widehat{\mathbf{v}}'^k_{j-1}$ as side-information. Since the encoding and decoding of all frequencies $\widehat{\mathbf{v}}_j^k$ in image $\widehat{\mathcal{I}}_j$ is identical and independent of other frequencies, we drop the superscript k from $\mathbf{v}_j^k$ in the sequel and describe the encoding and decoding procedure for a generic frequency vector.

4.1. Encoding Algorithm

Fig. 2(a) shows a block diagram of the LDPC encoder. Encoding of vector $\widehat{\mathbf{v}}_j$ (with q symbols) proceeds by converting it into its L-bit binary representation. Denote bit-plane l of $\widehat{\mathbf{v}}_j$ as $\widehat{\mathbf{v}}_j(l)$. Also, let $\mathbf{G}$ denote the $p \times q$ systematic generator matrix of the LDPC code. In our implementation, we use the algorithm in [9] to generate the parity-check matrix $\mathbf{H}$, and consequently $\mathbf{G}$, using a pseudo-random seed. The $p - q$ parity bits of the product $\widehat{\mathbf{v}}_j(l) \cdot \mathbf{G}$ constitute one coset packet. For each bit-plane l, multiple coset packets are generated by varying p. During the streaming session, an appropriate coset packet for each bit-plane of each frequency is transmitted to the decoder for each video frame $\mathcal{I}_j$. The coset packet that is transmitted depends upon the statistical characterization of the lossy channel. It is noted that while encoding frame $\mathcal{I}_j$, the encoder does not require knowledge of the decoder's reconstruction of frame $\mathcal{I}_{j-1}$, thereby providing state-free encoding.

4.2. Decoding Algorithm

Fig. 2(b) shows a block diagram of the LDPC decoder. Decoding of $\widehat{\mathbf{v}}_j$ proceeds by using the decoder reconstruction $\widehat{\mathbf{v}}'_{j-1}$ as side-information, and the received parity bits as the coset information. The bit-planes of $\widehat{\mathbf{v}}_j$ are sequentially decoded. The proposed sequential decoding structure of LDPC decoders is required since successful decoding of bitplane l yields information about the symbol to be reconstructed, which can be used as additional side-information for decoding the subsequent bitplanes $l+1$ through L. It should be noted that, in the context of transmission over lossy channels, the errors in the reconstructed sequence $\widehat{\mathbf{v}}'_j$ are limited to the errors incurred in the transmission of coset information for $\widehat{\mathbf{v}}_j$. Thus, the propagation of error due to predictive mismatch is eliminated.

5. PERFORMANCE ENCHANCEMENTS

In this section, we briefly outline some enchancements which significantly improve the rate-distortion performance of the basic codec described above.

Conventional predictive coding exploits the temporal redundancy between the video frame currently being encoded, and previously decoded video frames, by computing motion compensated differences. It then exploits the spatial redundancy in the computed residual by applying run-length coding followed by VLC coding. In the proposed codec, however, decoding of each frequency component of a video frame is performed independent of other frequency components. Secondly, it is assumed that the coefficients in any frequency vector are independent and identically distributed. Thus, the proposed codec only exploits the temporal redundancy, and makes no attempt to exploit the spatial redundancy. This is because the efficient use of VLCs in conjunction with iteratively decodable channel codes is still an open problem. Thus, the performance of the proposed approach should almost certainly be worse than that of predictive coding. We now describe a simple alteration, that significantly reduces this performance loss.

The key observation that we make is that in the above approach, the coset information transmitted conveys two pieces of information: 1) it *corrects for the errors* introduced by the channel, 2) It *conveys innovation information*, $\mathbf{v}_k - E[\mathbf{v}_k|\widehat{\mathbf{v}}_{k-1}]$, with respect to the corrected decoder reconstruction. Among these, the first is unknown to the encoder since the channel is statistical in nature. However, the second *is completely known* at the encoder. Thus, it would be judicious to encode $\mathbf{v}_k - E[\mathbf{v}_k|\widehat{\mathbf{v}}_{k-1}]$ using conventional

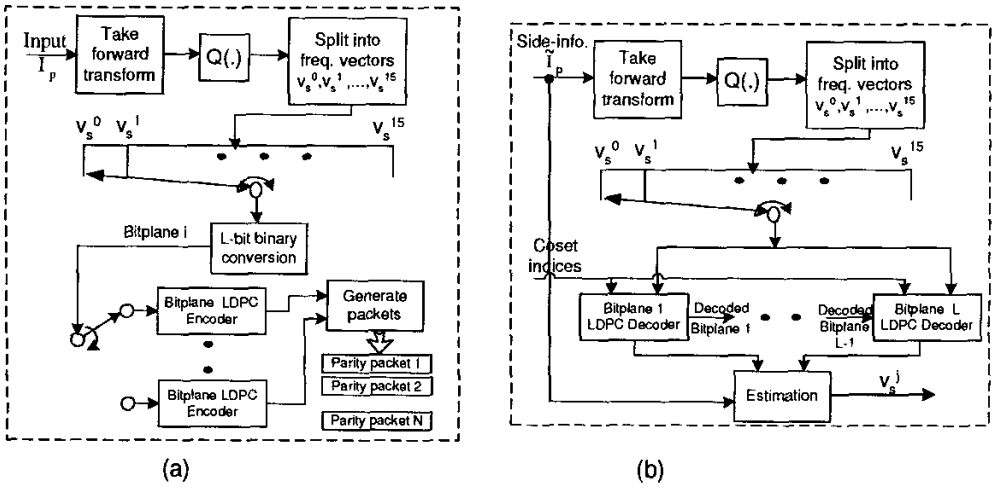

Fig. 2. (a) shows a block diagram of the LDPC encoder.(b) shows a block diagram of the LDPC decoder. The bitplane LDPC decoders have serial decoding structure.

source coding techniques that exploit both spatial and temporal redundancy.

Thus, in the modified codec, while predictively encoding $\mathbf{v}_k$, the encoder generates two pieces of information: a) $t_{1,k} = \mathbf{v}_k - \mathbf{v}_{k-1}$), b) the coset information $t_{2,k} = coset\ of \widehat{\mathbf{v}}_k$. Vector $t_{1,k}$ is then quantized to yield $\widehat{t}_{1,k}$ prior to transmission. The decoder estimates the predictor for $\mathbf{v}_k$ as $\widehat{t}_{1,k} + \widehat{v}'_{k-1}$ and uses this as the side-information while performing side-information decoding of $\mathbf{v}_k$.

We note that, as an additional benefit, the theoretical loss in rate for the modified codec over conventional predictive coding is only due to the deviation from Gaussianity of the channel losses—unlike the basic codec, no additional performance loss results due to deviation from Gaussianity of the the innovation in the source sequence.

Another modification to the basic codec that signifiacntly improved its rate-distortion performance is the estimation procedure while decoding $\widehat{\mathcal{I}}_j$. The estimation procedure makes an MMSE estimate of $\mathbf{v}_j$ based on the side-information $\widehat{\mathbf{v}}'_{j-1}$ and the LDPC decoded vector $\widehat{\mathbf{v}}_j$. As such, the MMSE estimate requires knowledge of the pdf of $\mathbf{v}_j - \widehat{\mathbf{v}}_{j-1}$. Simulation results demonstrate that the exponential pdf (with an impulse at the origin) approximates the desired pdf well. Based on this empirical evidence, the encoder transmits the requisite parameters of the parameterized pdf to the decoder. This parameterized pdf is used in the LDPC decoder and also for the estimation routine. Experimental results demonstrate that using the parametrized pdf results in 0.2 dB decrement in the MMSE distortion.

6. RESULTS

We investigate the overall performance of the proposed approach in this section. Experimental results are reported on two sequences –'foreman' and 'cheers'. The 'foreman' sequence is primarily a medium motion, low innovation sequence. The 'cheers' sequence, on the other hand, is a high motion, high innovation sequence. One hundred frames, at 30 frames per second, of each of these sequences were compressed using the proposed codec. For each of these sequences, the first frame was intra coded. All subsequent frames were encoded as P frames. The quantization routine of the H.26L codec was altered to reconstruct each point precisely, as opposed to the linear approximation made in the H.26L codec. This led to an improvement of approximately 0.1 dB in the overall performance of the system as compared to the H.26L codec. Loop filtering was

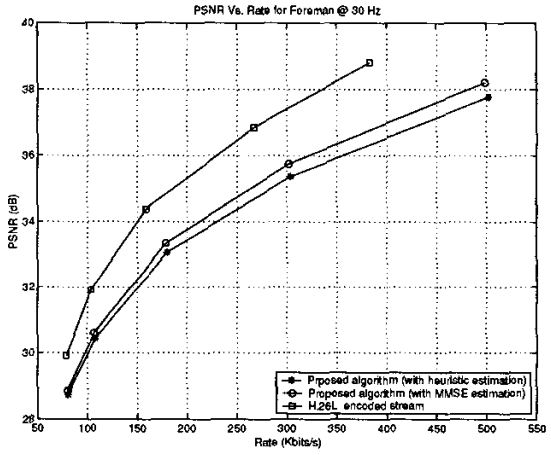

Fig. 3. Distortion-rate plots for 'foreman' sequence.

applied to each video frame just prior to displaying it so that the predictor frame is not loop filtered. This led to a marginal improvement in performance (less than 0.1 dB).

LDPC codes, as proposed in [9], were used in the proposed system. The parity check matrix of each code was generated using a pseudo-random seed. Following this, the algorithm proposed in [9] attempts to eliminate all cycles of length four in the bipartite graph of the code. Sixty-four such LDPC codes with varying redundancy were used in the simulations. The codebooks were made available to the encoder and the decoder prior to streaming. Each coset packet carried a 6-bit identification of the code using which it was encoded.

Figure 3 compares the performance of the proposed state-free encoder with the standard H.26L encoder for the 'Foreman' sequence. As can be discerned from the graph, the distortion-rate performance of the proposed approach is roughly 1 - 1.5 dB worse than that of the H.26L codec at low to medium bitrates. At higher bitrates the performance is 2 - 2.5 dB worse than that of H.26L. We feel that this loss in performance is but a small price to pay for the additional functionality that the proposed codec offers. Figure 3 also quantifies the loss in performance due to the usage of the estimated

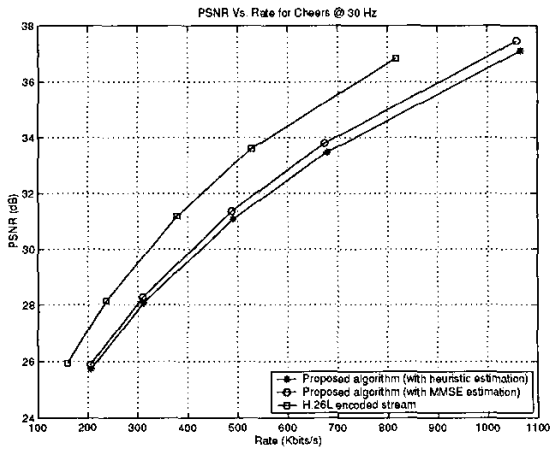

Fig. 4. Distortion-rate plots for 'cheers' sequence.

pdf as opposed to the actual pdf. As can be observed, this loss is quite small. It is also noted, that the coset information accounted for roughly half the bitrate of the entire compressed stream. The other half of the bit-rate comprised of the control and motion information, along with the compressed residual using the unquantized predictor. Lastly, we note that the LDPC based coset code operated at roughly 1.4-1.5 times the empirically calculated conditional entropy of the source. With superior LDPC code designs, the efficiency of the codec can only improve.

Figure 4, analogous to Figure 3 plots the corresponding quantities for the 'cheers' sequence. The performance gap between the proposed codec and the standard H.26L codec is slightly smaller for the 'cheers' sequence as opposed to the 'foreman' sequence. This is only expected. The performance of the proposed codec is inferior to that of the H.26L codec due to the inefficiencies in the compression of the coset information. Further, the coset information comprises mainly of corrections for the quantization errors (since each frame is predicted with an unquantized predictor and reconstructed using a quantized predictor). The bitrate required to compress quantization errors changes little across sequences. The larger amount of innovation in the 'cheers' sequence, however, implies that this quantization information constitutes a smaller fraction of the total bit-rate for 'cheers' as opposed to 'foreman', thus the improvement in the overall performance.

As an example of the functionality of state-free encoding, we note that the plots in Figures 3 and 4 of the proposed codec also represent its baseline performance for scalable compression. This can be easily seen as follows - since the proposed approach provides state-free compression, the state of the decoder is immaterial to the reconstruction of any given video frame. This is precisely the desired functionality of a layered codec. In comparison, layered compression of MPEG-4 suffers approximately 2-4 dB as compared to baseline MPEG-4.

7. CONCLUSIONS

In our opinion, the most important contribution of this paper is the design of a robust video encoding architecture based on the Wyner-Ziv side-information problem. In this regard, the proposed state-free encoding paradigm paves the way for further research on a number of application scenarios. These include, but are not limited to, application scenarios where video is communicated over error-prone media such as the Internet and the wireless channel, such as on-demand streaming, real-time streaming, multicasting, multiple description coding etc.. Also, as pointed out above, the proposed approach holds promise for layered coding, where the source compression algorithm is oblivious of the desired decodable bitrate, yet is required to be flexible enough to accommodate a wide range of bitrates.

It has recently been brought to our notice that while this paper was in review, Puri and Ramchandran [8] have proposed a video codec based on similar principles. Their codec performs 2 - 3 dB worse than H.263 in the high rate regime (results in the low-rate regime are not provided). In comparison, our results are 1 - 2.5 dB worse than H.26L across all bit-rates. Since the H.26L codec is itself superior to H.263 by 2-4 dB, the proposed codec can be expected to outperform the codec proposed in [8] by 3-5 dB. In our opinion, the superior performance of our codec is due to the artful integration of the H.26L encoder within the proposed framework.

8. ACKNOWLEDGEMENTS

We take this opportunity to thank Dr. Philip A. Chou, Microsoft Research, for his contributions during the germinal stages of this work. We also thank Prof. David MacKay and Prof. Radford McNeal for providing the C code for LDPC codes.

9. REFERENCES

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

[2] T. Wiegand, "H.26l test model long-term number 9 (tml-9) draft0," *ITU-T Video Coding Experts Group*, Dec. 2001, document VCEG-N83d1.

[3] A. Jagmohan, A. Sehgal, and N. Ahuja, "Predictive encoding using coset codes," in *IEEE International Conference on Image Processing*, 2002, vol. 2, pp. 29–32.

[4] A. Sehgal and N. Ahuja, "Robust predictive coding and the wyner-ziv problem," in *IEEE Data Compression Conference*, 2003, pp. 103–112.

[5] A. Aaron, S. Rane, R. Zhang, and B. Girod, "Wyner-ziv coding for video: Applications to compression and error resilience," in *IEEE Data Compression Conference*, 2003.

[6] A. Jagmohan, A. Sehgal, and N. Ahuja, "Wyze-pmd based multiple description video codec," in *IEEE Int. Conf. on Multimedia and Expo*, 2003.

[7] A. Sehgal, A. Jagmohan, and N. Ahuja, "Error correction coding for non-stationary channels," under review, IEEE Transactions on Multimedia and Signal Processing.

[8] R. Puri and K. Ramchandran, "Prism: A video encoding architecture based on distributed compression principles," ERL Berkeley Technical Report.

[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] 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.

[11] S.S. Pradhan, *Distributed source coding using syndromes (DISCUS)*, PhD dissertation, University of California at Berkeley, 2001.

[12] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *IEEE Data Compression Conference*, 2002, pp. 252–261.