

Systematic Lossy Error Protection of Video Signals

Shantanu Rane, *Member, IEEE*, Pierpaolo Baccichet, *Member, IEEE*, and Bernd Girod, *Fellow, IEEE*

Abstract— This paper proposes a scheme called Systematic Lossy Error Protection (SLEP) for robust transmission of video signals over packet erasure channels. The systematic portion of the transmission consists of a conventionally encoded video bit stream which is transmitted without channel coding. An additional bit stream generated by Wyner-Ziv encoding of the video signal is transmitted for error resilience. In the event of packet loss, this supplementary bit stream is decoded and allows the recovery of a coarsely quantized video signal, which is displayed in lieu of the lost portions of the primary video signal. The quantization mismatch results in a small, controlled loss in picture quality, but a drastic reduction in picture quality is avoided. An implementation of the SLEP system using the state-of-the-art H.264/AVC standard codec is described. Specifically, H.264/AVC redundant slices are used in conjunction with Reed-Solomon coding to generate the Wyner-Ziv bit stream. The received video quality is modeled as a function of the bit rates of the primary and redundant descriptions and the error resilience bit rate. The model is used to optimize the video quality delivered by SLEP. Via theoretical analysis and experimental simulation, it is shown that SLEP provides a flexible tradeoff between error resilience and decoded picture quality. By allowing the quality to degrade gracefully over a wider range of packet loss rates, SLEP mitigates the precipitous drop in picture quality suffered by traditional FEC-based systems.

Index Terms— Wyner-Ziv coding, distributed video coding, side information, systematic lossy source/channel coding, redundant slices, H.264/AVC

I. INTRODUCTION

In typical video transmission systems, a video signal is compressed, and the resulting bit stream is transmitted over an error-prone channel. The errors may consist of symbol errors caused by fading, as observed for wireless channels, or packet erasures caused by congestion, as observed in the Internet. Errors or losses in a received video packet must be concealed in order to provide a presentation with acceptable visual quality. Error concealment schemes alone cannot guarantee acceptable video quality at the error probabilities encountered in video transmission systems. Hence, a combination of Forward Error Correction (FEC) and feedback (if permissible for the given application) is used to protect the video packets, in exchange for an increase in latency and a small expansion of the transmitted bit rate. For example, MPEG-2 transport uses a (204,188) Reed-Solomon code to protect the packets in the broadcast stream. This code can correct a maximum of 8 byte errors or 16 byte erasures, but at high error probability, the

An earlier version of this work has been presented in part at the 2006 IEEE Packet Video Workshop [1] and the 2006 IEEE Picture Coding Symposium [2]. The present work includes a more detailed theoretical analysis of Systematic Lossy Error Protection (SLEP). Further, the new experimental settings are in accordance with suggestions received from the ITU-T/MPEG Joint Video Team (JVT) for a SLEP standardization experiment conducted from April to October 2006.

Copyright © 2008 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

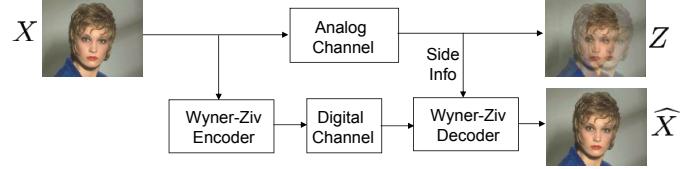


Fig. 1. The Wyner-Ziv decoder uses the degraded output of the analog channel as side information and delivers an output image of superior visual quality.

number of erroneous or erased bytes overwhelms the Reed-Solomon code. This results in rapid degradation of the picture quality, a phenomenon commonly referred to as the “cliff” effect.

In our recent work [1]–[3], we have proposed Systematic Lossy Error Protection (SLEP), a scheme which uses ideas from Wyner-Ziv coding to provide error resilience. This scheme achieves a graceful tradeoff between decoded video quality and resilience to transmission errors, effectively mitigating the cliff effect of FEC. A closed-form rate-distortion function has been derived for a simplified SLEP scheme applied to a first-order Markov source [4], and the graceful degradation property of SLEP has been verified for that case. In this work, we apply the SLEP principle to H.264/AVC video transmission. We use redundant slices and Reed-Solomon coding to construct a Wyner-Ziv codec, and analyze the performance of the proposed scheme. Our experimental results demonstrate that SLEP provides superior video quality than FEC-based schemes with similar buffering delay and a small increase in encoding/decoding complexity. It is therefore an attractive alternative for future digital video broadcasting systems. SLEP is based on the systematic lossy source/channel coding framework in which a source, X , is transmitted over an analog channel without coding. Owing to errors in the channel, a noisy version, Z , is received. As shown in Fig. 1, a second encoded version of X is sent over a digital channel as enhancement information. The noisy version, Z , serves as side information to decode the output of the digital channel and produce the enhanced version, \hat{X} . Thus, source coding with decoder side information, i.e., Wyner-Ziv coding [5], is performed to limit the degradation caused by the analog channel. The term “systematic coding” stems from systematic error-correcting channel codes and refers to a partially uncoded transmission. The information theoretic bounds and conditions for optimality of this configuration were derived by Shamai, Verdú, and Zamir in [6], and extended to the case of hierarchical source/channel coding by Steinberg and Merhav [7].

SLEP differs from other recently proposed schemes for distributed video coding [8]–[11]. The difference is that, in

these schemes, the Wyner-Ziv codec is a part of the video encoding and is necessary for both source coding efficiency and resilience to channel errors. In contrast, SLEP uses the Wyner-Ziv codec solely for error resilience and is, in principle, independent of the video compression scheme employed in the systematic transmission. Thus, SLEP is backward compatible with existing video transmission systems, because legacy decoders can simply discard the Wyner-Ziv bit stream.

The remainder of this paper is organized as follows. Section II reviews the principle of SLEP. Section III describes a SLEP system constructed using H.264/AVC redundant slices. In Section IV, the received video quality is modeled as a function of the bit rates of the (primary) video slices, redundant slices, and the Wyner-Ziv bit rate. In Section V, the accuracy of the model is verified by comparing the model's prediction with experimental results. The model is then used to select the aforementioned bit rates in order to optimize the received video quality. Using standardized internet error traces, the performance of SLEP is compared against that of FEC.

II. SYSTEMATIC LOSSY ERROR PROTECTION

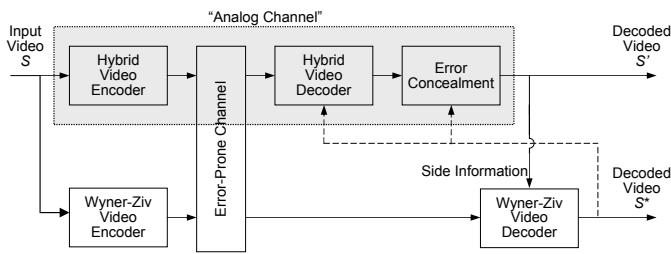


Fig. 2. A Wyner-Ziv decoder uses the decoded error-concealed video waveform as side information in a systematic lossy source/channel coding framework.

The principle of SLEP is illustrated in Fig. 2. At the transmitter, the input video S is compressed independently by a hybrid video encoder and a Wyner-Ziv encoder. As shown, the compressed video signal transmitted over the error-prone channel constitutes the systematic portion of the transmission. For robustness, the systematic portion is augmented by the Wyner-Ziv bit stream. The Wyner-Ziv bit stream can be thought of as a second description of S , but with coarser quantization. Thus the Wyner-Ziv bit stream contains a reduced quality description of the original video signal.

Without transmission errors, the Wyner-Ziv description is fully redundant, i.e., it can be regenerated bit-by-bit at the decoder, using the decoded video S' . When transmission errors occur, the decoder attempts to perform error concealment, but some portions of S' might still have unacceptably large errors. In this case, Wyner-Ziv bits allow reconstruction of the second description, using the decoded waveform S' as side information. This coarser second description and side information S' are combined to yield an improved decoded video S^* . In portions of S' that are unaffected by transmission errors, S^* is essentially identical to S' . However, in portions of S' that are degraded by transmission errors, the coarser second representation limits the maximum degradation in the current decoded frame. This "repaired" frame is then fed back to the video decoder to serve as a more accurate reference

for motion-compensated decoding of the subsequent frames. Since digital video transmission is being considered, there is no analog/digital channel separation, as was the case in Fig. 1. However, the role played by the hybrid video codec and the error-prone channel in Fig. 2 is analogous to the role played by the analog channel in Fig. 1.

Thus, SLEP is essentially a scheme which efficiently transmits an alternative representation of the video signal which may be used when portions of the main signal (also referred to as the "primary" description) are lost or corrupted during transmission. Since this alternative representation is coarsely quantized, each successful instance of Wyner-Ziv decoding results in a quantization mismatch between the high-quality primary description and the reduced quality second description. Owing to motion-compensated decoding of predictively coded frames, this quantization mismatch propagates to the subsequent frames. The more frequently channel errors occur, the more frequently Wyner-Ziv decoding is invoked and the larger is the quantization mismatch. To ensure that the ensuing video quality degradation is graceful, the quantization mismatch must be controlled, i.e., the quantization levels in the second description must be selected appropriately. This adds a degree of freedom in the design of a SLEP system, compared to the design of traditional FEC-based systems. For optimizing FEC, the designer has only to determine the percentage of the available bit rate that must be allocated for channel coding. In SLEP, it is not sufficient to determine the percentage of the available bit rate that must be allocated to Wyner-Ziv coding. Indeed, it is also essential to determine the quality (equivalently, the source coding bit rate) of the second description which can be recovered by Wyner-Ziv decoding.

III. SLEP BASED ON H.264/AVC REDUNDANT SLICES

A. SLEP Encoding

The implementation of the SLEP scheme, using the state-of-the-art H.264/AVC standard tools, is shown in Fig. 3. The following operations are performed on the encoder side:

- 1) *Generation of redundant slices*: Each macroblock is now redundantly encoded and grouped into a *redundant slice*. A redundant slice is an alternative representation of an already encoded (primary) video slice [12]. The redundant slices in a given frame constitute a *redundant picture*. In our implementation, the redundant slices have the same shape, coding modes, motion vectors and reference frames as the primary slices, but use coarser quantization. The restriction of using the same shape, coding modes, motion vectors and reference pictures is not imposed by H.264/AVC, but is imposed by us in order to simplify the implementation of the Wyner-Ziv decoder and limit the amount of side information that need to be transmitted.
- 2) *Reed-Solomon encoding*: Reed-Solomon codes perform the role of Slepian-Wolf coding [13] in this system. A Reed-Solomon code over $GF(2^8)$ is applied across the redundant slices, to generate parity slices, as shown in Fig. 4(a). The number of parity slices generated per frame depends upon the allowable error resilience bit

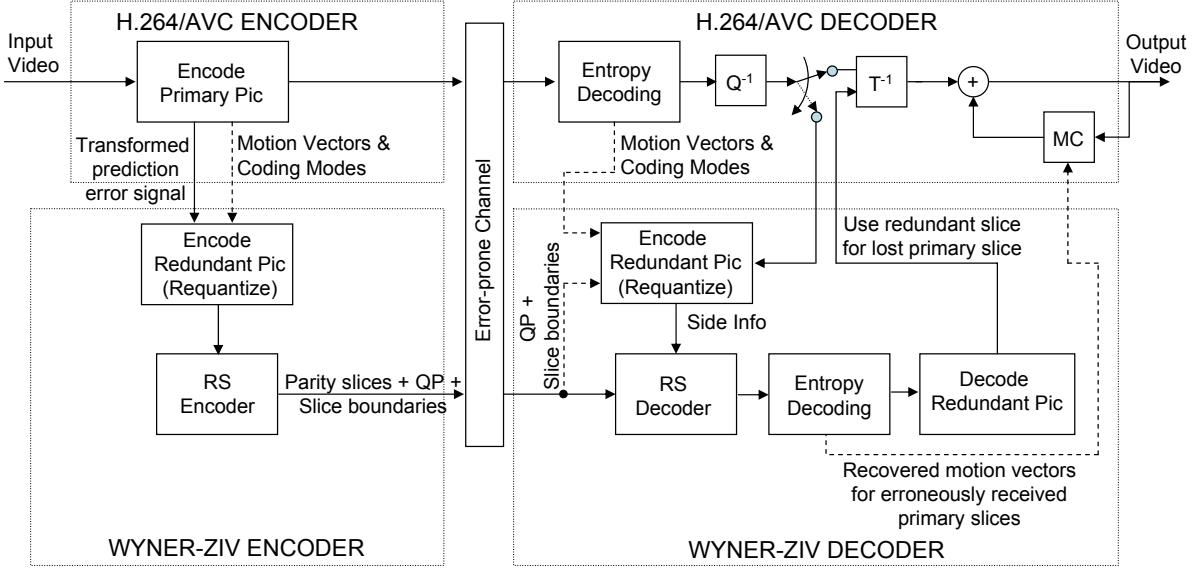


Fig. 3. Implementation of a Slepian-Wolf-like system using H.264/AVC redundant slices. Reed-Solomon codes applied across the redundant slices play the role of Slepian-Wolf codes in distributed source coding. At the receiver, the Wyner-Ziv decoder obtains the correct redundant slices using the error-prone primary slices as side information. The redundant description is used in lieu of the lost portions of the primary (systematic) signal.

rate, and can vary slightly from frame to frame. The redundant slices are then *discarded* and *only the parity slices* are included for transmission in the Wyner-Ziv bit stream. This is reminiscent of the analogy between Slepian-Wolf coding and traditional channel coding described in [3], where parity symbols corresponding to the source were transmitted in order to correct the errors in the side information.

- 3) *Wyner-Ziv bit stream generation:* In addition to the parity slices resulting from the previous step, we encode for each slice (1) The index of the first macroblock, (2) The number of macroblocks in a redundant slice (3) The difference between the quantization parameters (QPs) used for encoding the primary and redundant slices. This extra “helper information” is appended to the parity slices generated in the above step. Loss of a parity slice results in loss of the accompanying helper information. Within the H.264/AVC video coding standard [12], this helper information could travel in a Supplemental Enhancement Information (SEI) message, which can be optionally decoded by advanced decoders, but rejected by legacy receivers. A syntax and encoding method for an SEI message which carries this helper information have been proposed in our standardization contribution to the Joint Video Team (JVT) [14].

B. Slepian-Wolf Decoding

The following Wyner-Ziv decoding process is activated only when transmission errors result in the loss of one or more slices from the bit stream of the primary picture:

- 1) *Requantization to recover redundant slices:* This step involves the requantization of the received prediction residual signal of the primary coded picture, followed by entropy coding. This generates the redundant slices used as side information for the Wyner-Ziv decoder.

Note that redundant slices can be generated only for those portions of the frame, where the primary bit stream has not experienced channel errors. The redundant slices corresponding to the error-prone portions are treated as erasures. Since the coding modes for the redundant macroblocks are identical to those in the primary bit stream, the requantization procedure is straightforward.

- 2) *Reed-Solomon (Slepian-Wolf) decoding:* The parity slices received in the Wyner-Ziv bit stream are now combined with the redundant slices, and erasure decoding is performed to recover the slices which were erased from the redundant bit stream, as shown in Fig. 4(b). In the language of distributed video coding, the Reed-Solomon decoder functions as a Slepian-Wolf decoder, and recovers the correct redundant bit stream using the erasure-prone redundant bit stream as side information.
- 3) *Concealment of lost primary slices:* If Wyner-Ziv decoding succeeds, the lost portions of the prediction residual from the primary (systematic) signal, are replaced by the quantized redundant prediction error signal. The H.264/AVC decoder then performs motion compensation in the conventional manner, using the redundant prediction error signal and the motion vectors recovered from Wyner-Ziv decoding. The coarse fall-back operation results in a quantization mismatch which propagates to the future frames, but a drastic reduction in picture quality is avoided.¹

¹Strictly speaking, we apply Wyner-Ziv coding to the prediction error signal; the redundant slices are obtained via requantization and not via full re-encoding. Therefore, our implementation differs slightly from the principle of Slepian-Wolf described in Section II. Besides reducing the encoding complexity, this ensures that if a redundant slice is regenerated at the decoder, it will be identical to the corresponding untransmitted redundant slice at the encoder. The encoder, however, does not know the side information exactly because of the erasures that occur during the transmission of the primary slices and, in turn, in the regeneration of the redundant slices.

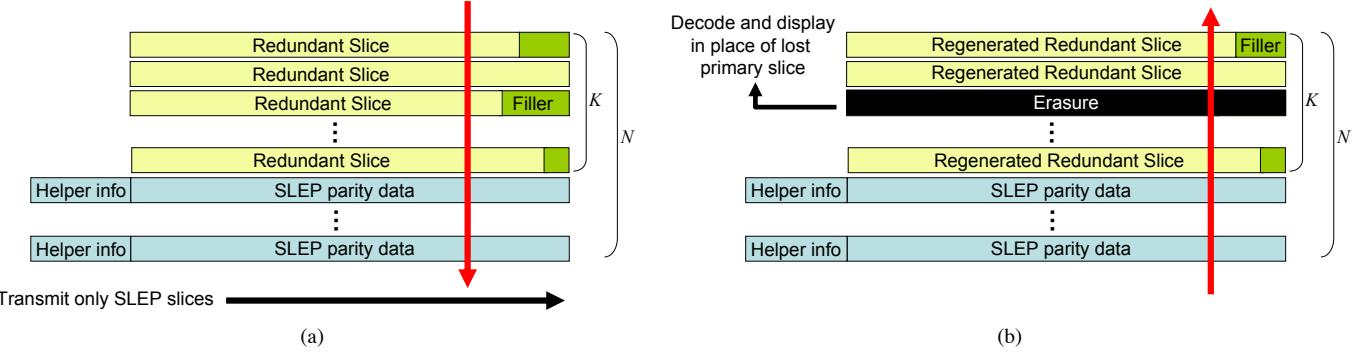


Fig. 4. (a) During Wyner-Ziv encoding, Reed-Solomon codes are applied across the redundant slices and only the parity slices are transmitted to the decoder. To each parity slice is appended helper information about the quantization parameter (QP) used in the redundant slices, and the shapes of the redundant slices. The parity slices, together with the helper information, constitute the Wyner-Ziv bit stream. (b) During Wyner-Ziv decoding, redundant slices corresponding to received primary slices are obtained by requantization, while those corresponding to the lost primary slices are treated as erasures. These are recovered by erasure decoding, using the parity slices and helper information received in the Wyner-Ziv bit stream. These recovered redundant slices are then decoded and displayed in lieu of the lost primary slices.



Fig. 5. Cut-outs of a video frame from the *Foreman* CIF sequence. The primary description is encoded at 408 kb/s, while the error resilience bit rate is fixed at 40 kb/s. Robustness increases with the quantization step in the redundant slices. With QP = 48, the robustness is maximum, but the increased quantization mismatch compromises the decoded picture quality.

From Fig. 3, it is clear that if the parity bit rate at the output of the Reed-Solomon encoder is fixed, then the coarser the quantization in the redundant slices, the greater is the error resilience. A visual example of this tradeoff between error protection and quantization mismatch is shown in Fig. 5 for the *Foreman* sequence. This tradeoff is modeled in Section IV and analyzed in further detail in the Appendix.

It is possible to apply SLEP to a region of interest within a video frame using Flexible Macroblock Ordering (FMO) [12]. This is especially beneficial for video sequences which contain a static or slow-moving background which can be concealed

satisfactorily using decoder-based error concealment. Then, a Wyner-Ziv bit stream need only be generated for the faster-moving foreground objects. For details on the implementation and performance of one such scheme, please refer to [15]. In the modeling and experiments which follow, we do not use FMO and treat only the case where SLEP is applied to the entire video frame.

IV. VIDEO DISTORTION MODEL

A. Motion Compensated Encoding

Since we are concerned with modeling error propagation, the ensuing treatment assumes a video sequence consisting only of predictively encoded frames (i.e., P frames). Let the original value of a pixel at location i in the n^{th} frame be denoted by a random variable X_n^i . This pixel is predicted from another pixel, say, at location j in the encoder's reconstruction of the previous frame, i.e., the previously reconstructed value of \hat{X}_{n-1}^j serves as the predictor for X_n^i . To prevent a source coding mismatch between the encoder and the decoder, the predictor used is the reconstructed pixel value \hat{X}_{n-1}^j and not the original pixel value X_{n-1}^j . Thus,

$$X_n^i = \hat{X}_{n-1}^j + V_n^i \quad (1)$$

where V_n^i denotes the error in the prediction of the current pixel, which is transformed, quantized, entropy-coded and transmitted to the decoder. The encoder also retains \hat{V}_n^i , which is obtained after inverse quantization and inverse transformation of V_n^i . Finally, \hat{X}_n^i the locally reconstructed value of X_n^i , is obtained as

$$\hat{X}_n^i = \hat{X}_{n-1}^j + \hat{V}_n^i$$

With an error-free channel, the decoder would receive \hat{V}_n^i and precisely reconstruct \hat{X}_n^i . The Mean Squared Error (MSE) distortion for the primary slices, resulting from the quantization of the prediction residual, is $\mathcal{D}_p = E(X_n^i - \hat{X}_n^i)^2 = E(V_n^i - \hat{V}_n^i)^2$.

With an error-prone channel, the decoder would receive $\tilde{V}_n^i \neq \hat{V}_n^i$ in general. Thus, the decoder's reconstruction of X_n^i is given by:

$$\tilde{X}_n^i = \tilde{X}_{n-1}^j + \tilde{V}_n^i$$

where \tilde{X}_{n-1}^j is a possibly error-prone reconstruction of X_{n-1}^j , the pixel used for motion compensation. In practice, motion estimation and motion compensation are performed on blocks of 4×4 , 8×8 , 16×16 , or 8×16 pixels. We consider pixel-level motion compensation because this makes it convenient to write expressions for MSE distortion. Our observation has been that such a simplified model tends to slightly underestimate the distortion at low loss probability and to slightly overestimate the distortion at high loss probability.

As explained in Section III, the redundant slices are constrained to use the same reference pixels as the corresponding primary slices in order to mitigate the quantization error propagation that would result when the redundant slice is decoded. Thus, the unquantized prediction error V_n^i for the redundant slices is the same as that in (1) above. However, due to coarser quantization, the redundant reconstructed prediction error is $\hat{V}_n^i \neq \tilde{V}_n^i$ in general. Then, the redundant locally reconstructed value of X_n^i , is obtained as

$$\hat{X}_n^i = \hat{X}_{n-1}^j + \hat{V}_n^i$$

and the MSE introduced due to the coarse quantization used in the redundant slices is $\mathcal{D}_r = E(X_n^i - \hat{X}_n^i)^2 = E(V_n^i - \hat{V}_n^i)^2$.

\hat{V}_n^i)². According to the SLEP scheme described in Section III, \hat{V}_n^i can be recovered at the receiver if Wyner-Ziv decoding is successful for the given bit rate assignment.

While modeling the end-to-end MSE, the following simplifying assumptions are made about the prediction error, the quantization error processes, and the process which introduces erasures during transmission:

- 1) It is assumed that, at the pixel location i , the prediction residual V_n^i , its quantized versions \hat{V}_n^i , \tilde{V}_n^i have zero mean over the duration of the sequence, i.e., $E V^i = E \hat{V}^i = E \tilde{V}^i = 0$
- 2) The quantization errors in the current sample, $V_n^i - \hat{V}_n^i$ and $V_n^i - \tilde{V}_n^i$ are assumed to be respectively independent of $X_{n-1}^i - \hat{X}_{n-1}^i$ and $X_{n-1}^i - \tilde{X}_{n-1}^i$, the errors in the past samples of X_n^i .
- 3) The quantization errors $V_n^i - \hat{V}_n^i$ and $V_n^i - \tilde{V}_n^i$ are assumed to be independent of the errors $\hat{X}_n^i - \tilde{X}_n^i$ and $\tilde{X}_n^i - \hat{X}_n^i$. Note that \tilde{X}_n^i can contain error energy contributed by (a) the quantization mismatch in the current sample $\hat{V}_n^i - \tilde{V}_n^i$ or (b) the erasure of both the current quantized prediction residuals \hat{V}_n^i and \tilde{V}_n^i . In addition they contain error energy propagating from the errors that have occurred previously. Therefore, when the primary video signal is received correctly,

$$\begin{aligned} E(X_n^i - \tilde{X}_n^i)^2 &= E(X_n^i - \hat{X}_n^i + \hat{X}_n^i - \tilde{X}_n^i)^2 \\ &= E(V_n^i - \hat{V}_n^i + \hat{X}_n^i - \tilde{X}_n^i)^2 \\ &= E(V_n^i - \hat{V}_n^i)^2 + E(\hat{X}_n^i - \tilde{X}_n^i)^2 \\ &= \mathcal{D}_p + E(\hat{X}_n^i - \tilde{X}_n^i)^2 \end{aligned} \quad (2)$$

Similarly, for the case in which the redundant slices are decoded instead of the primary slices,

$$E(X_n^i - \tilde{X}_n^i)^2 = \mathcal{D}_r + E(\hat{X}_n^i - \tilde{X}_n^i)^2$$

The expressions used to describe the encoding and decoding process will be now used to find the end-to-end MSE distortion at the decoder.

B. Distortion in the Decoded Video Sequence

The decoder performs different actions depending on whether the primary video packets are received or erased, and whether the number of received parity packets are sufficient for recovering the redundant slices, via Reed-Solomon decoding. Assume that packets are erased (lost), randomly and uniformly, with probability p . Assume also that the location of the macroblocks contained in the lost packet is known. Then, we have the following cases:

- 1) *Primary slices received correctly:* With probability $1-p$, a packet is received and decoded correctly by the main (primary) decoder and Wyner-Ziv decoding is unnecessary. The only source of error energy in this case is the error propagation from previous frames owing to

decoding errors in the past. Thus the mean-squared error in the decoded pixel value is given by:

$$\begin{aligned}\mathcal{D}_n^{EP} &= E(X_n^i - \tilde{X}_n^i)^2 \\ &= E(X_n^i - \hat{X}_n^i)^2 + E(\hat{X}_n^i - \tilde{X}_n^i)^2 \\ &\quad + 2 E[(X_n^i - \hat{X}_n^i)(\hat{X}_n^i - \tilde{X}_n^i)] \\ &= \mathcal{D}_p + E(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)^2 \\ &\quad + 2 E[(V_n^i - \hat{V}_n^i)(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)] \quad (3) \\ &= \mathcal{D}_p + \mathcal{D}_{n-1}^{EE} - \mathcal{D}_p = \mathcal{D}_{n-1}^{EE} \quad (4)\end{aligned}$$

where \mathcal{D}_{n-1}^{EE} denotes the overall MSE distortion in the previous frame. The second term in (3) is split into two terms according to (2), after noting that the error $X_n^j - \hat{X}_n^j$ introduced solely by quantization is independent of the error $\hat{X}_n^j - \tilde{X}_n^j$ introduced solely by the channel erasures. The third term in (3) vanishes because of the zero-mean assumption on the prediction residuals and the independence of $V_n^i - \hat{V}_n^i$ from the past sample difference $\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j$.

- 2) *Successful Wyner-Ziv decoding:* Wyner-Ziv decoding is invoked when the primary video slice is lost. The probability that Wyner-Ziv decoding succeeds, denoted by p_{WZ} , depends on the parameters N and K , of the Reed-Solomon code. Since the location of the lost packet(s) is known to the Wyner-Ziv decoder, the Reed-Solomon decoder has to perform erasure decoding. Similar to traditional erasure codes, Wyner-Ziv decoding succeeds if at least K out of N packets are received, but not otherwise. Since, the Reed-Solomon code is applied across K redundant slices (Fig. 4), we have:

$$p_{WZ} = \sum_{m=K}^{m=N-1} \binom{N-1}{m} (1-p)^m p^{N-1-m} \quad (5)$$

In the case of successful Wyner-Ziv decoding, error energy is contributed by the coarser quantization in the decoded packet as well as by error propagation from the previous frames. The distortion contribution is given by:

$$\begin{aligned}\mathcal{D}_n^{WZ} &= E(X_n^i - \tilde{X}_n^i)^2 \\ &= E(X_n^i - \hat{X}_n^i)^2 + E(\hat{X}_n^i - \tilde{X}_n^i)^2 \\ &\quad + 2 E[(X_n^i - \hat{X}_n^i)(\hat{X}_n^i - \tilde{X}_n^i)] \\ &= \mathcal{D}_r + E(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)^2 \\ &\quad + 2 E[(V_n^i - \hat{V}_n^i)(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)] \\ &= \mathcal{D}_r + \mathcal{D}_{n-1}^{EE} - \mathcal{D}_p \quad (6)\end{aligned}$$

where the assumptions used are identical to those used to derive (4).

- 3) *Decoder-based error concealment:* A decoder-based error concealment scheme must be used if a packet is lost in the systematic transmission and Wyner-Ziv protection is insufficient to reconstruct a redundant version of the lost video slice. In this case, we assume that the lost slice is concealed using its co-located slice in the previous frame. The error energy is now contributed by the process of error concealment of the current packet

as well as by the error propagation from the previous frames. The distortion contribution is then given by:

$$\begin{aligned}\mathcal{D}_n^{EC} &= E(X_n^i - \tilde{X}_n^i)^2 = E(X_n^i - \tilde{X}_{n-1}^i)^2 \\ &= E(X_n^i - \hat{X}_{n-1}^i)^2 + E(\hat{X}_{n-1}^i - \tilde{X}_{n-1}^i)^2 \quad (7) \\ &= E(X_n^i - \tilde{X}_n^i)^2 + E(\hat{X}_n^i - \tilde{X}_{n-1}^i)^2\end{aligned}$$

$$+ \mathcal{D}_{n-1}^{EE} - \mathcal{D}_p \quad (8)$$

$$\begin{aligned}&= \mathcal{D}_p + \text{MSE}(n, n-1) + \mathcal{D}_{n-1}^{EE} - \mathcal{D}_p \\ &= \text{MSE}(n, n-1) + \mathcal{D}_{n-1}^{EE} \quad (9)\end{aligned}$$

where $\text{MSE}(n, n-1)$ is the mean-squared error between the reconstructed current and previous frames. The equality in (7) assumes that the temporal pixel variations, $X_n^i - \hat{X}_{n-1}^i$, are independent of the errors, $\hat{X}_n^i - \tilde{X}_n^i$, introduced by the channel. The equality in (8) assumes that quantization errors, $X_n^i - \hat{X}_n^i$, are independent of the pixel variations and have zero mean as before. At the decoder, it is possible to use error concealment schemes that are more powerful than previous frame concealment. If such an advanced scheme is available, then the term, $\text{MSE}(n, n-1)$ in (9) may be replaced by the true average error energy obtained by the advanced scheme.

In summary, the decoded video packet in the n^{th} frame has a decoded prediction residual given by:

$$\tilde{V}_n^i = \begin{cases} \hat{V}_n^i & \text{w.p. } 1-p \\ \hat{V}_n^i & \text{w.p. } p p_{WZ} \\ \text{erasure} & \text{w.p. } p(1-p_{WZ}) \end{cases}$$

Each case results in a different MSE distortion which was evaluated above. Using (4), (6) and (9) with the corresponding probabilities, the end-to-end distortion in the n^{th} frame due to all of the above effects is then given by:

$$\mathcal{D}_n^{EE} = (1-p) \mathcal{D}_n^{EP} + p p_{WZ} \mathcal{D}_n^{WZ} + p(1-p_{WZ}) \mathcal{D}_n^{EC} \quad (10)$$

More details on the influence of the quantization mismatch on the end-to-end distortion are provided in the Appendix. Note that an intra coded video slice stops the propagation of the error energy associated with the quantization mismatch and the concealment artifacts. Thus, if a macroblock is intra-refreshed every M frames, the average distortion over the M frames is given by:

$$\mathcal{D} = \frac{1}{M} \sum_{n=1}^M \mathcal{D}_n^{EE} \quad (11)$$

C. Encoder Rate-Distortion Model

It now remains to model the distortion-rate tradeoff for encoding the primary and redundant slices. A number of such models have been developed for the purpose of rate-control in standardized video codecs [16]–[18]. Following the analysis of [19], a parametric model is used for the relationship

between the MSE distortion and the encoding bit rate. Thus:

$$\mathcal{D}_p = E(X_n^i - \hat{X}_n^i)^2 = D_{0p} + \frac{\theta_p}{\mathcal{R}_p - R_{0p}} \quad (12)$$

$$\mathcal{D}_r = E(X_n^i - \hat{\tilde{X}}_n^i)^2 = D_{0r} + \frac{\theta_r}{\mathcal{R}_r - R_{0r}} \quad (13)$$

where the encoding parameters, $(D_{0p}, R_{0p}, \theta_p)$ and $(D_{0r}, R_{0r}, \theta_r)$ are determined from trial encodings at the encoder. The parameters depend not only on the sequence being coded, but also on the encoding parameters and mode decisions, such as the frequency with which intra frames are inserted, and the number of reference frames used for predicting the current frame. As shown in Fig. 6, the parameters for the redundant slices are coupled to those of the primary slices because the redundant slices use the same reference pictures, coding modes and motion vectors as the primary slices. Thus, if the primary encoding bit rate is changed, D_{0r} , R_{0r} , and θ_r must be re-estimated. The parameters are updated at short intervals, such as a Group of Pictures (GOP) or one or more seconds of video, to ensure that rate and distortion values reflect the scene content. Note that the transmitted Wyner-Ziv bit rate, denoted by \mathcal{R}_{WZ} is different from the bit rate of the redundant slices \mathcal{R}_r . Wyner-Ziv encoding involves coarse quantization followed by Slepian-Wolf encoding. Coarse quantization reduces the encoding bit rate to $\mathcal{R}_r \leq \mathcal{R}_p$. Slepian-Wolf encoding, which is implemented by applying a channel code to the redundant bit stream and transmitting only the parity symbols, changes the bit rate to $\mathcal{R}_{WZ} \neq \mathcal{R}_r$.

Sometimes, it is more practical to buffer the redundant slices corresponding to one or more frames before applying Reed-Solomon coding. This is especially true when the total bit rate R_T is low, and an entire video frame fits into a single video packet. Assuming that L frames need to be buffered ($L \leq M$), the following optimization problem is solved for each group of M frames:

$$\begin{aligned} & \text{Minimize } \mathcal{D} \\ & \text{such that } \mathcal{R}_p + \mathcal{R}_{wz} \leq R_T \\ & \quad \mathcal{R}_{wz} = \frac{N - K_L}{K_L} \mathcal{R}_r \geq 0 \\ & \quad N \geq K_L \geq 1, \mathcal{R}_p \geq \mathcal{R}_r \geq 0 \end{aligned} \quad (14)$$

where \mathcal{D} is the average distortion from (11), R_T is the total bit rate constraint, K_L is the average number of redundant slices contained in L frames, and for the implementation of Fig. 3, \mathcal{R}_{wz} is the bit rate of the Reed-Solomon parity symbols. Recall that \mathcal{D} depends on \mathcal{D}_p and \mathcal{D}_r which, in turn, depend on the rates \mathcal{R}_p and \mathcal{R}_r via the encoder rate-distortion model in (12) and (13).

TABLE I

30 FRAMES/S SEQUENCES USED FOR SLEP SIMULATIONS.

Sequence	Resolution	Primary Bit Rate (kb/s)	Size of Primary Slices (bytes)
<i>Football</i>	SIF	1024	800
<i>Bus</i>	CIF	1024	800
<i>Mobile</i>	SIF	768	800
<i>Coastguard</i>	CIF	512	500

V. EXPERIMENTAL RESULTS

In this section, we compare the error resilience of SLEP with that of FEC under standardized test conditions, with emphasis on superior picture quality at high packet loss probabilities and graceful degradation with increasing packet loss rate. In our previous work [20], we have compared SLEP with Loss-Aware Rate-Distortion Optimization (LA-RDO) [21], a state-of-the-art error resilience tool in H.264/AVC, which chooses intra coding for those macroblocks that would result in large error propagation if lost. Note that graceful quality degradation can be achieved, albeit with lower coding efficiency, using layered coding with priority encoding transmission (PET) [22]. For a comparison of SLEP with layered video coding for MPEG-2 broadcasting, please refer to [23].

A. Experimental Setup

These experimental settings are based on the recommendations obtained from the Joint Video Team, during the course of the standardization effort for SLEP [14], [20], [24].

- *Video sequences and coding structure:* We use JVT version JM 11 [25] of the H.264/AVC video codec for our simulations. The experiments are carried out on the video sequences listed in Table I. The encoding bit rates for the primary (systematic) video signal are chosen based on the amount of scene complexity and motion present in the sequence. Each sequence consists of an I frame followed by P frames. There are no I frames inserted after the first frame, but intra macroblocks are inserted as explained below. To minimize the effect of instantaneous fluctuations in the channel characteristics on the average picture quality, the sequences are mirrored and concatenated to a length of 4000 frames.
- *Intra macroblock line refresh:* To mitigate error propagation resulting from lost slices, one row of macroblocks is encoded using the Intra mode in each frame. For a CIF frame, this is equivalent to a full frame refresh every 18 frames.
- *Rate control:* The rate control method provided in the H.264/AVC standard codec [12] is used at the encoder to determine the changes in the quantization parameters while encoding the macroblocks in the video sequence. Once the encoding rates for the primary and redundant slices have been decided by the model, the rate control algorithm for the primary picture proceeds independently from that for the redundant picture.
- *Packetization and slices:* The primary slices are constrained to a fixed length (see Table I) at the beginning of the simulation. We do not optimize the slice lengths, but choose them so as to obtain a reasonable tradeoff between header overhead and error resilience. Since each video slice travels inside an IP packet, the words “slice” and “packet” will be used interchangeably. The primary slices are not constrained to contain the same number of macroblocks, and can have different shapes. A redundant slice is constrained to contain the same number of macroblocks as its corresponding primary slice.

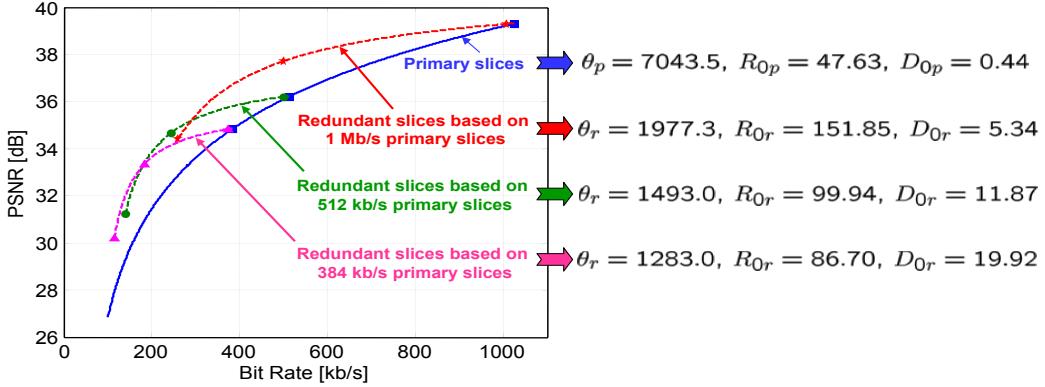


Fig. 6. A few trial encodings (data points) are used to find the parametric rate-distortion curves for the primary and redundant descriptions of the *Foreman* CIF sequence. The parameters for the redundant description depend upon the primary description used as reference.

- *Decoder-based error concealment:* At very high loss probabilities, the number of lost primary slices is too large and Wyner-Ziv decoding fails for some or all video frames. In this case, the non-normative error concealment scheme [26] included in the reference JVT codec is used to conceal the lost primary slices.
- *Erasure channel:* The video data travel to the decoder in the form of RTP packets. The application layer is assumed to be unaware of the particulars of the error control mechanisms used in the transport and link layers. If these mechanisms succeed, then a transmitted packet is correctly received and forwarded to the application layer. If they fail, then the application layer receives a notification that the packet is lost. Thus, irrespective of whether a video packet is (1) corrupted or (2) arrives too late, after its specified deadline or (3) dropped altogether due to congestion, the video decoder treats the phenomenon as an erasure.
- *Buffering delay:* From the perspective of buffering delay, it is best to buffer the redundant slices belonging to only one frame before applying Reed-Solomon encoding/decoding. However, this results in an inefficient Slepian-Wolf code. From the perspective of Slepian-Wolf coding efficiency, it is best to divide a frame into very small slices, so that K and N can be made large. However, this introduces a very large header overhead and compromises the source coding efficiency of the video coder. In practice, it is difficult to characterize the effect of the slice size and buffering delay on the decoded picture quality and hence to find their optimum values. Besides, our main goal is not to optimize slice size and buffering delay but to compare the error resilience of SLEP with that of traditional FEC, in which both schemes use exactly the same primary (systematic) video description. In the following experiments, the buffering delay is set to 1/3 second, based on recommendations from members of the Joint Video Team. Thus, for a sequence encoded at 30 frames/s, the encoder (decoder) buffers the redundant slices belonging to 10 frames before Reed-Solomon encoding (decoding).
- *Reed-Solomon coding:* Since the length of the primary slices is fixed, the number of primary slices, and hence the

number of the redundant slices per frame, varies depending upon the scene content and quantization parameter chosen by the rate control algorithm. Thus N and K can both vary during each buffering interval. N is chosen such that the constraints on the allowable Wyner-Ziv bit rate and the total transmitted bit rate are satisfied.

B. Comparison of SLEP with FEC

The accuracy of the model described in Sec. IV is verified in Fig. 7 by comparing the average received picture quality predicted by (11), with that obtained from experiments using the scheme in Fig. 3. It is now possible to compare the robustness of SLEP and FEC. Packets are erased randomly and uniformly, with a given constant loss probability and the decoded picture quality delivered by three schemes is compared: (a) A SLEP scheme, labeled “SLEP-25”, in which the redundant slices are encoded at 25% of the bit rate of the primary slices and the Wyner-Ziv bit rate is 10% of the bit rate of the primary slices (b) FEC, in which the primary description is identical to the redundant description, and the parity bit rate generated by the Reed-Solomon encoder is 10% of the bit rate of the primary slices (c) The non-normative H.264/AVC decoder-based error concealment scheme, labeled “EC”, in which no parity symbols are transmitted and all the available bit rate is allocated to the primary video signal. It is emphasized that the redundant slices are not transmitted, and therefore the total transmitted bit rate for the FEC and SLEP schemes is equal. The following properties of SLEP are observed:

- 1) At low packet loss rates, SLEP incurs a small reduction in average PSNR when compared with FEC, because Wyner-Ziv decoding introduces quantization mismatch between the primary and redundant slices. However, this mismatch does not result in unpleasant visual artifacts.
- 2) As the packet loss percentage increases, the end-to-end distortion is dominated by error-concealment artifacts resulting from the failure of FEC and Wyner-Ziv decoding. Recall that while both FEC and SLEP have the same error resilience bit rate, FEC applies error protection to the primary slices, while SLEP applies error protection to the redundant slices which are coarsely quantized. Therefore, for the given bit budget for error

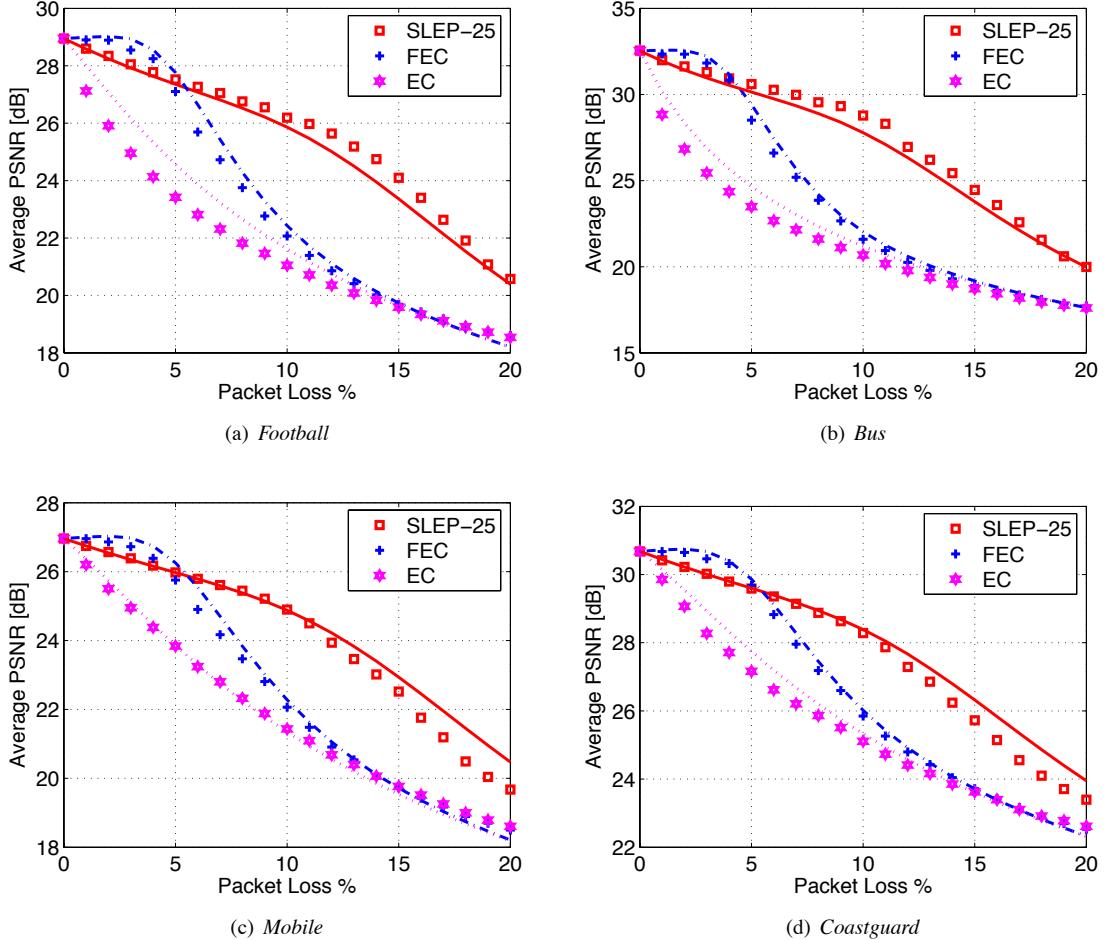


Fig. 7. Comparison of the robustness of SLEP and FEC to randomly inserted channel errors. The data points correspond to experimental simulation using the implementation of Fig. 3. The curves correspond to the values predicted by the model derived in Section IV. When the packet loss percentage is increased, SLEP provides gracefully degrading picture quality compared to FEC.

resilience, Wyner-Ziv protection in SLEP is stronger than conventional FEC protection. Thus, at high packet loss percentage, the average PSNR of SLEP is higher than that of FEC (Fig. 7).

- 3) Next, consider the instantaneous variation of the PSNR of the decoded video frame. In Fig. 8, the PSNR is plotted against the frame number for an experiment carried out with 10% packet loss. An arbitrarily chosen 200-frame window, consisting of frame numbers 651–850 from each simulation is displayed in the plots. Note that, in the case of Reed-Solomon codes with infinitely long block lengths, allocating 10% of the bit rate for parity information would be sufficient to provide FEC protection at 10% packet loss. However, owing to the use of finite block lengths in practical transmission systems, this allocated parity bit rate cannot always provide erasure protection. Thus, FEC fails in a few instances, and this results in a reduction in the frame PSNR. This effect is aggravated by error propagation, and is manifested as a drastic reduction of the *average* PSNR in Fig. 7. In contrast, as explained earlier, SLEP has stronger error protection at the same error resilience bit rate, because it uses smaller redundant descriptions.

Due to the quantization mismatch between the primary and redundant slices, the frame PSNR after successful SLEP decoding is slightly lower than that in the error-free case, but a drastic reduction in picture quality is avoided. For visual comparison, a video frame from both sequences is shown in Figs. 9 and 10, for a trace in which 10% of the packets are erased. It is observed that the subjective degradation due to the quantization mismatch from Wyner-Ziv decoding is not as severe as the degradation from error concealment artifacts.

C. Effect of Increasing Wyner-Ziv Bit Rate

In this experiment, a scheme labeled “SLEP-50” is considered, in which the bit rate of the redundant description is 50% of that of the primary video signal. For this fixed redundant description, the robustness increases when the Wyner-Ziv bit rate is increased from 10% to 20% of the bit rate of the primary video signal, as shown in Fig. 11. This trend is similar to that observed in conventional FEC-based systems, in which robustness increases when a stronger channel code is used. The performance of the equivalent FEC schemes is shown for comparison in Fig. 11.

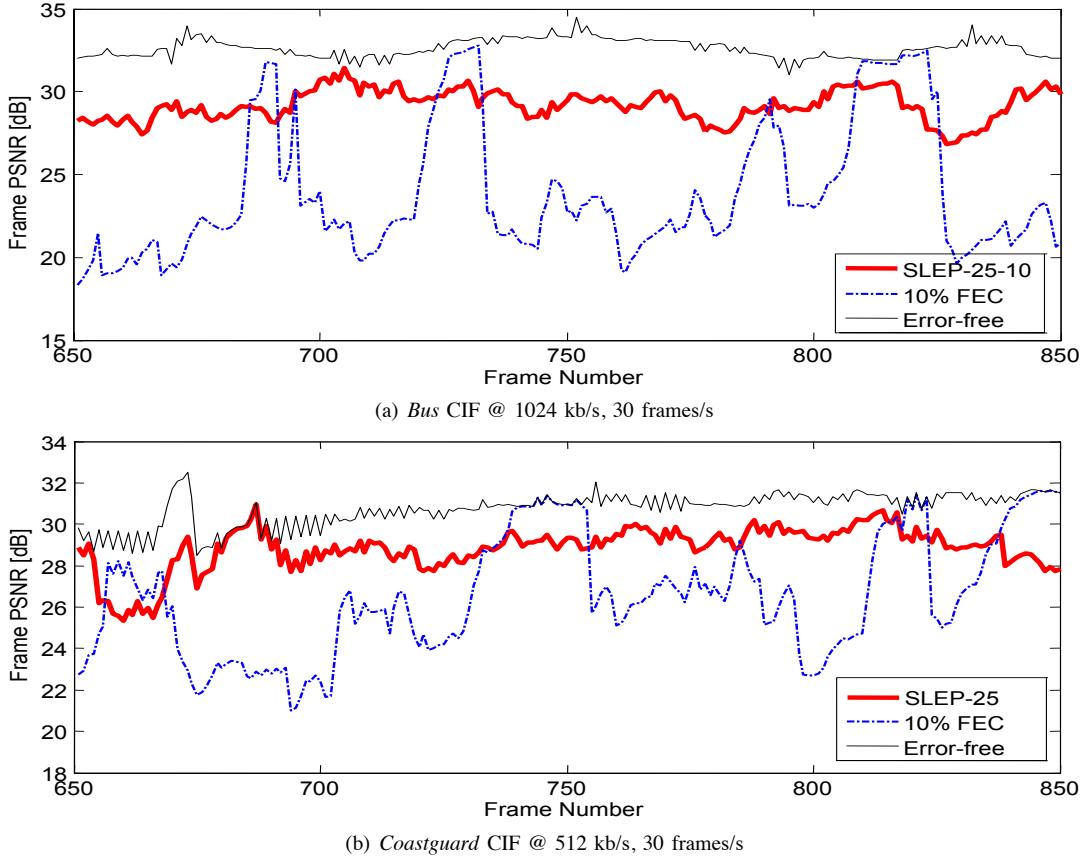


Fig. 8. When coarse quantization is used in the redundant description, there is a small reduction in the decoded frame PSNR compared to the error-free case. In return, drastic reduction in picture quality is avoided. At a high packet loss rate of 10%, SLEP-25-10 provides the smallest instantaneous fluctuation in frame PSNR, followed by SLEP-50-10 followed by FEC.



Fig. 9. Decoded frames of the *Bus CIF* sequence encoded at 1024 kb/s, when the parity bit rate is 10% of the primary source coding bit rate. With FEC (left), decoding fails for some portions of the frame, reducing the frame PSNR to 20.2 dB. With SLEP scheme for a redundant description encoded at 25% of the primary bit rate (right), successful Wyner-Ziv decoding results in a frame PSNR of 30.3 dB, much closer to the error-free PSNR of 32.2 dB.

D. Performance of Optimized SLEP

In this section, the model of Section IV is used to select the bit rates for encoding the primary and redundant slices as well as the Reed-Solomon parity slices, such that the average received PSNR is maximized. The optimization problem in (14) is solved at the encoder and is repeated after every second. This requires the parameter estimation in (12) to be repeated every second, but ensures that the rate-distortion functions used in the optimization reflect the changes in the scene

content. Similarly, the parameter estimation in (13) has to be repeated every second, but this latter parameter estimation can be performed at low complexity because the trial encodings for the redundant description only involve requantization followed by entropy coding, and do not require repeating the motion estimation algorithm. The optimization is performed using an exhaustive search, while varying \mathcal{R}_p and \mathcal{R}_r in steps of 50 kb/s subject to the constraints in (14). For all permissible pairs $(\mathcal{R}_p, \mathcal{R}_r)$, the encoder determines the number of redundant



Fig. 10. Decoded frames of the *Coastguard* CIF sequence encoded at 512 kb/s, when the parity bit rate is 10% of the primary source coding bit rate. With FEC (left), decoding fails for some portions of the frame, reducing the frame PSNR to 22.7 dB. With SLEP scheme for a redundant description encoded at 25% of the primary bit rate (right), successful Wyner-Ziv decoding results in a frame PSNR of 31.2 dB, much closer to the error-free PSNR of 32.9 dB.

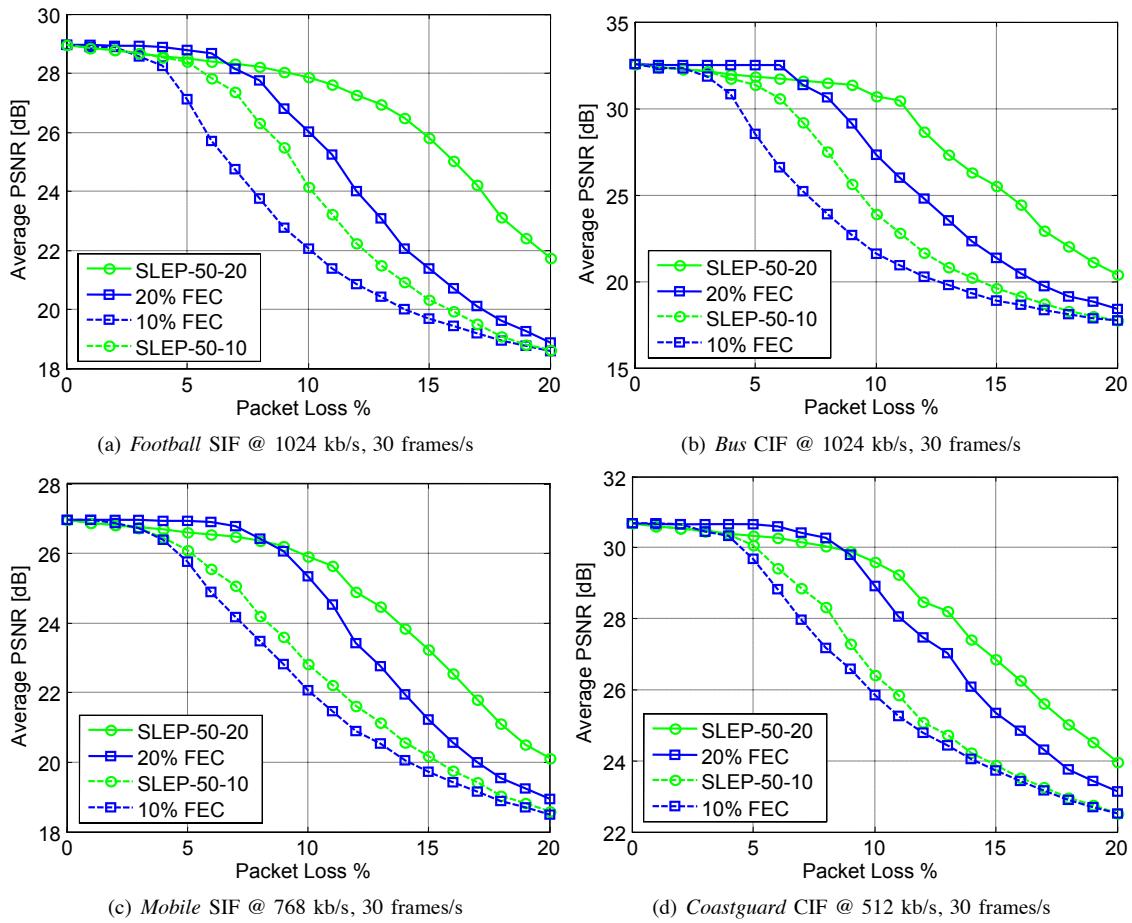


Fig. 11. With a fixed redundant description, increasing the Wyner-Ziv bit rate results in an increase in error resilience. FEC with the same parity bit rate is displayed for comparison.

slices K_L and calculates $\mathcal{R}_{WZ} = R_T - \mathcal{R}_p$. From K_L and \mathcal{R}_{WZ} , the encoder calculates N where $N - K_L$ is the number of parity slices. Finally, for this bit rate assignment, the end-to-end MSE distortion \mathcal{D} is calculated. The values of K_L and N depend upon the encoding bit rates. As an example, for the *Coastguard* sequence encoded at 512 kb/s at 30 frames/s with 500 bits/slice and a Wyner-Ziv bit rate of 51 kb/s, the average

values of K_L and N are 42 and 46 respectively. To simulate packet losses in Internet video transmission, we use error patterns generated from real internet measurements [27] at 3%, 5%, 10% and 20% packet loss. These error patterns are the common test conditions prescribed by the Joint Video Team for low-delay error resilience experiments [28]. In Fig. 12, the average video quality delivered by the optimized SLEP scheme

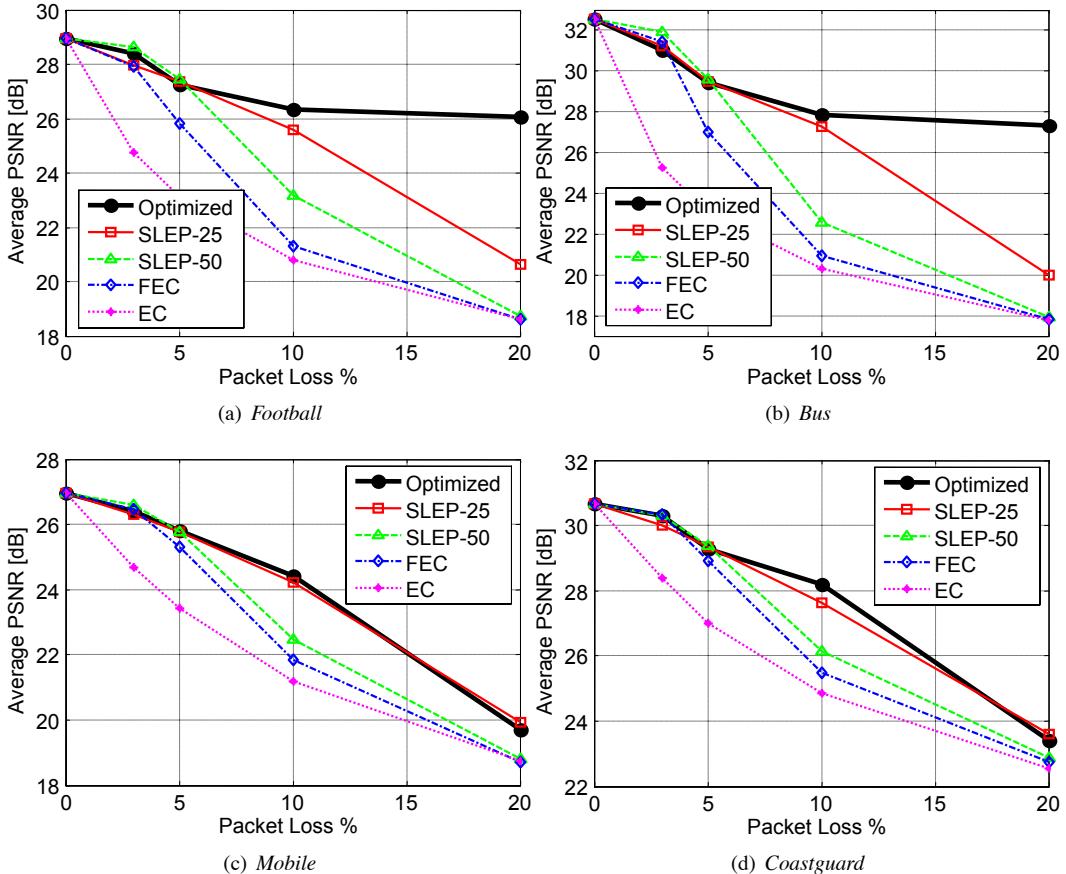


Fig. 12. Comparison of the robustness of optimized and unoptimized SLEP schemes. Using the model derived in Section IV, the bit rates R_p , R_r and R_{WZ} are selected such that the average MSE distortion \mathcal{D} is minimized. Moreover, the bit allocation is updated every second to account for changes in the scene content.

is plotted against the packet loss percentage. For comparison, the plots also include unoptimized SLEP schemes which use primary slices encoded at the bit rates given in Table I and redundant slices encoded at 25%, 50%, 100% of the bit rate of the primary slices. The error resilience bit rate in each of the unoptimized schemes is fixed to 10% of the primary slices bit rate. As observed from the plots, the optimized SLEP scheme provides the highest decoded video quality². The superior performance of the optimized SLEP scheme can be attributed to a combination of two effects: (a) The optimization of (14) forces $R_{WZ} > 10\%$ of R_p at the expense of a reduction in R_p , and (b) For a given R_p and R_{WZ} , the optimization scheme forces $R_r < 25\%$ of R_p . Both these actions result in the generation of a larger number of parity slices, which, in turn, increases the robustness to packet loss.

VI. CONCLUSIONS

A scheme for robust video transmission, known as Systematic Lossy Error Protection (SLEP) has been studied. Robustness to transmission errors is provided by transmitting a Wyner-Ziv encoded version of the video signal. When parts

²It is observed that, at some loss rates, the schemes in Fig. 12 have slightly inferior video quality than those in Fig. 7 at the same packet loss rates. The reason is that the experiments in Fig. 7 had uniform packet loss while those in Fig. 12 use packet loss traces from internet experiments which are non-uniform (and can be larger than the average loss rate) over short time intervals.

of the video sequence are lost, the Wyner-Ziv bit stream allows the recovery of a coarsely quantized version of the lost portions. A Wyner-Ziv codec has been implemented under the H.264/AVC specification, using redundant slices in conjunction with Reed-Solomon coding. A model is proposed which relates the average decoded picture quality to the bit rates of the original encoded video signal, the Wyner-Ziv bit stream and the untransmitted coarsely quantized redundant signal that can be recovered at the receiver by Wyner-Ziv decoding. The model equations are used to optimally select these three bit rates and to update them after a specified time window, which may be chosen depending upon the scene content and activity.

Traditional FEC schemes attempt to achieve the highest visual quality by performing source/channel bit allocation. By performing Wyner-Ziv coding instead of conventional channel coding, SLEP provides an additional degree of freedom: It inserts a small, bounded distortion in the protected signal, and gracefully trades off this distortion against the resilience to packet loss. Up to the present time, the only scheme which achieved this graceful degradation of received video quality was layered coding with unequal error protection. Our work has shown that, using Wyner-Ziv coding, it is possible to achieve graceful degradation without a layered video representation in the systematic portion of the transmission.

Hence, the SLEP scheme does not incur the loss in rate-distortion performance associated with layered video codecs. Furthermore, a Wyner-Ziv codec can be constructed out of well-understood components: quantizers, entropy coders and channel coders, at a small additional complexity cost compared with conventional FEC-based systems.

APPENDIX: RESIDUAL DISTORTION AFTER WYNER-ZIV DECODING

In the following, the video distortion model is used to derive an expression for the minimum increase in video distortion that must be tolerated after Wyner-Ziv decoding. We are interested in the distortion due to the quantization mismatch only, and not the distortion from error concealment. Therefore, in the following, it is assumed that the *average*³ Wyner-Ziv bit rate is just large enough to ensure that Wyner-Ziv decoding is successful, at the erasure probability p encountered by the system.

With this assumption, $p_{WZ} \rightarrow 1$ and $p_{EC} \rightarrow 0$ in (10). Then, the average distortion for a GOP of length M frames is given by

$$\begin{aligned}\mathcal{D} &= \frac{1}{M} \sum_{n=1}^M \mathcal{D}_n^{EE} = \left(1 - \frac{M+1}{2} p\right) \mathcal{D}_p + \frac{M+1}{2} p \mathcal{D}_r \\ &= \mathcal{D}_p + \frac{M+1}{2} p (\mathcal{D}_r - \mathcal{D}_p) = \mathcal{D}_p + \Delta\end{aligned}$$

where Δ is the residual error energy due to the quantization mismatch. Clearly, to minimize the quantization mismatch between the primary and redundant descriptions, the encoding bit rate \mathcal{R}_r of the redundant description must be as close as possible to the primary description bit rate \mathcal{R}_p . The remaining bit rate, $R_T - \mathcal{R}_p$ is then allocated to the Wyner-Ziv bit stream. Thus,

$$\mathcal{R}_{WZ} = R_T - \mathcal{R}_p = \frac{N-K}{K} \mathcal{R}_r \simeq \frac{p}{1-p} \mathcal{R}_r$$

where the third expression indicates that the Wyner-Ziv bit rate depends upon the parameters of the Reed-Solomon code and the encoding bit rate of the redundant description. The last expression above assumes that N and K are large enough to ensure that the Reed-Solomon code operates at its maximum efficiency⁴. Now, the maximum allowable bit rate for encoding the redundant description is given by:

$$\mathcal{R}_r = \min \left((R_T - \mathcal{R}_p) \frac{1-p}{p}, \mathcal{R}_p \right)$$

Thus, at packet erasure probability p , a redundant description encoded at bit rate \mathcal{R}_r increases the MSE distortion by:

$$\Delta = p \frac{M+1}{2} (\mathcal{D}_r - \mathcal{D}_p) \quad (15)$$

³The Wyner-Ziv bit rate, i.e., the bit rate of the parity slices, changes slightly over the duration of a video sequence because the number of redundant slices per frame is not constant. In this section, we consider the average Wyner-Ziv bit rate to simplify the analysis.

⁴This assumption is made only in the Appendix, the goal being to investigate the tradeoff between the quantization mismatch and the robustness while being oblivious to other design considerations. In the remainder of this paper, the inefficiency associated with the use of short block lengths in the Reed-Solomon coder is captured in the term p_{WZ} , which is evaluated in (5).

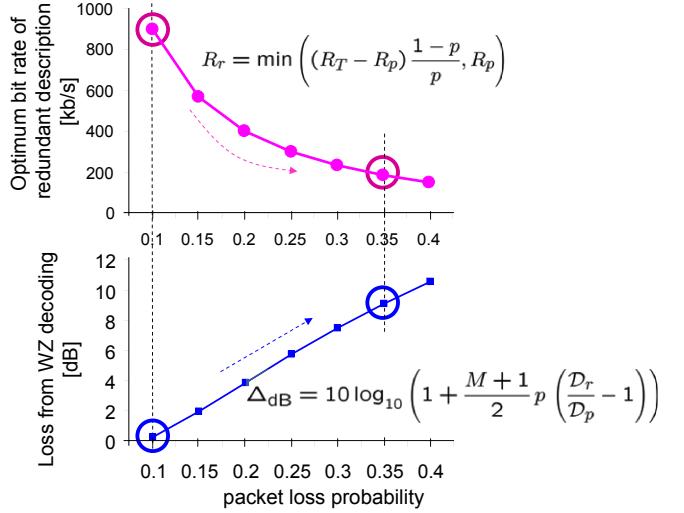


Fig. 13. As the erasure probability increases, redundant descriptions encoded at a lower bit rate must be used to provide robustness to loss. The increased resilience is achieved at the cost of increased quantization mismatch after Wyner-Ziv decoding.

where \mathcal{D}_r and \mathcal{D}_p depend on \mathcal{R}_r and \mathcal{R}_p through (12) and (13). The drop in video quality in dB, resulting from the usage of the redundant description rather than the primary description, is given by:

$$\begin{aligned}\Delta_{dB} &= 10 \log_{10} \frac{255^2}{\mathcal{D}_p} - 10 \log_{10} \frac{255^2}{\mathcal{D}_p + \Delta} = 10 \log_{10} \frac{\mathcal{D}_p + \Delta}{\mathcal{D}_p} \\ &= 10 \log_{10} \left(1 + \frac{M+1}{2} p \left(\frac{\mathcal{D}_r}{\mathcal{D}_p} - 1 \right) \right)\end{aligned}$$

where Δ is obtained from (15). Fig. 13 plots this loss in dB, at various packet erasure rates, for $R_T = 1.1$ Mb/s, $\mathcal{R}_p = 1$ Mb/s, and $M = 15$ for the *Foreman CIF* sequence. The plots indicate that error resilience at high erasure probability is achieved at the price of increased distortion from the quantization mismatch between the redundant and primary descriptions. Observe that the residual distortion also depends on the number of frames over which the quantization mismatch propagates due to motion-compensated decoding. After M frames, the quantization error propagation is stopped by an intra coded video slice. This is reminiscent of the theoretical analysis of SLEP for a first-order Markov source [4], in which the residual distortion after Wyner-Ziv decoding was expressed as a function of the erasure probability, the quantization mismatch and the error propagation resulting from the temporal correlation in the source samples.

ACKNOWLEDGMENTS

This work has been supported in part by NSF Grant No:CCR-0310376. P. Baccichet was supported by the Max Planck Center for Visual Computing and Communication at Stanford University.

REFERENCES

- [1] P. Baccichet, S. Rane, and B. Girod, "Systematic Lossy Error Protection using H.264/AVC Redundant Slices and Flexible Macroblock Ordering," *Journal of Zhejiang University, Science*, vol. 7, no. 5, pp. 727–736, May 2006.

- [2] S. Rane, P. Baccichet, and B. Girod, "Modeling and Optimization of a Systematic Lossy Error Protection System," in *Picture Coding Symposium*, Beijing, China, Apr. 2006.
- [3] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed Video Coding," *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, pp. 71–83, Jan. 2005.
- [4] S. Rane, D. Rebollo-Monedero, and B. Girod, "High-Rate Analysis of Systematic Lossy Error Protection of a Predictively Encoded Source," in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2007, pp. 263–272.
- [5] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
- [6] S. Shamai, S. Verdú, and R. Zamir, "Systematic Lossy Source/Channel Coding," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.
- [7] Y. Steinberg and N. Merhav, "On Hierarchical Joint Source-Channel Coding with Degraded Side Information," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 886–903, Mar. 2006.
- [8] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv Video Coding with Hash-based Motion Compensation at the Receiver," in *Proc. IEEE International Conference on Image Processing*, vol. 5, Singapore, Oct. 2004, pp. 3097–3100.
- [9] A. Aaron, D. Varodayan, and B. Girod, "Wyner-Ziv Residual Coding of Video," in *Proc. Picture Coding Symposium*, Beijing, China, Apr. 2006.
- [10] R. Puri and K. Ramchandran, "PRISM: A New Robust Video Coding Architecture based on Distributed Compression Principles," in *Proc. Allerton Conference on Communication, Control, and Computing*, Allerton, IL, Oct. 2002.
- [11] A. Jagmohan, A. Sehgal, and N. Ahuja, "Predictive Encoding using Coset Codes," in *Proc. IEEE International Conference on Image Processing*, vol. 2, Rochester, NY, Sept. 2002, pp. 29–32.
- [12] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, *Advanced Video Coding for Generic Audiovisual services, ITU-T Recommendation H.264 - ISO/IEC 14496-10(AVC)*. ISO/IEC MPEG & ITU-T VCEG, Oct. 2003.
- [13] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Transactions on Information Theory*, vol. IT-19, pp. 471–480, July 1973.
- [14] S. Rane, P. Baccichet, and B. Girod, "Systematic Lossy Error Protection based on H.264/AVC Redundant Slices and Flexible Macroblock Ordering," in *Input documents to 19th ITU-T/MPEG JVT Meeting*, no. JVT-S025, Geneva, Switzerland, Apr. 2006.
- [15] S. Rane, "Systematic Lossy Error Protection of Video Signals," Ph.D. Dissertation, Stanford University, 2007.
- [16] Z. He and S. Mitra, "Optimum Bit Allocation and Accurate Rate Control for Video Coding via Rho-domain Source Modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 840–849, Oct. 2002.
- [17] T. Chiang and Y. Q. Zhang, "A New Rate Control Scheme using Quadratic Rate-Distortion Model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–250, Feb. 1997.
- [18] J. Ribas-Corbera and S. Lei, "Rate Control in DCT Video Coding for Low-Delay Communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172–185, Feb. 1999.
- [19] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of Video Transmission over Lossy Channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–1032, June 2000.
- [20] S. Rane, P. Baccichet, and B. Girod, "Systematic Lossy Error Protection based on H.264/AVC Redundant Slices," in *Input documents to 21st ITU-T/MPEG JVT Meeting*, no. JVT-U057, Hangzhou, China, Oct. 2006.
- [21] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-Distortion Optimization for JVT/H.26L Video Coding in Packet Loss Environment," in *Proceedings of the 2002 Packet Video Workshop*, Pittsburgh, PA, Apr. 2002.
- [22] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority Encoding Transmission," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.
- [23] S. Rane and B. Girod, "Systematic Lossy Error Protection versus Layered Coding with Unequal Error Protection," in *Proc. SPIE Image and Video Communications and Processing*, vol. 5685, San Jose, CA, Jan. 2005, pp. 663–671.
- [24] S. Rane, P. Baccichet, and B. Girod, "Systematic Lossy Error Protection based on H.264/AVC Redundant Slices," in *Input documents to 20th ITU-T/MPEG JVT Meeting*, no. JVT-T093, Klagenfurt, Austria, July 2006.
- [25] K. Suehring, "JVT JM reference software home page (online)," <http://iphome.hhi.de/suehring/tm1/>.
- [26] Y. Wang, M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj, "The Error Concealment Feature in the H.26L Test Model," in *Proc. IEEE International Conference on Image Processing*, vol. 2, Rochester, NY, Sept. 2002, pp. 729–732.
- [27] S. Wenger, "Error Patterns for Internet Experiments," VQEG Q15-I-16r1, 2002.
- [28] ——, "Common Test Conditions for Wireline Low Delay IP/UDP/RTP Packet Loss-Resilient Testing," ITU-T SG16, VQEG-N79r1, 2001.



Shantanu Rane received the Ph.D. degree in Electrical Engineering from Stanford University in 2007. He has a B.E. degree in Instrumentation Engineering from the Government College of Engineering, Pune University, India and a M.S. degree in Electrical Engineering from the University of Minnesota, Minneapolis. In 1998–99, he worked at the National Center for Radio Astrophysics (Tata Institute of Fundamental Research) in Pune. His research interests include distributed source coding, theory and practice of video compression, and image communication in error-prone environments. He is currently working at Mitsubishi Electric Research Laboratories in Cambridge, MA.



Pierpaolo Baccichet received the M.S and the Ph.D. degree in Computer Science from the University of Milan, Italy, in 2001 and 2006 respectively. During his Ph.D. studies, he worked on error concealment algorithms for whole-frame loss and low-complexity protection schemes for error resilient video transmission at the National Research Center of Italy in Turin, funded by ST Microelectronics. At present, he is a visiting assistant professor at Stanford University with the Max Planck Center for Visual Computing and Communication. His research interests include error resilient video coding and real-time streaming over peer-to-peer networks.



Bernd Girod is Professor of Electrical Engineering and (by courtesy) Computer Science in the Information Systems Laboratory of Stanford University, California, since 1999. Previously, he was Professor of Telecommunications in the Electrical Engineering Department of the University of Erlangen-Nuremberg. His current research interests are in the areas of video compression and networked media systems. He has published over 400 conference and journal papers, as well as 5 books, receiving the EURASIP Best Paper Award in 2002 and the IEEE Multimedia Communication Best Paper Award in 2007, as well as the EURASIP Technical Achievement Award in 2004. As an entrepreneur, Professor Girod has been involved with several startup ventures as founder, director, investor, or advisor, among them Polycom (Nasdaq:PLCM), Vivo Software, 8x8 (Nasdaq: EGBT), and RealNetworks (Nasdaq: RNWK). He holds an Engineering Doctorate from University of Hannover, Germany, and an M.S. Degree from Georgia Institute of Technology. Prof. Girod is a Fellow of the IEEE and a member of the German Academy of Sciences (Leopoldina).