Analysis of Packet Loss for Compressed Video: Effect of Burst Losses and Correlation Between Error Frames

Yi J. Liang, John G. Apostolopoulos, *Fellow, IEEE*, and Bernd Girod, *Fellow, IEEE*

***Abstract*—Video communication is often afflicted by various forms of losses, such as packet loss over the Internet. This paper examines the question of whether the packet loss pattern, and in particular, the burst length, is important for accurately estimating the expected mean-squared error distortion resulting from packet loss of compressed video. We focus on the challenging case of low-bit-rate video where each P-frame typically fits within a single packet. Specifically, we: 1) verify that the loss pattern does have a significant effect on the resulting distortion; 2) explain why a loss pattern, for example a burst loss, generally produces a larger distortion than an equal number of isolated losses; and 3) propose a model that accurately estimates the expected distortion by explicitly accounting for the loss pattern, inter-frame error propagation, and the correlation between error frames. The accuracy of the proposed model is validated with H.264/AVC coded video and previous frame concealment, where for most sequences the total distortion is predicted to within ± 0.3 dB for burst loss of length two packets, as compared to prior models which underestimate the distortion by about 1.5 dB. Furthermore, as the burst length increases, our prediction is within ± 0.7 dB, while prior models degrade and underestimate the distortion by over 3 dB. The proposed model works well for video-telephony-type of sequences with low to medium motion. We also present a simple illustrative example, of how knowledge of the effect of burst loss can be used to adapt the schedule of video streaming to provide improved performance for a burst loss channel, without requiring an increase in bit rate.**

***Index Terms*—Distortion modeling, error propagation, error resilience, H.264/AVC, packet loss, rate-distortion optimization.**

I. Introduction

VIDEO communications over bit-rate-limited and error-prone channels, such as packet networks and wireless links, require both high compression and high error resilience. Important applications within this context include video streaming over the Internet and wireless video to handheld devices such as with the emerging Third Generation (3G) cellular system. To achieve high compression, most current video compression systems employ motion-compensated prediction between frames to exploit the temporal redundancy, followed by a spatial transform to exploit the spatial redundancy, and the resulting parameters are entropy-coded to produce the compressed bitstream. These algorithms provide significant compression, however the compressed signal is highly vulnerable to losses when transmitted over error-prone channels. Probably the most important problem that afflicts video compressed with these coders is the *error propagation problem*. Specifically, inter-frame prediction provides significant compression, however, the decoder must have the same reference frame as used by the encoder in order to perform correct decoding. A channel error can cause the reconstructed frame at the decoder to be incorrect, which can lead to significant error propagation to subsequent frames.

The problem of error-resilient video communication has received significant attention in recent years, and a variety of techniques have been proposed to mitigate the effects of packet loss and inter-frame error propagation, and thereby to increase the robustness of video communication over lossy networks [1]–[7]. Examples of recent work in this area includes intra/inter-mode switching [8]–[10], dynamic control of prediction dependency using multiframe memory [11]–[13], forward error correction (FEC) [14]–[16], channel-adaptive packet scheduling [17]–[20], and the use of multiple description coding and packet path diversity [21]–[25]. These approaches for reliable or error-resilient video communication use models of the effect of losses on the reconstructed video to motivate and direct the design of the different approaches, and also to adapt their operation to transmission conditions. In addition, many of these recent algorithms use rate-distortion (R-D) optimization techniques to improve the performance over lossy channels. The goal of these optimization techniques is to minimize the expected distortion due to both compression and channel losses subject to the bit-rate constraint. Their performance crucially depends on the accuracy with which they can predict the distortion that results for different loss events. Therefore, it is very important to have accurate models for predicting the distortion resulting from packet loss.

Developing an accurate model for the effect of packet loss on the reconstructed video quality is critical for designing and accurately predicting the performance of video communication systems. An important question along these lines is whether the expected distortion as seen by the client depends only on the average packet loss rate, or whether it also depends on the specific pattern of the loss. For example, does packet loss burst length matter, or is the resulting distortion equivalent to an equal number of isolated losses?

Manuscript received November 18, 2003; revised February 6, 2007. This paper was recommended by J. Arnold.

Y. J. Liang is with Mobim Technologies, Shanghai 201203, China (e-mail: yiliang@stanfordalumni.org).

J. G. Apostolopoulos is with the Mobile and Media Systems Lab, HP Labs, Palo Alto, CA 94304 USA (e-mail: japos@hpl.hp.com).

B. Girod is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: bgirod@stanford.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2008.923139

1051-8215/$25.00 © 2008 IEEE

Most of the prior work on modeling the effect of losses model the expected distortion as being proportional to the number of losses that occur [26]–[28]. For example, [26] and [27] analyze and model the distortion for a single (isolated) loss and model the mean-squared error (MSE) expected distortion for multiple losses as being proportional to the number of losses that occur, e.g., the average packet loss rate. This model accounts for error propagation, intra-refresh rate, and spatial filtering performed by the motion compensation loop of the decoder. In this model, linearity and superposition are assumed for multiple errors, specifically, the total distortion increases linearly with the average packet loss rate. Therefore, in this model, the estimated distortion does not depend on the specific loss pattern or the burst length of a loss, and only depends on the number of losses that occur. Such approaches focus on the average packet loss rate as the most important parameter to consider, and implicitly assume that burst length does not matter. This model is accurate when the losses can be considered to have independent effects, for example, when single (isolated) losses are spaced sufficiently far apart (with respect to the intra-update interval). This may be true when the losses are isolated and not bursty, or when each frame is coded into many packets and any loss event is restricted to losing packets of a single frame. However, in many important communication situations for example, video communication over the Internet or over a wireless link, the losses may be bursty.

The effect of burst losses is particularly pronounced for low bit rate video. For example, in the practically important and challenging case of low bit-rate video communication over lossy wired or wireless packet networks, a burst loss generally leads to the loss of multiple video frames, e.g., QCIF resolution video (144×176 pixels/frame) coded with a conventional H.263, MPEG-4, or an H.264/AVC video coder at less than about 150 kb/s where each P-frame can fit in a single packet. These cases of low bit rate video over a lossy packet network are both practically important and technically quite challenging and have lead to significant research on error-resilient video coding. On the other hand, the effect of burst losses are generally much less pronounced for high bit rate video, for instance MPEG-2 video at 4 Mb/s (e.g., [29]), where each frame may be coded into about 10 packets, and, therefore, much longer burst losses are required to have similar effects.

A model for estimating the overall end-to-end distortion for pre-encoded video is proposed in [30], [31] to aid R-D optimized streaming, by using partial derivative approximations on a limited order. In [32], a statistical model to estimate channel error induced distortion for different channel conditions is proposed for wireless video coding, based on a theoretical analysis of the distortion caused by channel errors as well as inter-frame error propagation. In [29] and [33], the quality of compressed video transmitted over a packet network is monitored from the perspective of a network service provider, by extracting sequence-specific information including spatio–temporal activity and the effects of error propagation. All of these schemes use an overall packet loss rate and do not explicitly consider the pattern of the losses.

In [21] the length of a burst loss was shown to have an important effect on the resulting distortion, where longer burst lengths generally led to larger distortions. Furthermore, the effect of a burst loss was also identified as an important feature for comparing the relative merits of different error-resilient coding schemes. This was extended in [34] where a simple model was proposed that distinguishes loss events based on the length of the burst loss (e.g., single loss of length one, burst loss of length two, burst loss of length three) and explicitly accounts for the different distortions that result for different burst runlengths. This model provides some improvement over the prior additive model in the sense that it accurately accounts for the different effects of burst losses as compared to isolated losses. It also provides a simple mechanism that accounts for the different distortions that result for different burst lengths. However, this model also shares some of the disadvantages of the prior additive model. For example, it does not account for more general loss patterns, such as two losses spaced apart by a short lag.

An understanding of the effect of packet loss on the reconstructed video quality and developing accurate models for predicting the distortion for different loss events, is clearly very important for designing, analyzing, and operating video communication systems over lossy networks. This paper examines the question of whether the loss pattern, and in particular the burst length, is important for accurately estimating the expected distortion. Understanding that the effects of a single packet loss and burst losses may be very different, we study and model the distortion resulting from more general and complex loss patterns, including burst losses and losses separated by a certain lag. We: 1) verify that the packet loss pattern does, in fact, have a significant effect on the resulting distortion; 2) explain why a loss pattern, for example, a burst loss, generally produces a larger distortion than an equal number of isolated losses; and 3) propose a model that accurately estimates the expected distortion by explicitly accounting for the loss pattern. To estimate the expected distortion the proposed model explicitly considers the effect of different loss patterns, including burst losses and separated (non-consecutive) losses spaced apart by a lag, and accounts for inter-frame error propagation and the correlation between error frames. The proposed model provides a significantly more accurate estimate of the mean-squared error distortion resulting from different loss events, compared to prior models. The accuracy of the proposed model is validated for four standard video test sequences coded with the H.264/AVC standard.

This paper is structured as follows. Section II presents the proposed model, and specifically focuses on the cases of burst losses and separated (non-consecutive) losses spaced apart by some lag. Experimental results that illustrate and validate the accuracy of the proposed model are presented in Section IV. We conclude in Section V by presenting a simple illustrative example of how knowledge of the effect of burst loss can be used to adapt the packet scheduling to provide improved performance for burst loss channels.

II. Loss Modeling Considering Error Correlation

The goal of this section is to develop models that accurately estimate the distortion for more general loss patterns, including burst losses and separated (non-consecutive) losses where the separation is less than that required to make the losses independent.

Throughout this paper we consider an H.264/RTP/UDP/IP scenario, where the H.264 packetization is performed such that packets are independently decodable. The packetized data are encapsulated into RTP payloads and delivered over the IP network through RTP/UDP. At the receiver side packet loss is identified through the sequence number.

We also assume that each predictively coded frame (P-frame) is coded into a single packet, so that the loss of a packet corresponds to the loss of an entire frame. This corresponds, for example, to the practically important case of low bit rate video communication over lossy wired or wireless packet networks, e.g., QCIF resolution video coded with an H.263, MPEG-4, or an H.264/AVC video coder, and packetized with the conventional packet framing option of sending each new coded frame in a new packet. The results in this paper can also be extended to the case where each frame is coded into multiple packets, by accounting for the portion of each coded frame that depends on each packet and the longer burst loss required to lose an entire frame.

The original video signal is a discrete space-time signal denoted by $s[x, y, k]$, where $k \in \mathcal{Z}$ is the frame index. To simplify notation, the 2-D array of $M = M_1 \times M_2$ pixels in each frame k are sorted in the 1-D vector $f[k]$ (of length M) in line-scan order. We use the 1-D vector $f[k]$ to represent an original video frame, $\widehat{f}[k]$ to denote the loss-free reconstruction of the frame, and $g[k]$ to denote the reconstruction at the decoder after possible loss concealment. An error frame introduced by a channel loss is defined as

$$e[k] = g[k] - \widehat{f}[k] \tag{1}$$

which is also a 1-D vector. Since our primary concern is the effect of channel loss, quantization error is not included in the error signal being studied. Assuming the error frame $e[k]$ to be a stationary process, its mean-squared error (MSE) is given by

$$\frac{(e^T[k] \cdot e[k])}{M} = d[k]. \tag{2}$$

The distortion that would result from a single loss, as a function of the specific frame that it afflicts, can be evaluated at the encoder by simulating the corresponding loss event and decoding the sequence. Note that these distortions can be straightforwardly computed and stored, and we refer to these distortions as "pre-measured" distortions in the remainder of this work. We will show that by using these pre-measured distortions, we are able to accurately estimate the distortion from more general loss patterns using the models proposed in this work. We denote the initial error frame resulting from a *single* lost frame k by $e_S[k]$, and its MSE by $d_S[k]$; while $e[k]$ and $d[k]$ are used to represent the error frame and the MSE at time k resulting from a more general loss pattern.

While the MSE above for each individual frame quantifies the initial error power introduced by a channel loss, it does not include the effect of error propagation. In order to describe the overall effect of a particular loss event, we also define *total distortion* to be the sum of the MSEs of all the frames during the entire error recovery period, given by

$$D[\mathcal{K}] = \sum_{i=k_1}^{\infty} d[i] \tag{3}$$

where $\mathcal{K} = \{k_1, k_2, \ldots\}$ with k_1 being the index of the first lost frame, denotes the indices of the frames in error, and ∞ indicates the sequence should be long enough to cover the entire recovery period. Correspondingly $D_S[k]$ is used to denote the total distortion for a single frame loss at frame k. In Section II-A–C, we study the MSE and the total distortion for different loss patterns.

A. Burst Losses of Length Two

With the notations defined above, we first study burst losses of length two.

1) MSE of the Lost Frame: Assuming a simple loss concealment scheme where the lost frame is replaced by the previous frame at the decoder output, the error frames of single losses at $k-1$ and k are, respectively, given by

$$e_S[k-1] = g[k-1] - \widehat{f}[k-1] = \widehat{f}[k-2] - \widehat{f}[k-1] \tag{4}$$

$$e_S[k] = g[k] - \widehat{f}[k] = \widehat{f}[k-1] - \widehat{f}[k]. \tag{5}$$

In the case of a burst loss of length two afflicting frames $k-1$ and k, the residual error of frame k is given by

$$e[k] = g[k] - \widehat{f}[k] = \widehat{f}[k-2] - \widehat{f}[k] \tag{6}$$

$$= e_S[k-1] + e_S[k] \tag{7}$$

where the last equality is recognized by considering the sum of (4) and (5). The corresponding MSE of the error frame k is

$$\begin{aligned} d[k] &= (\widehat{f}[k-2] - \widehat{f}[k])^T \cdot \frac{(\widehat{f}[k-2] - \widehat{f}[k])}{M} \\ &= (e_S[k-1] + e_S[k])^T \cdot \frac{(e_S[k-1] + e_S[k])}{M} \\ &= d_S[k-1] + d_S[k] + 2\rho_{k-1,k} \cdot \sqrt{d_S[k-1] \cdot d_S[k]} \end{aligned} \tag{8}$$

where

$$\rho_{k-1,k} = \frac{\frac{\left(e_S^T[k-1] \cdot e_S[k]\right)}{M}}{\sqrt{d_S[k-1] \cdot d_S[k]}}$$

is the correlation coefficient between error frames $k-1$ and k.

In (8), the distortion of a burst loss of length two is expressed as a function of the distortion of two single and independent losses. Note that the MSE of the loss-affected frame in (8) is not just the sum of the MSEs of two independent losses, unlike what the additive model predicts. Specifically, the first two terms in (8) express the distortion when the two error frames are uncorrelated, and the third term expresses the change that results when the two error frames are correlated. Note that as a simple loss concealment scheme of previous frame replacement is used, in this distortion model, the amount of motion from frame to frame is implicitly accounted for through $d_S[k]$ and $d[k]$.

2) Modeling of the Total Distortion: As defined above, the total distortion expresses the distortion of the lost frame and the subsequent error propagation. We model the error propagation process in a typical video decoder with a geometric attenuation factor resulting from spatial filtering, and a linear attenuation factor from Intra update. With an Intra update period of N, if a single error is introduced at k with an MSE of $d[k]$, the power of the propagated error at $k+l$ is given by

$$d[k+l] = \begin{cases} d[k] \cdot r^l \cdot \left(1 - \frac{l}{N}\right), & \text{for } 0 \leq l \leq N \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The attenuation factor r ($r < 1$) accounts for the effect of spatial filtering, and $1 - l/N$ for Intra update, in reducing the error power. It is assumed that the error is completely removed by Intra update after N frames. In this model, linear attenuation as a result of Intra update is intuitive since the update is assumed to apply in a periodic manner. In [26] and [27], a $1/(1 + r' \cdot k)$ factor, is used to model the attenuation effect of spatial filtering, derived from calculating the signal power spectrum density and modeling the spatial filter as a linear system. The geometric attenuation we use here is a simplification based on that for simpler computation to be introduced later. The simplification is based on the fact that usually the attenuation effect of the spatial filter is relatively weaker than that of the Intra update, as shown by simulations.

For a single error introduced at k, and considering an entire period that is sufficiently long for complete error recovery, the total distortion is

$$\begin{aligned} D_S[k] &= \sum_{i=k}^{\infty} d[i] = \sum_{i=0}^{N-1} r^i \left(1 - \frac{i}{N}\right) \cdot d_S[k] \\ &= \left[1 + r\left(1 - \frac{1}{N}\right) + r^2\left(1 - \frac{2}{N}\right) + \cdots \right. \\ &\qquad \left. + r^{N-1}\left(1 - \frac{N-1}{N}\right)\right] \cdot d_S[k] \\ &= \frac{r^{N+1} - (N+1)r + N}{N(1-r)^2} d_S[k] \\ &= \alpha \cdot d_S[k], \end{aligned} \tag{10}$$

where $d[k] = d_S[k]$ is the initial error power introduced at k, and $\alpha = D_S[k]/d_S[k]$ is the ratio between the total distortion and the MSE of frame k, where the error is introduced. In (10)r is a parameter describing how effective the spatial filter is in reducing the introduced error power, and is dependent on the strength of the loop filter of the codec and the power spectrum density (PSD) of the input error signal. Since the variation of r from frame to frame is low, it is assumed that, for a fixed error burst length, r (and α) is constant for the entire recovery period, and independent of frame index k. In Section II-B, we will discuss the effect of the error burst length and the shape of the PSD of the input error signal on r.

We now study the total distortion D of two losses at $k-1$ and k. According to (3)

$$\begin{aligned} D[k-1, k] &= \sum_{i=k-1}^{\infty} d[i] \\ &= d_S[k-1] + \sum_{i=0}^{N-1} r^i \left(1 - \frac{i}{N}\right) \cdot d[k] \\ &= d_S[k-1] + \alpha \cdot d[k]. \end{aligned} \tag{11}$$

According to (8) and (10), (11) can be re-expressed as

$$\begin{aligned} D[k-1, k] = d_S[k-1] &+ D_S[k-1] + D_S[k] \\ &+ 2\rho_{k-1,k} \cdot \sqrt{D_S[k-1] \cdot D_S[k]} \end{aligned} \tag{12}$$

which is again the sum of two uncorrelated total distortions, plus a cross-correlation term, plus the distortion for frame $k-1$. Specifically, the cross-correlation term distinguishes the proposed model in this work from the previous additive model. The total distortion of a burst of losses is not only the sum of the distortion of independent losses at the same locations, but also largely affected by the correlation between the error frames.

B. Burst Losses of Length Greater Than Two

In Section II-A we developed analytical expressions for modeling the total distortion for a burst loss of length two. We now extend this to a model for a burst loss of length $B (B \geq 2)$. With the loss of B consecutive frames from $k - B + 1$ to k,

$$e[k] = \hat{f}[k-B] - \hat{f}[k] = \sum_{i=k-B+1}^{k} e_S[i],$$

and its MSE

$$\begin{aligned} d[k] =& (\hat{f}[k-B] - \hat{f}[k])^T \cdot \frac{(\hat{f}[k-B] - \hat{f}[k])}{M} \qquad (13) \\ =& \sum_{i=k-B+1}^{k} d_S[i] + 2 \cdot \sum_{i=k-B+1}^{k} \sum_{j=i+1}^{k} \rho_{i,j} \cdot \sqrt{d_S[i] \cdot d_S[j]} \end{aligned} \tag{14}$$

which is the sum of the MSEs of independent losses and the cross-correlation terms.

Next we derive the total distortion of burst losses afflicting frames $k - B + 1$ through k. According to the definition by (3), the total distortion is

$$D[k-B+1, \ldots, k] = \sum_{i=k-B+1}^{\infty} d[i] = \sum_{i=k-B+1}^{k-1} d[i] + D[k]. \tag{15}$$

With $d[k]$ obtained from (14), we are able to derive $D[k]$ according to (10).

However, as the burst length B varies, the shape of the PSD of the input error signal is different due to the concealment mechanism, which leads to the variation of the ratio α (or r) in (10). Since a lost frame is replaced with the last good frame received by concealment at the decoder, as the burst length increases, the PSD of the error frame has larger low-frequency component. The process of error power reduction by loop filtering can be modeled with a linear system, and r is the proportion of the power of the introduced error passing through the system. In [26], the loop filter is approximated in the base band by a Gaussian low-pass filter. Hence r increases as the PSD of the error is more concentrated in the lower band. In other words, α increases as the burst length B increases.

From the simulations presented in Section IV, it is found that the variation of α as a function of B is near-linear. For this reason, we approximate α as a linear function of B:

$$\alpha(B) = \alpha_0 + c \cdot (B-2) \tag{16}$$

where α_0 is the ratio for $B=2$, c is the slope of the increase, and $B \geq 2$. The equation described by (16) can be determined by fitting at least two measured values of α for different Bs.

With the obtained α, (16), the total distortion is given by

$$D[k] = \alpha(B) \cdot d[k]$$

or

$$D[k-B+1,\ldots,k] = \sum_{i=k-B+1}^{k-1} d[i] + \alpha(B) \cdot d[k]. \tag{17}$$

C. Two Separated Losses With a Short Lag

To study the distortion of a loss with a general and arbitrary pattern, we also want to analyze the effect of two losses, where the two losses are not consecutive (i.e., not a burst loss), but they are also not far enough apart to have the effect of independent losses. We define the *lag* between two losses at frames i and j $(j > i)$ as $l = j - i$. With $l > 1$, the first error propagates before the second error occurs. We study the MSE of two separated losses at $k-l$ and k.

For a single loss at $k-l$, (4) takes a more general form as

$$e[k-1] = g[k-1] - \widehat{f}[k-1]$$

where the error in $e[k-1]$ is propagated from the error of a single loss $e_S[k-l]$. Still assuming the simple concealment scheme of copying the previous frame, the error of a single loss at k is

$$e_S[k] = g[k] - \widehat{f}[k] = \widehat{f}[k-1] - \widehat{f}[k].$$

With two losses occurring at $k-l$ and k, the error at k is

$$e[k] = g[k] - \widehat{f}[k] = g[k-1] - \widehat{f}[k] = e[k-1] + e_S[k] \tag{18}$$

and the corresponding MSE is

$$\begin{aligned} d[k] &= (g[k-1] - \widehat{f}[k])^T \cdot \frac{(g[k-1] - \widehat{f}[k])}{M} && (19)\\ &= (e[k-1] + e_S[k])^T \cdot \frac{(e[k-1] + e_S[k])}{M} \\ &= d[k-1] + d_S[k] + 2\rho_{k-1,k} \cdot \sqrt{d_S[k-1] \cdot d_S[k]} && (20) \end{aligned}$$

where

$$\rho_{k-1,k} = \frac{\frac{e[k-1]^T \cdot e_S[k]}{M}}{\sqrt{d_S[k-1] \cdot d_S[k]}} \tag{21}$$

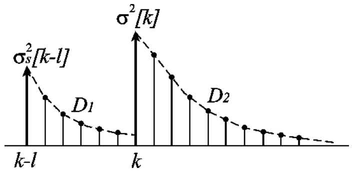

Fig. 1. Model of the distortion from two losses with a lag.

is the correlation coefficient between error frames $k-1$ (propagated from $k-l$) and k.

We now examine the total distortion for two losses occurring at $k-l$ and k, with an arbitrary lag of $1 < l \leq N$. N is the distance such that for $l > N$, the two losses are sufficiently far apart so that their effects can be treated as independent. For instance, N can be the period of Intra update. In other words, for $l > N$ the total distortion is additive, i.e., the total distortion from two losses spaced $l > N$ apart is equal to the sum of the total distortions corresponding to each independent loss event.

As illustrated in Fig. 1, we split the distortion into two parts, the distortion before and the distortion after the introduction of the second loss at k. The first distortion is given by

$$\begin{aligned} D_1 &= \sum_{i=k-l}^{k-1} d[i] \\ &= \sum_{i=0}^{l-1} r^i \left(1 - \frac{i}{N}\right) \cdot d_S[k-l] \\ &= \frac{(N-l+1)r^{l+1} - (N-l)r^l - (N+1)r + N}{N(1-r)^2} d_S[k-l] \\ &= \frac{(N-l+1)r^{l+1} - (N-l)r^l - (N+1)r + N}{r^{N+1} - (N+1)r + N} D_S[k-l] \end{aligned} \tag{22}$$

for $l \leq N$. At k, the second error is introduced, and the MSE of frame k, $d[k]$, is obtained from (20). The second component of the total distortion is

$$\begin{aligned} D_2 &= \sum_{i=k}^{\infty} d[i] = \sum_{i=0}^{N-1} r^i \left(1 - \frac{i}{N}\right) \cdot d[k] \\ &= \frac{r^{N+1} - (N+1)r + N}{N(1-r)^2} d[k] \\ &= \alpha \cdot d[k] = \frac{d[k]}{d_S[k]} D_S[k] \end{aligned} \tag{23}$$

given $D_S[k] = \alpha \cdot d_S[k]$.

Hence, the total distortion is shown in (24) and (25) at the bottom of the page. Note that, in (25), the total distortion is

$$\begin{aligned} D[k-l,k] &= D_1 + D_2 && (24)\\ &= \frac{(N-l+1)r^{l+1} - (N-l)r^l - (N+1)r + N}{N(1-r)^2} d_S[k-l] + \alpha \cdot d[k] \\ &= \frac{(N-l+1)r^{l+1} - (N-l)r^l - (N+1)r + N}{r^{N+1} - (N+1)r + N} D_S[k-l] + \frac{d[k]}{d_S[k]} D_S[k] && (25) \end{aligned}$$

re- expressed as a function of the distortion of two single and independent losses. The scaling of these two distortions, which is a function of the lag and the correlation between the error frames, is what distinguishes this model from the prior additive model.

III. Parameter Estimation and Algorithms

Before applying the model, the model parameters need to be estimated for a particular video sequence. In this section, we present two approaches for parameter estimation: local estimation and global estimation.

In calculating the MSE d of an arbitrary frame in error, and the total distortion D of the error event, we need the MSE of a single loss d_S and the ratio $\alpha = D_S/d_S$. Note that d_S is frame content dependent and varies as the frame index k. For a video sequence with a total of L frames, if $d_S[k]$ is pre-measured and stored for a single loss occurring to each frame in the sequence, e.g., for $k = 0, 1, \ldots, L-1$, a loss in a general pattern occurring at any particular location in the sequence may be accurately obtained. This approach is referred to as the local estimation, since the parameters are estimated and stored for localized error events. Local estimation is useful when the estimation of loss occurring at particular locations is desired, such as in the applications of real-time channel-adaptive R-D optimization.

When estimating the average distortion of a video sequence afflicted by stationary error events, or by errors occurring at undetermined random locations, an averaged parameter $\overline{d}_S$ for the entire sequence is required. In this case, a smaller number of simulations and decodings are needed, for single loss events at only a subsampled of frames in the sequence, for instance, at frames $k = 10, 20, 30, \ldots$ only. $\overline{d}_S$ and α are estimated by averaging the measured MSEs, for instance, $d_S[10], d_S[20], d_S[30], \ldots$. This approach is referred to as the global estimation, and fewer global parameters need to be stored for a particular sequence. Global estimation gives a low-complexity alternative for the estimation of the distortion averaged over a sequence.

Next we provide more details for parameter estimation in the cases of different loss patterns.

In the case of two separated losses, using local parameter estimation, we perform L simulations, each time with a single loss occurring at a particular location k, where $k = 0, 1, \ldots, L-1$. In each simulation, e.g., a frame loss at k, we measure the MSE of the frame in error, $d_S[k]$, and the resulting total distortion, $D_S[k]$. We also estimate the MSE $d[k,l]$ (or equivalently the correlation coefficient in (21)) according to (19), for the frame l frames after, if it should get lost. Here we count $l = 1, 2, \ldots, N$, since we consider two losses as independent if they are more than N frames apart. We estimate $d[k,l]$ according to (19), and assume the reconstructed frames $\widehat{f}$ are available at the encoder. The ratio α is obtained as the ratio of the averaged D_S and the averaged d_S (both averaged over k), and r can be solved from (10). In summary, to estimate the required model parameters, $L \times N$ decodings during pre-measurement are required in total. A total of $L + L \times N + 1$ parameters need to be stored: including $d_S[k]$, $d[k,l]$ and α (or equivalently, r), for $k = 0, 1, \ldots, L-1$ and $l = 1, 2, \ldots, N$. With the obtained parameters, the total distortion of a particular loss event can be calculated using the model by (25).

Using global estimation, we perform L' decodings, with a single loss occurring only at subsampled locations, for instance, at frames $k = 10, 20, 30, \ldots$. In each simulation, $d_S[k]$, $D_S[k]$ and $d[k,l]$ are measured in the same way as using the local estimation. However only the average $\overline{d}_S$ and average $\overline{d}[l]$ are stored, which are the values of $d_S[k]$ and $d[k,l]$ averaged over k, respectively. The ratio α is obtained in a similar way, except the D_S and the d_S used to calculate α is averaged over L' samples instead of L. In summary, for global estimation, $L' \times N$ decodings are required. A total of $1 + N + 1$ parameters need to be stored: $\overline{d}_S$, $\overline{d}[l]$ and α (or equivalently, r), for $l = 1, 2, \ldots, N$.

In the case of burst loss with $B > 2$, to pre-measure the ratio $\alpha(B)$ using local estimation, loss events at all L locations, e.g., $k = 0, 1, \ldots, L-1$, are considered. For each k, where the loss takes place, 2 simulations with two different error burst length B_1 and B_2 are performed to measure the resulting D and d. $\alpha(B)$ is obtained as the ratio of the averaged D and the averaged d, for a particular B. With the obtained $\alpha(B_1)$ and $\alpha(B_2)$, $\alpha(B)$ can be calculated according to (16). The MSE of the frames in error, $d[k]$, can be calculated from the reconstructed frames using (13) online, without pre-decoding the sequence. In summary, to estimate the parameters, $L \times 2$ decodings during pre-measurement are required. A total of 2 parameters are stored for a particular sequence: α_0 and c in (16) for the estimation of α. With these parameters, the total distortion is obtained from (17). Burst loss with $B = 2$ is a special case of the above, in which α is a constant and the only parameter to store, and only L decodings are needed.

Using global estimation, the ratio α is obtained in a similar way, except that loss events are simulated at subsampled locations. As a result, only $L' \times 2$ decodings are needed. The MSE of the frames in error, $d[k]$, is estimated in the same way. Two parameters, α_0 and c, are stored.

IV. Simulation Results

To validate the accuracy of the proposed model and to compare it with prior models, we simulate different loss patterns on standard video test sequences, and measure the resulting total distortion at the decoder. Specifically, we compare the measured distortion with that predicted by the proposed model and predicted by the additive model described in Section I.

To perform our tests, we use the H.264/AVC emerging video compression standard [35]. Specifically, all test are performed with the JVT JM 2.0 codec implementation of this emerging standard. We evaluate the performance on four standard video-telephony-type sequences in QCIF format, *Foreman*, *Mother-Daughter*, *Salesman* and *Claire*, where each has 280 frames and is coded at 30 fps. Note that these standard test sequences represent a range of video complexity, from little motion in *Mother-Daughter* or *Claire* to significant motion in *Foreman*. Each sequence is coded at a constant quality (constant PSNR determined by a constant quantizer stepsize), and the bit-rate and PSNR for each sequence is shown in Table I. The first frame of a sequence is Intra-coded, followed by Inter-coded frames. Every 4 frames, a slice is Intra updated for increased error-resilience and reducing possible error power, which corresponds to an Intra-frame update in a period of $N = 4 \times 9 = 36$ frames. When a loss occurs, the lost frame is

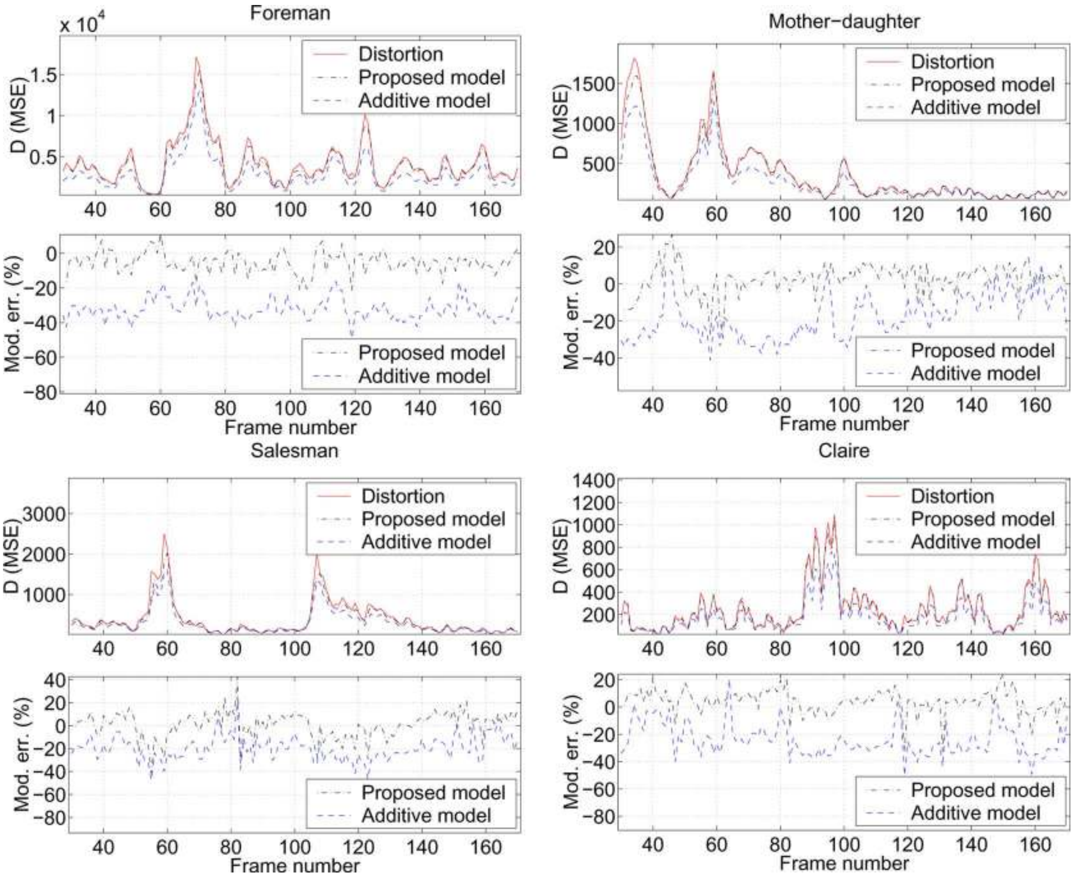

Fig. 2. Measured versus estimated total distortion, D (in MSE), and modeling error, for burst loss of length two. *Foreman, Mother-Daughter, Salesman* and *Claire* sequences. Local parameter estimation.

TABLE I
BIT RATE AND PSNR FOR SEQUENCES TESTED

Sequence	Foreman	Mother-Daughter	Salesman	Claire
Bit rate (Kbps)	153.0	70.13	68.03	38.95
PSNR (dB)	35.72	36.20	34.93	39.57

TABLE II
AVERAGE MODELING ERROR (DB) FOR BURST LOSSES OF LENGTH TWO, GIVEN BY THE ADDITIVE MODEL, PROPOSED MODEL WITH LOCAL PARAMETER ESTIMATION (LE) AND GLOBAL ESTIMATION (GE). ALSO LISTED ARE THE RATIO α, AND THE AVERAGE MSE FOR A SINGLE LOSS

Sequence	Foreman	Mother-Daughter	Salesman	Claire
Additive model	-1.51	-1.40	-1.31	-1.47
Proposed (LE)	-0.29	-0.18	-0.41	0.07
Proposed (GE)	-0.32	-0.11	-0.71	-0.18
$\alpha = D_S/d_S$	14.7	11.5	14.8	16.0
$\bar{d}_S$ (MSE)	98.0	17.4	9.9	5.4

replaced by the last correctly decoded frame, as implemented in the JVT JM 2.0 codec. The distortion we study in this work, including the lost-frame MSE and the total distortion, only includes channel-induced error, as defined in (1)–(3). We use both the local estimation (LE) and the global estimation (GE) to obtain the model parameters, as described in Section III. For LE of the parameter, $L = 140$ frames is used; while for GE, $L' = 30$.

1) Distortion for Burst Losses of Length Two: Fig. 2 plots the total distortion for burst losses of length two as we vary the frame where the burst loss begins. It is observed that the proposed model using (12), predicts the measured distortion quite well, while the additive model generally underestimates the distortion due to the prevailing positive correlation [as expressed in (8)] between the two adjacent error frames. Table II lists the average modeling error for the two methods.

2) Distortion as a Function of Burst Length: Fig. 3 shows the total distortion for burst losses of varying lengths. For each burst length, we simulate the loss event occurring at different frames in the video sequence, decode the sequence and compute the resulting total distortion. The averaged distortion is computed over all loss realizations. The averaged total distortion is then normalized to the distortion resulting from a single loss (also averaged over all loss realizations), and presented on the logarithmic scale.

It is observed from Fig. 3 that as the burst length increases, the total distortion is much greater than the sum of the distortion of individual losses, unlike what is predicted by the additive model. These plots clearly illustrate that burst length matters, and that the total distortion is not equivalent to an equal number of isolated losses. Furthermore, the proposed model accurately accounts for the effect of burst length, as shown by its accuracy in predicting the total distortion for burst losses.

Note that the observation found above can be very different from the scenario of high rate video over ATM networks [36], where it is found that "short burst cell loss causes greater video degradation compared to long burst cell loss." This is because,

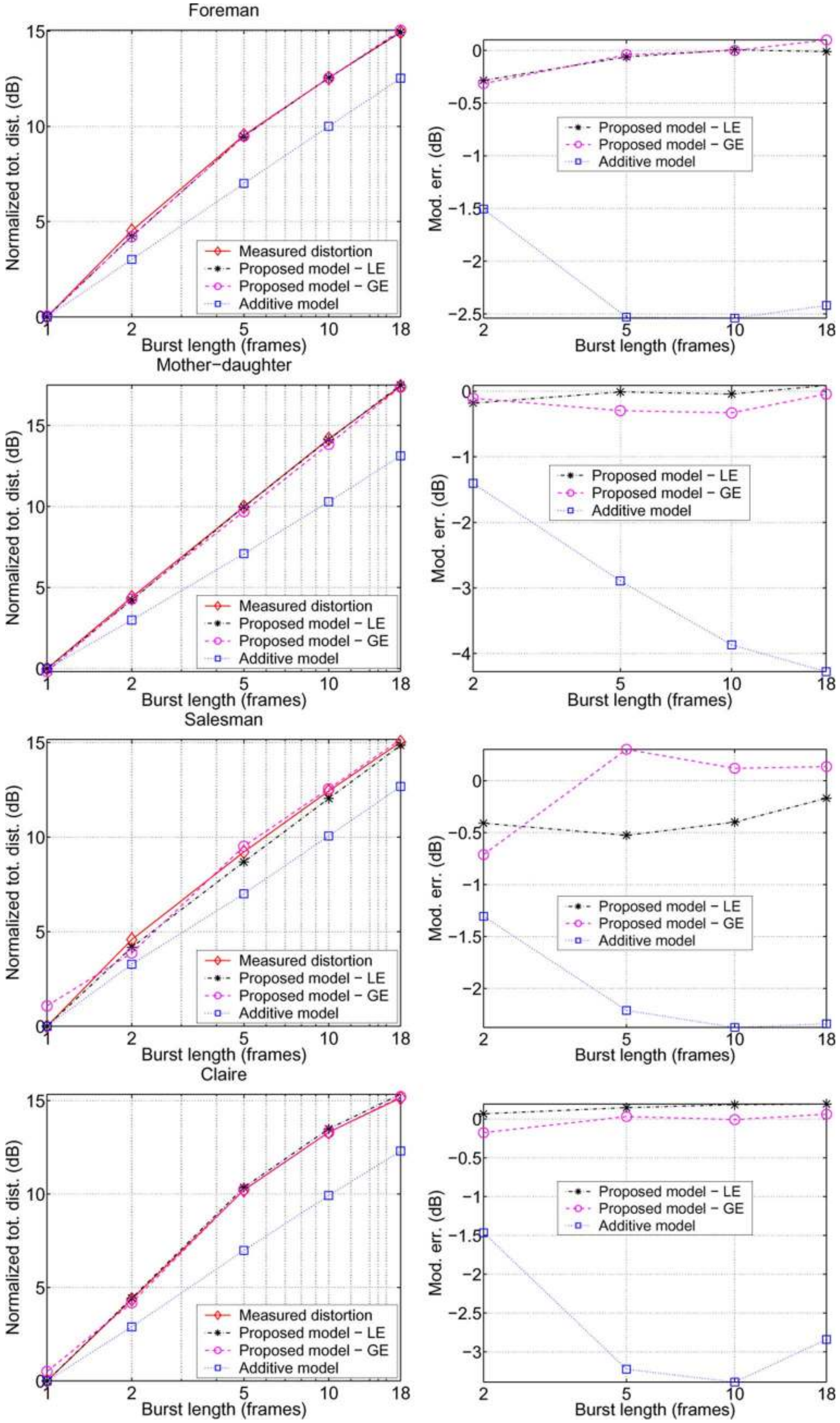

Fig. 3. Measured versus estimated total distortion, D, and modeling error, as a function of burst length, averaged for loss realizations at different frames in a sequence, and normalized by the average distortion of a single loss. *Foreman, Mother-Daughter, Salesman* and *Claire* sequences.

in [36], ATM cells are small in relative to video frames, and they do not contain independently decodable packets. A compressed video frame typically has to be distributed across many cells. In that scenario, the loss of a single cell resulted in the

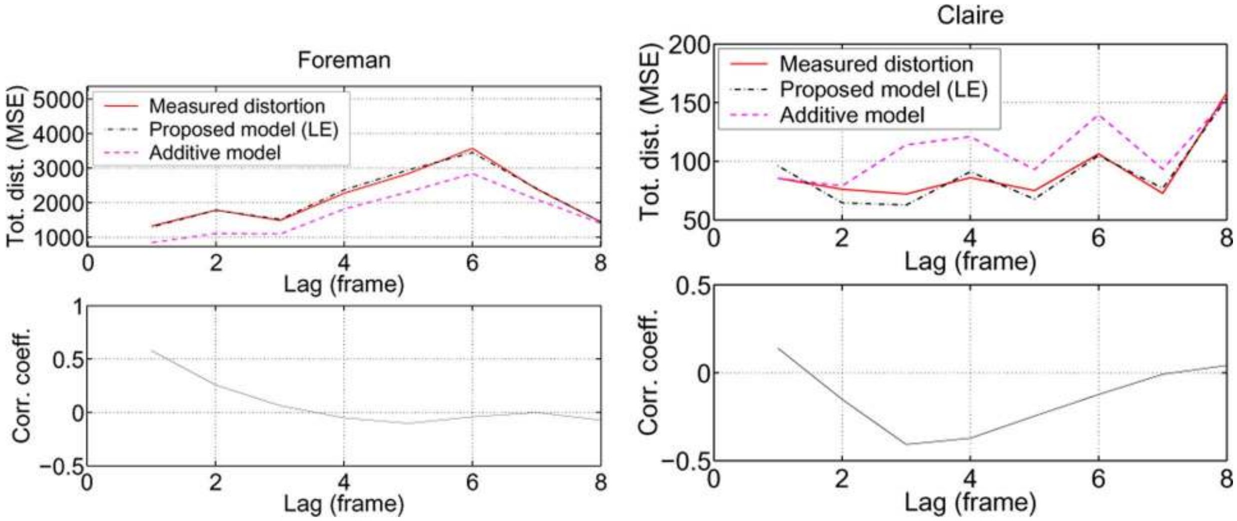

Fig. 4. Total distortion and error correlation of two losses with a lag. First loss at Frame 80, and second loss at frame 80 + lag. *Foreman* and *Claire* sequences.

loss of an entire frame, hence a single loss or a burst loss within a single frame would have the same effect. It is therefore more beneficial to have a longer burst losses afflicting one frame instead of distributed short bursts at the same average loss rate afflicting multiple frames. However, in this paper we consider a fundamentally different situation in our H.264/RTP/UDP/IP scenario.

3) Distortion as a Function of the Lag Between Two Losses: Figs. 4 and 5 show the distortion for two separated losses as a function of the gap between the losses. Fig. 4 plots the distortion and error frame correlation, calculated according to (25) and (21), respectively, for one particular realization in which the first loss occurs at Frame 80. When the lag is small, the additive model underestimates the distortion for *Foreman* due to the positive correlation. However, it overestimates the distortion for *Claire* due to the negative correlation, which is in accordance with (20).

Fig. 5 shows the distortion for different lags, averaged over all loss realizations, as well as the modeling error. The averaged total distortion is also normalized, and given on the logarithmic scale. Note that for *Foreman*, the proposed model (LE) underestimates the error by up to 0.24 dB, while the additive model underestimates the error by up to 1.64 dB. Furthermore, for *Claire*, the proposed model (LE) estimates the distortion to an accuracy of within ±0.09 dB for all lags, while the additive model underestimates the distortion by 1.57 dB for some lags and overestimates it by 0.86 dB for other lags. To summarize the results for this figure, the proposed model provides much higher accuracy, especially for small lags. The additive model does not take the lag into consideration, and is accurate only for large lags when the two losses are isolated and can be treated independently.

Note that the distortion models work well in the sense that as soon as a loss afflicts overlapping portions of consecutive frames the error signals will be correlated. While the specific video source material may affect the modeling results, the first-order effect is the number of packets per frame as compared to the burst length. When the ratio of the average burst length to the average number of packets per frame is greater than 1 the model is applicable, and becomes more important as the above ratio increases. The correlation between error frames typically does not come into effect when the ratio is smaller than 1. Since the model considers first and second-order loss events, it works well for video-telephony-type of sequences with low to medium motion. Corresponding modeling of losses for high-motion content is the target for further study.

V. Application to Delay-Distortion Optimized Packet Interleaving

Section II describes that a burst loss generally produces greater total distortion than an equivalent number of isolated losses. This suggests that when communicating over a channel that exhibits burst losses, it would be beneficial to use interleaving to convert the burst losses into an equal number of isolated losses that in general are easier to recover from and produce lower total distortion. In this section, we explore a simple packet scheduling scheme, packet interleaving, to achieve this goal. We apply the proposed loss model to the design of an optimal packet interleaver that maximizes the performance (minimizes total distortion) given knowledge of the burst loss characteristics of the channel. Compared to other types of error-resilience techniques, packet interleaving provides the advantages of: 1) being simple and 2) not requiring any increase in bit rate.

A. Packet Interleaver

A simple block interleaver is used at the sender to interleave the packets before transmission. Packets are first read into the interleaver in rows, with each row corresponding to a block of n packets. Packets are transmitted as soon as m rows of packets fill up, and are transmitted by columns. Here n is referred to as the *block size* and m is the *interleaving depth* of the interleaver. Fig. 6 shows an (n, m) interleaver.

A simple packet interleaver permutes the locations of losses in order to convert burst losses into isolated losses. The effectiveness of the interleaver depends on the block size and the interleaving depth of the interleaver, and the loss characteristics of the channel. A larger interleaver is more effective in that it can convert a longer burst loss into isolated losses or increase

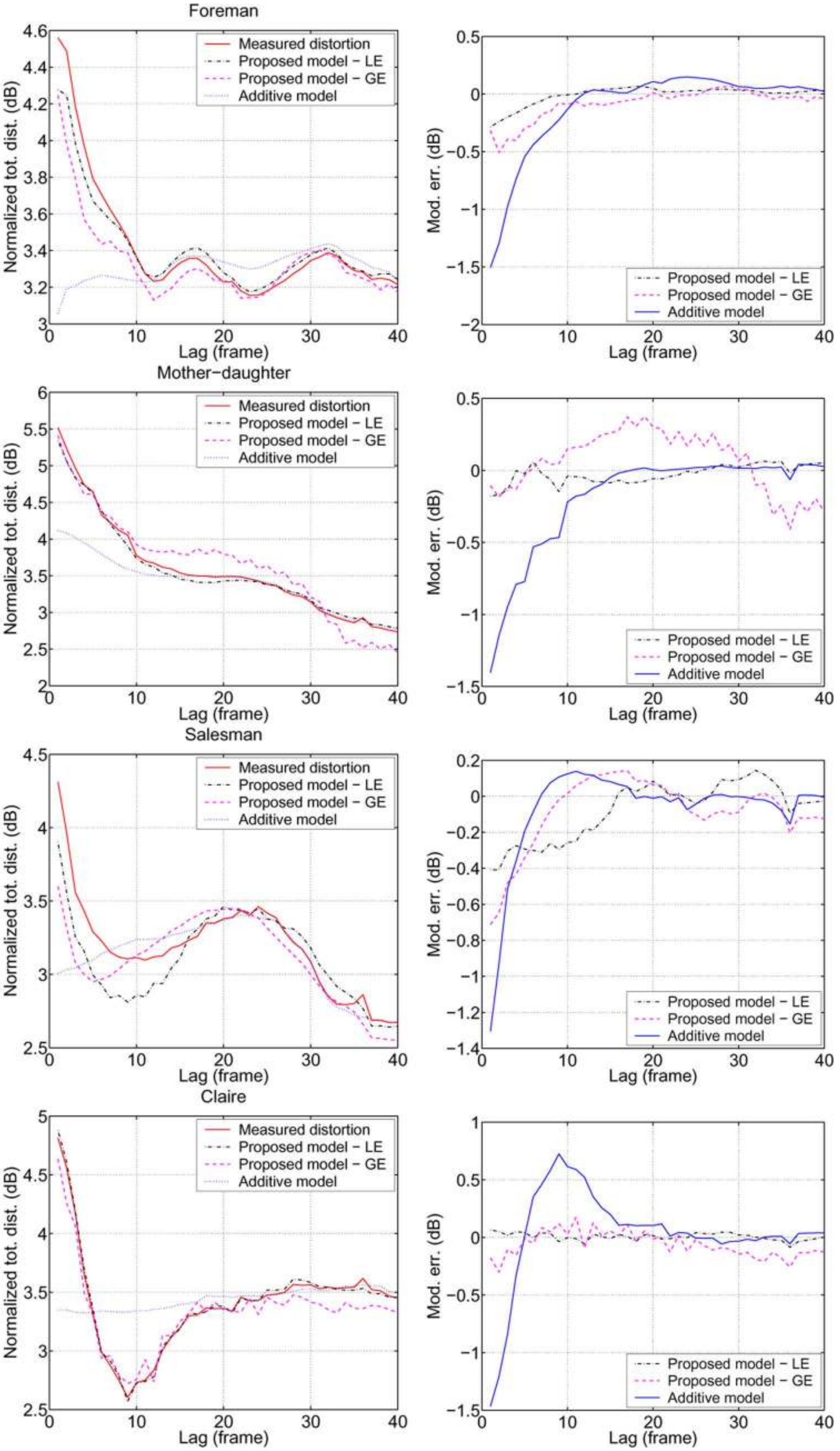

Fig. 5. Measured versus estimated total distortion, D, and modeling error, for two losses separated by a lag, averaged for loss realizations at different frames in a sequence, and normalized by the average distortion of a single loss. Model parameters obtained using LE and GE. *Foreman, Mother–Daughter, Salesman* and *Claire* sequences.

the separation of the converted isolated losses. However, this is at the cost of higher latency. At the client, an interleaved packet received cannot be used until all the packets it depends on are received. For an (n, m) interleaver, the n-th packet in

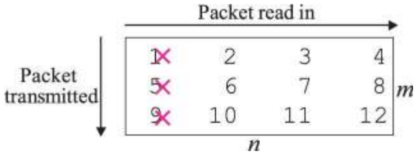

Fig. 6. Block interleaver with block size $n = 4$ and interleaving depth $m = 3$.

the original order suffers from the highest delay, which has to be transmitted in the $((n-1)\cdot m+1)$th place. Hence, the decoding delay corresponding to an (n, m) interleaver is $(n-1)\times(m-1)$, and a trade-off exists between the effectiveness in permuting the packets and the latency. It should be noted that generally a large delay is not required since, as will be shown in Section V-C, as $n \times m$ increases beyond a certain point, further increase in n or m does not necessarily improve the performance, i.e., a larger interleaver is not always better. In Section V-B, we determine the optimal interleaver (n, m) under certain delay constraints.

B. Optimal Block Interleaving

We use the set $K_{\text{orig}} = \{k_1, k_2, \ldots\}$ to denote the indices of the original lost packets when transmitted over the channel with no interleaving. With interleaving, the losses are redistributed across packets, and the loss indices are a function of the interleaver parameters. We use $\mathcal{K} = I(n, m, K_{\text{orig}})$ to denote the indices of the lost packets when an (n, m) interleaver is used, where $I(\cdot)$ is the functional representation of the interleaver (n, m), and K_{orig} denotes the indices of the lost packets before interleaving.

The total distortion D of the decoded video sequence, which depends on the loss pattern, is a function of the lost packets $\mathcal{K}$, and hence a function of the interleaver used, (n, m). If the channel loss statistics are known (for instance, the distribution of B is known) we are able to determine the optimal interleaver $(n_{\text{opt}}, m_{\text{opt}})$ that achieves the lowest distortion given a delay constraint. The problem is formally stated as follows: given the channel loss characteristics, and the delay constraint C_{delay}, determine the optimal interleaver $(n_{\text{opt}}, m_{\text{opt}})$, such that the total distortion of the decoded video sequence $D[I(n, m, K_{\text{orig}})]$ is minimized, i.e.,

$$\begin{aligned}&(n_{\text{opt}}, m_{\text{opt}})\\ &\quad = \arg\min\nolimits_{n,m:(n-1)\times(m-1)\le C_{\text{delay}}} D[I(n, m, K_{\text{orig}})]. \qquad (26)\end{aligned}$$

This is a delay-distortion optimization problem. To solve for the optimal n and m, we need to estimate the distortion that results for different loss patterns $\mathcal{K}$. This is achieved using our proposed loss model presented in Section II. When the characteristics of a channel are known, e.g., the probability distribution for burst loss length B, the distortion in (26) is the expected distortion. The optimal interleaver is then selected to minimize the expected distortion.

Given an estimate of the channel loss characteristics, we can estimate the probability of different loss patterns and hence the associated loss events K_{orig}. For a given delay constraint C_{delay}, we determine all factorizations of n and m, such that $(n-1)\times(m-1) \le C_{\text{delay}}$, which correspond to eligible interleavers with acceptable delay constraints. For each set of eligible interleaver parameters (n, m), we calculate the indices $I(n, m, K_{\text{orig}})$ of the redistributed losses. For a particular loss event $\mathcal{K} = I(n, m, K_{\text{orig}})$, we are able to estimate the corresponding total distortion, $D[\mathcal{K}]$, using the loss model discussed in Section II. The estimated distortion for a particular loss event $\mathcal{K}$, and for a particular video sequence, can also be stored at the sender or streaming server for future use.

C. Simulation Results

We illustrate the potential performance gain that may be achieved by using the simple interleaving scheme for a channel that exhibits a significant amount of burst loss. In addition, we investigate the trade-off between performance gain from larger interleavers and the corresponding delay. A simple bursty channel model is used to illustrate the effects. We simulate that time is divided into 100-ms intervals, with each interval corresponding to 3 packets (frames) for a frame rate of 30 fps. Each interval may be in either a good state or a bad state. In a good state, 3 consecutive packets are received; while in a bad state, 3 are lost. Each time interval is assumed to be independent and identically distributed (Bernoulli), with the probability that a time interval is in the bad state is 0.10. The average packet loss rate is therefore 10%. Our primary reason for choosing this simple channel model is that it simplifies interpretation of the results. The experimental conditions are the same as in Section IV. The distortions are obtained by averaging the results for 6 random channel loss realizations shifted across the whole sequence, or, a total of 280×6 loss realizations.

For different delay constraints, all of the eligible interleavers are identified and their performances are then estimated. The PSNRs for *Foreman* and *Claire* with different interleavers as a function of delay constraint are shown in Fig. 7. Note that the PSNRs shown in Fig. 7 are the averaged results for all frames, including both good and error-afflicted frames, in a sequence; while the quantities previously shown in Fig. 3 are the normalized total distortion of the error-afflicted frames only, where the total distortion is defined by (10). For a particular delay constraint C_{delay}, an optimal interleaver $(n_{\text{opt}}, m_{\text{opt}})$ is determined using the algorithm in Section V-B. Although many eligible interleavers are tested, only those providing optimal performance are marked with circles in the plot. For example, for $C_{\text{delay}} = 12$ frames, $(n_{\text{opt}}, m_{\text{opt}}) = (7, 3)$ is found. For $C_{\text{delay}} > 12$, increasing the interleaver size further more does not improve the effectiveness. In particular, for short burst lengths, a small interleaver with low latency is sufficient to provide most of the gain. This is because, for a given burst loss model, as the size of the interleaver increases beyond a certain point, the burst losses are isolated and spaced apart far enough from each other such that they can be considered independent. Further separation of the losses will not bring additional gain.. The PSNR curve in the plot is stair-cased, which is the outer bound of all data points tested.

It is observed from Fig. 7 that using interleaver (5, 3) with a delay of 8 frames (267 ms) provides a gain of 0.67 dB over the case of no interleaving for *Foreman*. Using interleaver (7, 3) with a delay of 12 frames (400 ms), increases the gain to 0.72 dB. For *Claire* sequence, gains of 0.81 and 0.93 dB are achieved for delays of 333 and 533 ms, respectively.

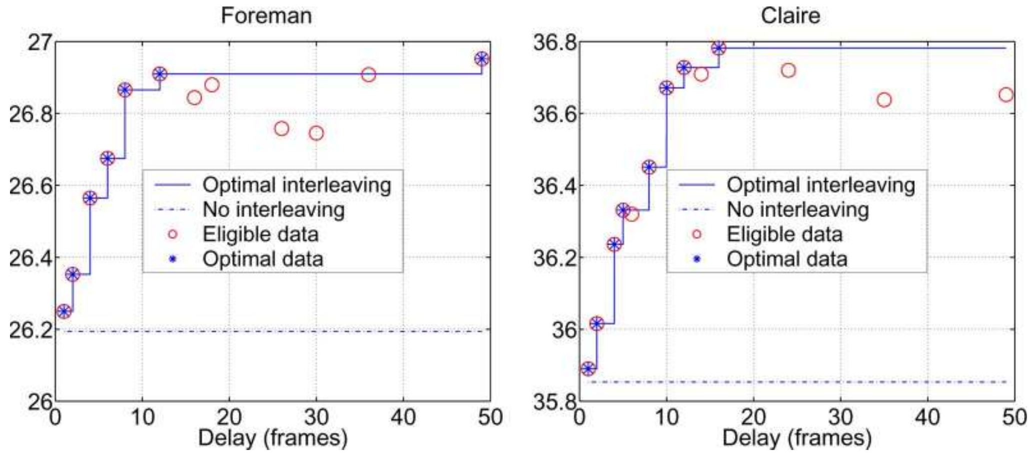

Fig. 7. PSNR of the optimal interleaver versus delay constraint. Optimal experimental data points are marked with circles.

TABLE III
GAIN IN PSNR (dB) PROVIDED BY THE OPTIMAL INTERLEAVER FOR DIFFERENT DELAY CONSTAINTS

Delay (frame/ms)	Foreman	Mother	Salesman	Claire
8/267	0.67	0.16	0.32	0.60
12/400	0.72	0.24	0.36	0.87
16/533	0.72	0.24	0.36	0.93

The gains in PSNR for all four video test sequences examined in the experiments are listed in Table III, for different delay constraints and corresponding optimal interleavers. Note that these gains are obtained without requiring any increase in bit rate. The optimal interleavers for delays of 8 and 12 frames are (5, 3) and (7, 3), respectively, for all sequences, which indicates the optimal interleaver's weak dependence on the sequence.

VI. CONCLUSION

This paper examined the question of whether the loss pattern, and in particular the burst length, is important for accurately estimating the expected distortion for video communication over error-prone channels using previous-frame concealment. We verified that the loss pattern of packet loss does in fact have a significant effect on the resulting distortion and, therefore, should be accounted for. This is consistent with prior work [21], [34]. We proposed a model for estimating the expected distortion that explicitly accounts for the loss pattern. This model explains why a loss pattern, such as a burst loss, generally produces a larger total distortion than an equal number of isolated losses. This model was shown to provide significant improvements in accurately predicting the distortion that results for different loss patterns (loss events). The proposed model can be used to estimate the distortion for general and complex loss patterns, including burst losses and separated (non-consecutive) losses. By accounting for the inter-frame error propagation, explicitly accounting for the correlation between error frames, and modeling spatial filtering at the decoder as a linear system, the proposed model provides a significantly more accurate estimate of the resulting distortion, as compared to prior models. The proposed model is validated with H.264/AVC coded video, and experiments on four video–telephony-type test sequences that represent a diverse range of video content. Specifically, for most sequences, the proposed model accurately predicts the total distortion to within ± 0.3 dB for two packet losses, as compared to the prior additive model that could underestimate the distortion by about 1.6 dB or overestimate by about 0.9 dB. Furthermore, the accuracy of our prediction is within ±0.7 dB as the length of a burst loss increases, while that of the prior model degrades and may underestimate the total distortion by over 3 dB. The proposed model works well for video-telephony-type of sequences with low to medium motion. We also present a simple illustrative example, of how knowledge of the effect of burst loss and accurate modeling can be used to adapt the schedule of video streaming to provide improved performance for a burst loss channel, without requiring an increase in bit-rate. We expect that the use of this more accurate loss model can improve the design and performance of various error-resilient video communication schemes.

ACKNOWLEDGMENT

The authors would like to thank W. (Dan) Tan and S. Wee of HP Labs for their contributions to this work.

REFERENCES

[1] S. Wenger, G. D. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 7, pp. 867–877, Nov. 1998.

[2] R. Talluri, "Error-resilient video coding in the iso MPEG-4 standard," *IEEE Commun. Mag.*, pp. 112–119, Jun. 1998.

[3] N. Färber, B. Girod, and J. Villasenor, "Extension of ITU-T recommendation H.324 for error-resilient video transmission," *IEEE Commun. Mag.*, pp. 120–128, Jun. 1998.

[4] E. Steinbach, N. Färber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 6, pp. 872–881, Dec. 1997.

[5] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. , no. 10, pp. 1707–1723, Oct. 1999.

[6] W. Tan and A. Zakhor, "Real-time Internet video using error resilient scalable compression and TCP-friendly transport protocol," *IEEE Trans. Multimedia*, vol. , no. 2, pp. 172–186, Jun. 1999.

[7] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[8] J. Y. Liao and J. D. Villasenor, "Adaptive intra update for video coding over noisy channels," in *Proc. IEEE Int. Conf. on Image Process.*, Lausanne, Switzerland, Sep. 1996, vol. 3, pp. 763–6.

[9] R. O. Hinds, T. N. Pappas, and J. S. Lim, "Joint block-based video source/channel coding for packet-switched networks," in *Proc. SPIE VCIP 98*, San Jose, CA, Oct. 1998, vol. 3309, pp. 124–33.

[10] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas in Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[11] T. Wiegand, N. Färber, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Sel. Areas in Commun.*, vol. 18, no. 6, pp. 1050–1062, Jun. 2000.

[12] M. Budagavi and J. D. Gibson, "Multiframe video coding for improved performance over wireless channels," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 252–265, Feb. 2001.

[13] Y. J. Liang and B. Girod, "Network-adaptive low-latency video communication over best-effort networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 72–81, Jan. 2006.

[14] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pt. 1, pp. 1737–1744, Nov. 1996.

[15] W. Tan and A. Zakhor, "Video multicast using layered FECs: And scalable compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 373–87, Mar. 2001.

[16] P. A. Chou, A. E. Mohr, A. Wang, and S. Mehrotra, "Fec and pseudo-ARQ for receiver-driven layered multicast of audio and video," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2000, pp. 440–449.

[17] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, no. 1, Feb. 2001, submitted for publication.

[18] J. Chakareski, P. A. Chou, and B. Aazhang, "Computing rate-distortion optimized policies for streaming media to wireless clients," in *Proc. Data Compression Conf.*, Snowbird, UT, Apr. 2002, pp. 53–62.

[19] M. Kalman, E. Steinbach, and B. Girod, "R-D optimized media streaming enhanced with adaptive media playout," in *Proc. IEEE Int. Conf. on Multimedia, ICME-2002*, Lausanne, Switzerland, Aug. 2002, vol. 1, pp. 869–872.

[20] S. J. Wee, W. Tan, J. G. Apostolopoulos, and M. Etoh, "Optimized video streaming for networks with varying delay," in *Proc. of the IEEE Int. Conf. on Multimedia and Expo (ICME)*, Lausanne, Switzerland, Aug. 2002, vol. 2, pp. 89–92.

[21] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," *Proc. Visual Commun. and Image Process.*, pp. 392–409, Jan. 2001.

[22] N. Gogate, D.-M. Chung, S. S. Panwar, and Y. Wang, "Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 9, pp. 777–792, Sep. 2002.

[23] D. Comas, R. Singh, and A. Ortega, "Rate-distortion optimization in a robust video transmission based on unbalanced multiple description coding," in *Proc. IEEE 4th Workshop on Multimedia Signal Process.*, Cannes, France, Oct. 2001, pp. 581–586.

[24] Y. Wang, S. S. Panwar, S. Lin, and S. Mao, "Wireless video transport using path diversity: Multiple description vs. layered coding," in *Proc. of the IEEE Int. Conf. on Image Processing*, Rochester, NY, Sep. 2002.

[25] Y. J. Liang, E. Setton, and B. Girod, "Network-adaptive video communication using packet path diversity and rate-distortion optimized reference picture selection," *J. VLSI Signal Process. Syst. Signal, Image, and Video Technol.*, vol. 41, no. 3, Nov. 2005.

[26] N. Färber, K. Stuhlmüller, and B. Girod, "Analysis of error propagation in hybrid video coding with application to error resilience," in *Proc. IEEE Int. Conf. on Image Processing*, Kobe, Japan, Oct. 1999, vol. 2, pp. 550–4.

[27] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas in Commun.*, vol. 18, no. 6, pp. 1012–32, Jun. 2000.

[28] I.-M. Kim and H.-M. Kim, "A new resource allocation scheme based on a PSNR criterion for wireless video transmission to stationary receivers over Gaussian channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. , pp. 393–401, Jul. 2002.

[29] A. R. Reibman, Y. Sermadevi, and V. Vaishampayan, "Quality monitoring of video over the Internet," in *Proc. 36th Asilomar Conf. Signals, Syst., and Comput.*, Nov. 2002, vol. 2, pp. 1320–1324.

[30] R. Zhang, S. L. Regunathan, and K. Rose, "End-to-end distortion estimation for rd-based robust delivery of pre-compressed video," in *Proc. 35th Asilomar Conf. on Signals, Syst., and Comput.*, Pacific Grove, CA, Nov. 2001, vol. 1, pp. 210–14.

[31] R. Zhang, S. L. Regunathan, and K. Rose, "Optimized video streaming over lossy networks with real-time estimation of end-to-end distortion," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*, Lausanne, Switzerland, Aug. 2002, vol. 1, pp. 861–4.

[32] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.

[33] A. R. Reibman and V. Vaishampayan, "Quality monitoring for compressed video subjected to packet loss," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*, Jul. 2003, vol. 1, pp. 17–20.

[34] J. G. Apostolopoulos, W. Tan, S. J. Wee, and G. W. Wornell, "Modeling path diversity for multiple description video communication," in *Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Process. (ICASSP '02)*, Orlando, FL, May 2002.

[35] *Advanced Video Coding (AVC) for Generic Audiovisual Services*, ITU-T Recommendation H.264, ITU, 2003.

[36] L. Zhang, D. Chow, and C. H. Ng, "Cell loss effect on QoS for MPEG video transmission in ATM networks," in *Proc. IEEE Int. Conf. on Commun. (ICC '99)*, Vancouver, BC, Canada, Jun. 1999, vol. 1, pp. 147–151.

Yi Liang received the B.Eng. degree from Tsinghua University, Beijing, China, and the Ph.D. degree in electrical engineering from Stanford University, Palo Alto, CA.

He is the founder and currently Chief Technology Officer of Mobim Technologies, Shanghai, China. At Mobim Technologies, he and his team pioneered the mobile instant messaging application enabled with live interactive video communication over very-low-speed GPRS networks. Built-in innovative mobile video technologies, handset devices featuring Mobim's multimedia solutions have shipped to over twenty countries worldwide. From 2003 to 2007, he held positions at Qualcomm CDMA Technologies, San Diego, CA, and was responsible for the design and development of video and display system architecture for multimedia handset chipsets. From 2000 to 2001, he conducted research with Netergy Networks, Inc., Santa Clara, CA, on voice over IP systems that provide superior quality over best-effort networks. From 2001 to 2003, he had led the Stanford—Hewlett-Packard Labs low-latency video streaming project, in which he and his colleagues developed error-resilience techniques for rich media communication over IP networks at low latency. In 2002 at Hewlett-Packard Labs, Palo Alto, CA, he contributed to the development of the pioneering mobile streaming media content delivery network (MSM—CDN) that delivers rich media over third-generation wireless. His research interests include networked and embedded multimedia systems, real-time voice and video communication, and low-latency media streaming over wireless and wire-line networks.

Dr. Liang was the recipient of the IEEE Multimedia Communications Best Paper Award in 2007 (with Nikolaus Färber and Bernd Girod).

John G. Apostolopoulos (S'92-M'97-SM'06-F'08) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge.

He joined Hewlett-Packard Laboratories in 1997, where he is currently a Distinguished Technologist and Lab Director for the Multimedia Communications and Networking Lab. He also teaches and conducts joint research at Stanford University, where he is a Consulting Associate Professor of electrical engineering. In graduate school, he worked on the U.S. Digital TV standard and received an Emmy Award Certificate for his contributions. His work on media transcoding in the middle of a network while preserving end-to-end security (secure transcoding) has recently been adopted by the JPEG-2000 Security (JPSEC) standard. His research interests include improving the reliability, fidelity, scalability, and security of media communication over wired and wireless packet networks.

Dr. Apostolopoulos received a Best Student Paper award for part of his Ph.D. thesis, the Young Investigator Award (best paper award) at VCIP 2001 for his work on multiple description video coding and path diversity, was named "one of the world's top 100 young (under 35) innovators in science and technology" (TR100) by Technology Review in 2003, and was co-author for the best paper award at ICME 2006 on authentication for streaming media. He currently serves as chair of the IEEE IMDSP and member of MMSP technical committees, and recently was general co-chair of VCIP'06 and technical co-chair for ICIP'07.

Bernd Girod (F'98) received the M. S. degree in electrical engineering from Georgia Institute of Technology, Atlanta, in 1980 and his Doctoral degree "with highest honors" from University of Hannover, Germany, in 1987.

He is Professor of Electrical Engineering in the Information Systems Laboratory of Stanford University, Palo Alto, CA. He also holds a courtesy appointment with the Stanford Department of Computer Science. He serves as Director both of the Stanford Center for Image Systems Engineering (SCIEN) and the Max Planck Center for Visual Computing and Communication. His research interests include video coding and networked media systems. Until 1987 he was a member of the research staff at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover. In 1988, he joined Massachusetts Institute of Technology, Cambridge, first as a Visiting Scientist with the Research Laboratory of Electronics, then as an Assistant Professor of Media Technology at the Media Laboratory. From 1990 to 1993, he was Professor of Computer Graphics and Technical Director of the Academy of Media Arts, Cologne, Germany, jointly appointed with the Computer Science Section of Cologne University. He was a Visiting Adjunct Professor with the Digital Signal Processing Group at Georgia Institute of Technology, Atlanta, in 1993. From 1993 until 1999, he was Chaired Professor of Electrical Engineering/Telecommunications at University of Erlangen-Nuremberg, Germany, and the Head of the Telecommunications Institute I, co-directing the Telecommunications Laboratory. He served as the Chairman of the Electrical Engineering Department from 1995 to 1997, and as Director of the Center of Excellence "3-D Image Analysis and Synthesis" from 1995 to 1999. He was a Visiting Professor with the Information Systems Laboratory of Stanford University, Stanford, CA, during the 1997/1998 academic year. As an entrepreneur, he has worked successfully with several start-up ventures as founder, investor, director, or advisor. Most notably, he has been a co-founder and Chief Scientist of Vivo Software, Inc., Waltham, MA (1993–1998); after Vivo's acquisition, 1998-2002, Chief Scientist of RealNetworks, Inc. He has served on the Board of Directors for 8×8, Inc., Santa Clara, CA, 1996–2004, and for GeoVantage, Inc., Swampscott, MA, 2000–2005. He is currently an advisor to start-up companies Mobilygen, Santa Clara, CA, and to NetEnrich, Inc., Santa Clara, CA. Since 2004, he also serves as Chairman of the Steering Committee of the new Deutsche Telekom Laboratories at the Technical University of Berlin. He has authored or co-authored one major text-book (printed in 3 languages), three monographs, and over 400 book chapters, journal articles and conference papers, and he holds over 20 US patents.

Prof. Girod has been a member of the IEEE Image and Multidimensional Signal Processing Technical Committee from 1989 to 1997 and has served on the Editorial Boards for several journals in his field, among them as Area Editor for Speech, Image, Video & Signal Processing of the IEEE TRANSACTIONS ON COMMUNICATIONS. He has served on numerous conference committees, e.g., as Tutorial Chair of ICASSP-97 in Munich and again for ICIP-2000 in Vancouver, as General Chair of the 1998 IEEE Image and Multidimensional Signal Processing Workshop in Alpbach, Austria, as General Chair of the Visual Communication and Image Processing Conference (VCIP) in San Jose, CA, in 2001, and General Chair of Vision, Modeling, and Visualization (VMV) at Stanford, CA, in 2004. He was elected Fellow of the IEEE in 1998 'for his contributions to the theory and practice of video communications.' He has been named 'Distinguished Lecturer' for the year 2002 by the IEEE Signal Processing Society. He received the 2002 EURASIP Best Paper Award (with J. Eggers) and the 2004 EURASIP Technical Achievement Award, and the IEEE Multimedia Communication Best Paper Award in 2007. He was elected a member of the German Academy of Sciences (Leopoldina) in 2007.