Submitted to ICASSP 2003, Hong Kong

ANALYSIS OF PACKET LOSS FOR COMPRESSED VIDEO: DOES BURST-LENGTH MATTER?

Yi J. Liang[*†], *John G. Apostolopoulos*[◇] *and Bernd Girod*[†]

[◇]Streaming Media Systems Group
Hewlett-Packard Labs, Palo Alto, CA 94304

[†]Information Systems Laboratory
Stanford University, Stanford, CA 94305

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. 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 JVT/H.26L coded video and previous frame concealment, where for most sequences the total distortion is predicted to within ± 0.25 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.

1. INTRODUCTION

The problem of error-resilient video communication has received significant attention in recent years, and a variety of techniques have been proposed, including intra/inter-mode switching [1, 2], dynamic control of prediction dependencies [3], forward error correction [4], and multiple description coding [5]. These approaches are designed and operated based on models for the effect of losses on the reconstructed video quality. For example, rate-distortion optimization techniques crucially depend on the accuracy of these models when they attempt to minimize the expected distortion for different loss events.

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. An important question along these lines is whether the expected distortion 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? Most prior work implicitly assumed that burst length does not matter, and focused on the average packet loss rate as the most important attribute to consider. Recently, [5, 6] identified that burst length is important and should be explicitly considered.

In this paper, we (1) verify that the packet 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. To estimate the 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 distortion resulting from different loss events, compared to prior models. The accuracy is validated for four video test sequences coded with the emerging JVT/H.26L standard.

This paper continues in Section 2 by reviewing prior models for estimating the distortion produced by packet loss. Section 3 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 which illustrate and validate the accuracy of the proposed model are presented in Section 4.

2. PREVIOUS LOSS MODELS

Prior work on modeling the effect of losses generally model the distortion as being proportional to the number of losses that occur [2, 7]. For example [2] carefully analyzes and models the distortion for a single (isolated) loss (accounting for error propagation, intra refresh, and spatial filtering), and model the effect of multiple losses as the superposition of multiple independent losses. With this linear or *additive model*, the expected distortion is proportional to the average packet loss rate. This model is accurate when single losses occur that are spaced sufficiently far apart with respect to the intra-refresh period, for example when the loss rate is low and the losses are not bursty. However, in many important communication situations, for example video communication over the Internet or over a wireless link, the losses may be bursty. In [5] 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 [6] where a simple model was proposed that distinguishes loss events based on the length of the burst loss and explicitly accounts for the different distortions that result for different burst lengths.

[*]This work was performed during a summer internship at HP Labs. The authors would also like to thank Wai-tian (Dan) Tan and Susie Wee of HP Labs for their contributions to this work.

This model provides some improvements over the prior additive model in the sense that it accurately accounts for the different effects of burst losses as opposed to isolated losses, and provides a simple mechanism for accounting for the different distortions for different burst lengths. However, it does not account for more general loss patterns, such as two losses spaced apart by a short lag.

3. PROPOSED LOSS MODEL CONSIDERING ERROR CORRELATION

This section proposes a model that can accurately estimate the distortion for more general loss patterns. Throughout this paper we 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. The results in this paper can also be extended to the case when each frame is coded into multiple packets where the loss of one packet does not result in the loss of 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 loss concealment. The initial error frame introduced by a loss at frame k is defined as

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

which is also a 1-D vector. Since our primary concern is the effect of channel loss, quantization error is not included in our study. Assuming the error frame $e[k]$ to be a zero-mean process, its variance equals its Mean Square Error (MSE), given by

$$(e^T[k] \cdot e[k])/M = \sigma^2[k].$$

The distortion that would result from a single loss, as a function of the specific frame that it afflicts, is measured at the encoder and stored by simulating the corresponding loss event, decoding the sequence, and computing the distortion. These distortions are referred to as "pre-measured" distortions in this paper. We show that by using these pre-measured distortions for single and independent losses, we can accurately estimate the distortion for 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 $\sigma_S^2[k]$; while $e[k]$ and $\sigma^2[k]$ are used for losses with more general patterns.

The above MSE quantifies the error power introduced in the initial lost frame, but it does not include the effect of error propagation to subsequent frames. We define the *total distortion*, denoted by D, to be the sum of the MSEs over all the frames in the entire error recovery period. Correspondingly $D_S[k]$ denotes the total distortion that results for a single frame loss at frame k.

3.1. Burst Losses of Length Two

Modeling the Distortion for Initial Lost Frames. In the following, we assume a simple loss concealment scheme where the lost frame is replaced by the previous frame at the decoder output. To study burst losses of length two, first consider the error frames that result for *single* losses at $k-1$ and k which are given by

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

and

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

respectively. Therefore, a burst loss of length two afflicting frames $k-1$ and k has a residual error frame k given by

$$\begin{aligned} e[k] &= g[k] - \widehat{f}[k] = \widehat{f}[k-2] - \widehat{f}[k] \\ &= e_S[k-1] + e_S[k]. \end{aligned}$$

The corresponding MSE of error frame k is

$$\sigma^2[k] = \sigma_S^2[k-1] + \sigma_S^2[k] + 2\rho_{k-1,k} \cdot \sigma_S[k-1] \cdot \sigma_S[k], \quad (1)$$

where

$$\rho_{k-1,k} = \frac{(e_S^T[k-1] \cdot e_S[k])/M}{\sigma_S[k-1] \cdot \sigma_S[k]}$$

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

In (1), 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 (1) is not just the sum of the MSEs of two independent losses, unlike what the additive model predicts. Specifically, the first two terms in (1) 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.

Modeling of the Total Distortion. To estimate the total distortion, we model the error propagation process in a typical video decoder with a geometric attenuation factor and a linear attenuation factor to account for the spatial filtering and intra update, respectively. With an intra update period of N, if a single error is introduced at k with an MSE of $\sigma^2[k]$, the power of the propagated error at $k+l$ is given by

$$\sigma^2[k+l] = \begin{cases} \sigma^2[k] \cdot r^l \cdot (1 - l/N), \text{ for } 0 \le l \le N; \\ 0, \text{ otherwise.} \end{cases} \quad (2)$$

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.

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

$$\begin{aligned} D_S[k] &= \sum_{i=k}^{\infty} \sigma^2[i] = \sum_{i=0}^{N-1} r^i (1 - \frac{i}{N}) \cdot \sigma_S^2[k] \\ &= \frac{r^{N+1} - (N+1)r + N}{N(1-r)^2} \sigma_S^2[k] = \alpha \cdot \sigma_S^2[k], \quad (3) \end{aligned}$$

where $\sigma^2[k] = \sigma_S^2[k]$ is the initial error power introduced at k, and $\alpha = D_S[k]/\sigma_S^2[k]$ is the ratio between the total distortion and the MSE of frame k. In (3) 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.

The total distortion D of two losses at $k-1$ and k is

$$\begin{aligned} D[k-1,k] &= \sum_{i=k-1}^{\infty} \sigma^2[i] = \sigma_S^2[k-1] + \alpha \cdot \sigma^2[k] \\ &= \sigma_S^2[k-1] + D_S[k-1] + D_S[k] \\ &\quad + 2\rho_{k-1,k} \cdot \sqrt{D_S[k-1] \cdot D_S[k]}, \end{aligned}$$

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 and the distortion for frame $k-1$ distinguish the proposed model from the previous additive model.

3.2. Burst Losses of Length Greater than Two

We now extend the above to model burst losses of length B ($B \geq 2$). For the loss of B consecutive frames from $k-B+1$ to k,

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

and its MSE

$$\sigma^2[k] = \sum_{i=k-B+1}^{k} \sigma_S^2[i] + 2 \cdot \sum_{i=k-B+1}^{k} \sum_{j=i+1}^{k} \rho_{i,j} \cdot \sigma_S[i] \cdot \sigma_S[j], \tag{4}$$

which is the sum of the MSEs of independent losses and the cross-correlation terms. The total distortion is given by

$$D[k-B+1, ..., k] = \sum_{i=k-B+1}^{\infty} \sigma^2[i] = \sum_{i=k-B+1}^{k-1} \sigma^2[i] + D[k].$$

With $\sigma^2[k]$ obtained from (4), we can derive $D[k]$ from (3).

However, as the burst length B varies, the shape of the initial error signal's PSD also varies, which leads to a variation in α (or r) in (3). 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 [2], the loop filter is approximated by a Gaussian low-pass filter. Hence, as B increases, r (and α) increases as the PSD of the error is more concentrated in the lower band. Fortunately, the simulations in Section 4 showed that the variation of α is relatively small and can be approximated as a linear function of B, that is $\alpha(B) = \alpha_0 + c \cdot (B-2)$, where α_0 is the ratio for $B=2$, c is the slope of the increase, and $B \geq 2$. α can be determined by two measured values for different Bs. With the obtained α, the total distortion is given by

$$D[k-B+1, ..., k] = \sum_{i=k-B+1}^{k-1} \sigma^2[i] + \alpha(B) \cdot \sigma^2[k]. \tag{5}$$

3.3. Two Losses Separated by 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 separated by a *lag*, denoted by l, where the lag is shorter than that required to make the losses independent. We study the distortion of two separated losses at $k-l$ and k, with an arbitrary lag of $1 < l \leq N$. For $l > N$, the two losses are treated as independent, and the total distortion is additive.

It can be shown that the total distortion can be expressed as

$$D[k-l, k] = \frac{(N-l+1)r^{l+1} - (N-l)r^l - (N+1)r + N}{r^{N+1} - (N+1)r + N} D_S[k-l] + \frac{\sigma^2[k]}{\sigma_S^2[k]} D_S[k] \tag{6}$$

where $\sigma^2[k]$ corresponds to the MSE of frame k resulting from both the loss of frame k and error propagation from the loss of frame $k-l$. Note that the total distortion in (6) is 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. With the two models derived above, the distortion of losses in general patterns can be obtained by using those models concatenated and combined.

4. SIMULATION RESULTS

To validate the accuracy of the proposed model, and to compare it versus the prior models, we simulate different loss patterns on standard video test sequences, and compare the measured distortion with that predicted by the proposed model and by the additive model described in Section 2. Video sequences are coded using JM 2.0 of the emerging JVT/H.26L video compression standard. Four standard test sequences in QCIF format are used, *Foreman, Mother-Daughter, Salesman* and *Claire.* Each has 280 frames at 30 fps, and is coded with a constant quantization level at an average PSNR of about 36 dB. The first frame of each sequence is intra-coded, followed by P-frames. Every 4 frames a slice is intra updated to improve error-resilience by reducing error propagation (as recommended in JM 2.0), corresponding to an intra-frame update period of $N = 4 \times 9 = 36$ frames.

The model parameters are estimated and stored for each video sequence using two approaches for parameter estimation: *local estimation (LE)* and *global estimation (GE)*. With local estimation, to calculate the σ^2 and D of an arbitrary error event, the MSE of a *single* loss σ_S^2 and the total distortion D_S are pre-measured for every frame, e.g. for $k = 0, 1, ..., L-1$, where L is total number of frames to be studied in the sequence. Since the parameters are estimated and stored for localized error events, a loss in a general pattern occurring at any location in the sequence may be accurately obtained. To estimate the required model parameters, $\sigma_S^2[k]$ and α, L decodings are required for two losses and $L \times 2$ decodings required for $B > 2$, so that $\alpha(B)$ can be calculated. With the obtained parameters, the total distortion can be calculated using the model by (5) or (6). The global estimation method gives a low-complexity alternative for estimating the distortion averaged over a sequence without considering the local frame content. An averaged parameter $\overline{\sigma}_S^2$ for the entire sequence is used, and a smaller number of simulations and decodings are needed, for single loss events at only a subsampled set of L' frames in the sequence, e.g., at frames $k = 10, 20, 30, ...$ only. In our simulations, $L = 140$ frames is used for LE, and $L' = 30$ for GE.

Fig. 1 shows the total distortion for burst losses of varying lengths. For each burst length, we simulate the loss event starting at different frames in the video sequence and decode and compute the resulting total distortion for each starting frame. The averaged distortion for each burst length is then computed by averaging over all these loss realizations. This averaged total distortion is then normalized by the total distortion resulting from a single loss (also averaged over all loss realizations), and presented on a log scale.

It is observed from Fig. 1 that as the burst length increases, the measured total distortion is much greater than the sum of the distortions for an equal number of individual losses, unlike what is predicted by the additive model. These plots clearly illustrate that burst length matters, in the sense that it has a significant effect on the reconstructed video quality, and that its effect (total distortion)

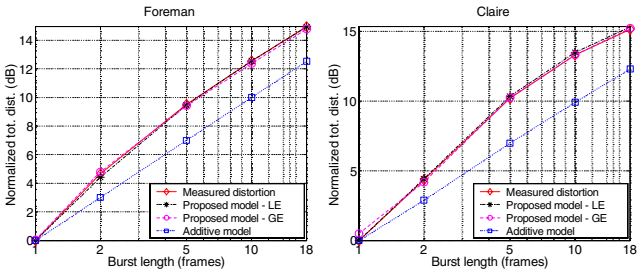

Fig. 1. Measured versus estimated total distortion as a function of burst loss length, normalized by total distortion for a single loss.

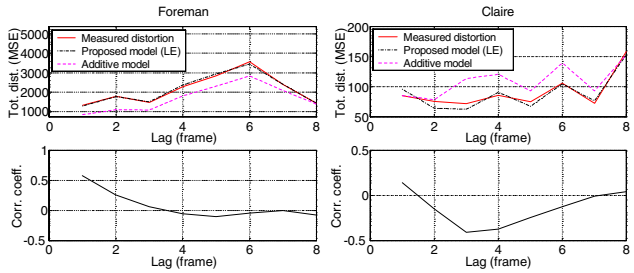

Fig. 2. Total distortion and error correlation of two losses with a lag. First loss at Frame 80, and second loss at frame 80+lag.

Table 1. Averaged 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).

Sequence	Foreman	Mother	Salesman	Claire
Additive	-1.64	-1.40	-1.31	-1.47
Proposed (LE)	-0.24	-0.18	-0.41	0.07
Proposed (GE)	0.14	-0.11	-0.71	-0.18

is not equivalent to an equal number of isolated losses. This is consistent with [5, 6]. 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. Table 1 lists the modeling error for the special case of $B = 2$, and it is clear that the proposed model estimates the total distortion to within ± 0.25 dB for most sequences while the additive model underestimates it by about 1.5 dB.

Fig. 2 plots the measured versus estimated distortion for two losses separated by different lags, as well as the error correlation, 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; while it overestimates the distortion for *Claire* due to the negative correlation. Fig. 3 plots the distortion for two losses separated by different lags, averaged over all loss realizations. 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, in particular 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.

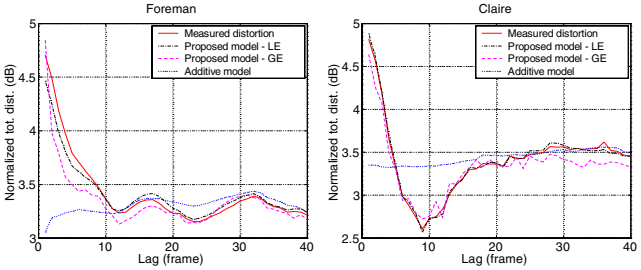

Fig. 3. Measured versus estimated total distortion for two losses separated by a lag, normalized by total distortion for a single loss.

5. CONCLUSIONS

We have shown that the packet loss pattern, and in particular the burst length, is important for accurately estimating the distortion for video communication over lossy packet networks. We proposed a model that explains why a loss pattern, such as a burst loss, generally produces a larger distortion than an equal number of isolated losses. This model enables a significant improvement in accurately estimating the distortion for different loss events. Specifically, for most sequences, the proposed model accurately predicts the total distortion to within 0.25 dB for burst loss of length two, as compared to the prior additive model which underestimates by about 1.5 dB. Furthermore, our accuracy is within 0.7 dB as the burst length increases, while that of the prior model degrades and may underestimate the distortion by over 3 dB. We expect that the use of this more accurate loss model can improve the design and performance of error-resilient video communication schemes.

6. REFERENCES

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

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

[3] Y.J. Liang and B. Girod, "Low-latency streaming of pre-encoded video using channel-adaptive bitstream assembly," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, Aug. 2002.

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

[5] J.G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. SPIE, VCIP*, Jan. 2001, pp. 392–409.

[6] 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. Acoust., Speech, and Signal Processing, ICASSP'02*, May 2002.

[7] 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. 3, pp. 393–401, July 2002.