Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

A channel distortion model for video over lossy packet networks[*]

CHENG Jian-xin[†], GAO Zhen-ming, ZHANG Zhi-chao
(*School of Information Science and Engineering, Shandong University, Jinan 250100, China*)
[†]E-mail: jxcheng.cn@gmail.com

Received Dec. 15, 2005; revision accepted Feb. 22, 2006

Abstract: Error-resilient video communication over lossy packet networks is often designed and operated based on models for the effect of losses on the reconstructed video quality. This paper analyzes the channel distortion for video over lossy packet networks and proposes a new model that, compared to previous models, more accurately estimates the expected mean-squared error distortion for different packet loss patterns by accounting for inter-frame error propagation and the correlation between error frames. The accuracy of the proposed model is validated with JVT/H.264 encoded standard test sequences and previous frame concealment, where the proposed model provides an obvious accuracy gain over previous models.

Key words: Channel distortion, Packet loss, Inter-frame error propagation, Correlation, Video communication
doi:10.1631/jzus.2006.AS0048 **Document code:** A **CLC number:** TN919.8

INTRODUCTION

Video transmission over Internet and wireless networks has much significant attention in recent years. But difficulties for video transport over Internet and wireless networks lie in the fluctuation of available bandwidth, high error rate and long delay. Video communication over lossy packet switched networks is often impaired by packet losses due to congestion, erasures and late delivery. In order to account for the channel introduced losses and maximize the reconstructed video quality at the receiver, various techniques have been proposed, including intra/inter-mode switching (Zhang *et al*., 2000), dynamic control of prediction dependency (Liang and Girod, 2002), forward error correction (Tan and Zakhor., 2001), multiple description coding (Apostolopoulos *et al*., 2002), and Rate-Distortion Optimized (RaDiO) packet scheduling (Chakareski *et al*., 2002). All these approaches are designed and operated based on models for the effect of losses on the reconstructed video quality. Therefore, the accuracy of the employed distortion models is one of the crucial factors affecting the performance of error-resilient video communication.

Previous work on modelling the effect of losses generally models the distortion as being directly proportional to the number of losses that occur (Stuhlmuller *et al*., 2000). This additive model is accurate as long as the burst loss does not lead to the loss of more than a single frame, where the number of lost frames depends on the number of packets per frame relative to the burst length in packets, e.g. when the loss rate is low and the losses are not bursty. However, the additive model proved to be not accurate enough for low-bit-rate video communication over the Internet or a wireless link. Liang *et al*.(2003) proposed two general packet loss patterns and distortion models for the general packet loss patterns by accounting for inter-frame error propagation and the correlation between error frames, concatenated and combined those models to obtain the general patterns of the distortion of losses. Compared to the previous additive models, the distortion models provide great improvements as they accurately account for the different effects of different loss patterns although they do not account for the correlation between general

---

[*] Project (No. Y2001G05) supported by the Natural Science Foundation of Shandong Province, China

loss patterns. Actually, the proposed two general packet loss patterns: burst losses of length B ($B\geq 1$) and two losses separated by a short lag can be replaced with a general loss pattern: burst losses of length B ($B\geq 1$) and a short lag.

In this paper, we propose a new model for predicting the distortion for different loss events by concatenating general loss patterns and accounting for the correlation between general loss patterns. The proposed model provides an obviously more accurate estimate of the distortion resulting from different loss events, compared to previous models. The accuracy is validated for two video test sequences coded with the JVT/H.264 standard.

PROPOSED PACKET LOSS MODEL

We first introduce some necessary notation and background. We follow the notation and background of previous work (Liang *et al*., 2003). We analyze the case where sequences with an I-P-P-P frame structure that have a certain number of macroblocks periodically intra updated. For simplicity, we assume that each P-frame is coded into a single packet, so that the loss of a packet corresponds to the loss of an entire frame. However, the results in this paper can also be extended to the case when each frame is coded into multiple packets.

The original video signal is denoted by $s[x,y,n]$, where $n\in\mathbb{Z}$ is the frame index. To simplify notation, we use the 1D vector $\boldsymbol{f}[n]$ to represent an original video frame, $\hat{\boldsymbol{f}}[n]$ to denote the loss-free reconstruction of the frame, and $\boldsymbol{g}[n]$ to denote the reconstruction at the decoder after loss concealment. The error frame at frame n introduced by one or more packet losses that occurred earlier is defined as

$$\boldsymbol{e}[n]=\boldsymbol{g}[n]-\hat{\boldsymbol{f}}[n],$$

which is also a 1D vector. We assume previous frame loss concealment. Therefore, if frame n is the first occurrence of packet loss, then $\boldsymbol{g}[n]=\hat{\boldsymbol{f}}[n-1]$. Since our primary concern is the effect of channel loss, quantization error is not included in our study. So the mean square error (MSE) associated with error frame $\boldsymbol{e}[n]$ is given by

$$\sigma^2[n]=(\boldsymbol{e}^{\mathrm{T}}[n]\cdot\boldsymbol{e}[n])/M,$$

where M is the number of pixels in frame n.

The above MSE quantifies the error power introduced in a single frame due to previous packet losses. It 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. Using these pre-measured distortions for single and independent losses, we can accurately estimate the distortion for more general loss patterns using the model proposed in this paper. We denote the initial error frame resulting from a single lost frame n by $\boldsymbol{e}_{\mathrm{S}}[n]$, and its MSE by $\sigma_{\mathrm{S}}^2[n]$; while $\boldsymbol{e}[n]$ and $\sigma^2[n]$ are used for losses with more general patterns. 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, $D_{\mathrm{S}}[n]$ denotes the total distortion resulting from a single frame loss at frame n.

At the decoder output, the total distortion D resulted by a loss is from the distortion for initial lost frames D_{d} and the distortion for error propagation D_{p}. So

$$D=D_{\mathrm{d}}+D_{\mathrm{p}}.$$

Distortion for initial burst consecutive losses

In the following, we assume a simple loss concealment scheme where the lost frame is replaced by the previous frame at the decoder output. We first consider the error frames that result for single losses at n–1 and n which are given by

$$\boldsymbol{e}_{\mathrm{S}}[n-1]=\hat{\boldsymbol{f}}[n-2]-\hat{\boldsymbol{f}}[n-1], \tag{1}$$

and

$$\boldsymbol{e}_{\mathrm{S}}[n]=\hat{\boldsymbol{f}}[n-1]-\hat{\boldsymbol{f}}[n], \tag{2}$$

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

$$\boldsymbol{e}[n]=\hat{\boldsymbol{f}}[n-2]-\hat{\boldsymbol{f}}[n]=\boldsymbol{e}_{\mathrm{S}}[n-1]+\boldsymbol{e}_{\mathrm{S}}[n]. \tag{3}$$

The corresponding MSE of error frame n is

$$\sigma^2[n]=\sigma_{\mathrm{S}}^2[n-1]+\sigma_{\mathrm{S}}^2[n]+2\rho_{n-1,n}\sigma_{\mathrm{S}}[n-1]\sigma_{\mathrm{S}}[n], \quad (4)$$

where $\rho_{n-1,n}=\dfrac{(\boldsymbol{e}_{\mathrm{S}}^{\mathrm{T}}[n-1]\boldsymbol{e}_{\mathrm{S}}[n])/M}{\sigma_{\mathrm{S}}[n-1]\sigma_{\mathrm{S}}[n]}$ is the correlation coefficient between error frames *n*–1 and *n*.

We extend the above to model burst losses of length *B* ($B\geq 1$). For the loss of *B* consecutive frames from *n*–*B*+1 to *n*,

$$\boldsymbol{e}[n]=\hat{\boldsymbol{f}}[n-B]-\hat{\boldsymbol{f}}[n]=\sum_{i=n-B+1}^{n}\boldsymbol{e}_{\mathrm{S}}[i],$$

and its MSE

$$\sigma^2[n]=\sum_{i=n-B+1}^{n}\sigma_{\mathrm{S}}^2[i]+2\sum_{i=n-B+1}^{n}\sum_{j=i+1}^{n}\rho_{i,j}\sigma_{\mathrm{S}}[i]\sigma_{\mathrm{S}}[j], \quad (5)$$

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

The distortion for initial consecutive *B* lost frames is

$$D_{\mathrm{d}}[n-B+1,\ldots,n]=\sum_{i=n-B+1}^{n}\sigma^2[i]. \quad (6)$$

Distortion for error propagation

We model the error propagation process in a typical video decoder with a geometric attenuation factor (Liang *et al*., 2003) and a linear attenuation factor to account for the spatial filtering and intra update (Stuhlmuller *et al*., 2000), respectively. With an intra update period of *N*, if a single error is introduced at *n* with an MSE of $\sigma^2[n]$, the power of the propagated error at *n*+*l* is given by

$$\sigma^2[n+1]=\begin{cases}\sigma^2[n]r^l(1-l/N), & 0\leq l\leq N,\\ 0, & \text{otherwise.}\end{cases} \quad (7)$$

The attenuation factor *r* ($0<r<1$) 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.

If there are no losses in consecutive *L* frames following frame *n* and the frame *L*+1 ($\leq N$) is lost, the distortion for error propagation introduced by the loss at frame *n* is

$$D_{\mathrm{p}}[n]=\sum_{i=n+1}^{L}\sigma^2[i]=\sum_{i=1}^{L}r^i\left(1-\frac{i}{N}\right)\sigma^2[n]. \quad (8)$$

If $L+1>N$, the effect of error propagation introduced by the loss at frame *n* is completely removed,

$$D_{\mathrm{p}}[n]=\sum_{i=n+1}^{N-1}\sigma^2[i]. \quad (9)$$

Modelling of total distortion

We now estimate the total distortion for general loss patterns. A general loss pattern consists of consecutive *B* frames losses and an arbitrary lag of $1<L\leq N$. According to Eqs.(5), (6), (8) and (9), the total distortion for consecutive *B* frames losses is

$$D[n-B+1,\ldots,n]=\sum_{i=n-B+1}^{n}\sigma^2[i]+D_{\mathrm{p}}[n].$$

If the distance between current losses and next losses is *L* frames and $L<N$,

$$D[n-B+1,\ldots,n]=\sum_{i=n-B+1}^{n}\sigma^2[i]+\frac{(N-L)r^{L+2}-(L-N+1)r^{L+1}-(N+1)r+N}{N(1-r)^2}\sigma^2[n]. \quad (10)$$

If $L\geq N$, we can let $L=N$. So

$$D[n-B+1,\ldots,n]=\sum_{i=n-B+1}^{n}\sigma^2[i]+\frac{r^{N+1}-(N+1)r+N}{N(1-r)^2}\sigma^2[n], \quad (11)$$

where *r* is a parameter describing how effective the spatial filter is in reducing the introduced error power, and dependent on the strength of the loop filter of the codec and power spectrum density (PSD) of the input error signal. The method determining the *r* value is given in Section 2.4.

Total distortion model of different loss events

If the distance between different loss events is *L* frames and $L\geq N$, these distortions are independent and the correlation is 0. The total distortion is the sum of the distortions resulting from different loss events.

If the distance between different loss events is shorter than the intra update period ($L<N$) and the distortion at the last frame of the previous loss event is $\sigma^2[n]$, the distortion for error propagation at the first frame of the current loss event is

$$\sigma^2[n+L]=\sigma^2[n]r^L(1-L/N).$$

If the distortion of the initial lost frame at the first frame of the current loss event is $\sigma_{\mathrm{S}}^2[m]$, then the distortion introduced at the current frame is

$$\sigma^2[m]=\sigma^2[n+L]+\sigma_{\mathrm{S}}^2[m]+2\rho_{n+L,m}\sigma[n+L]\sigma_{\mathrm{S}}[m], \quad (12)$$

where

$$\rho_{n+L,m}=\frac{(\boldsymbol{e}^{\mathrm{T}}[n+L]\boldsymbol{e}_{\mathrm{S}}[m])/M}{\sigma[n+L]\sigma_{\mathrm{S}}[m]}.$$

$\boldsymbol{e}[n+L]$ and $\sigma^2[n+L]$ can be obtained by simulating the corresponding loss event at the encoder. Now we can estimate the distortion for arbitrary loss events.

In this model, it is very important to determine r value. According to Eq.(7), for a single error at n, and considering a period that is sufficiently long for complete error recovery, the total distortion is

$$D_{\mathrm{S}}[k]=\frac{r^{N+1}-(N+1)r+N}{N(1-r)^2}\sigma_{\mathrm{S}}^2[n]=\alpha\sigma_{\mathrm{S}}^2[n],$$

where $\alpha=D_{\mathrm{S}}[k]/\sigma_{\mathrm{S}}^2[n]$ is the ratio between the total distortion and the MSE of frame n. Since the variation of r from frame to frame is low, it is assumed that, for a fixed burst losses length, r (and α) is constant for the entire recovery period, and independent of frame index n. However, as the burst length B varies, the shape of the initial error signal's PSD also varies, which leads to a variation in r (or α). Researches showed that the process of error power reduction by loop filtering can be modelled with a linear system and that r is the proportion of the power of the introduced error passing through the system. The loop filter can be approximated by a Gaussian low-pass filter (Stuhlmuller *et al.*, 2000). Hence, as B increases, r (and α) increases as the PSD of the error is more concentrated in the lower band. So we can determine r as that obtained in previous work (Liang *et al.*, 2003). As B increases, the variation of α is relatively small and can be approximated as a linear function of B, that is $\alpha(B)=\alpha_0+c(B-2)$, where α_0 is the ratio for $B=2$, c is the slope of the increase, and $B\geq2$. α can be determined by two measured values for different B_{S}. With the obtained α, r is determined.

SIMULATION RESULTS

To validate the accuracy of the proposed model, and to compare it versus previous models, we simulate different loss patterns on standard video test sequences, and compare the measured distortion with that predicted by the proposed models, the previous model (Liang *et al.*, 2003) and the additive model (Stuhlmuller *et al.*, 2000). Video sequences are coded using JM 8.6 of the JVT/H.264 video compression standard (Telecommunication Standardization Sector of ITU, 2003). Two standard test sequences, Foreman and Claire in QCIF format are used. 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, corresponding to an intra-frame update period of N=36 frames.

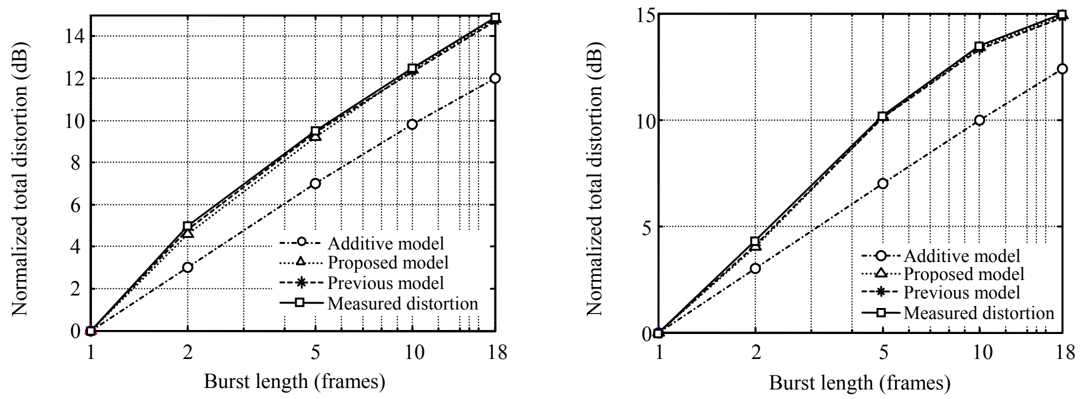
In our simulations, we use 140 frames for parameters estimations in each test sequence. To calculate the distortion of an arbitrary loss event, the MSE of a single loss $\sigma_{\mathrm{S}}^2[n]$ and $D_{\mathrm{S}}[n]$ are pre-measured for every frame. To estimate the required model parameters r, L decodings are required for two losses and $2L$ decodings required for $B>2$, so r can be calculated. With the obtained parameters, the total distortion can be calculated using the model in Eqs.(10)~(12).

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

Fig.1 shows that the proposed model and the previous model can both more accurately estimate the total distortion than the additive model, and that these two models are similar in performance for consecutive burst losses.

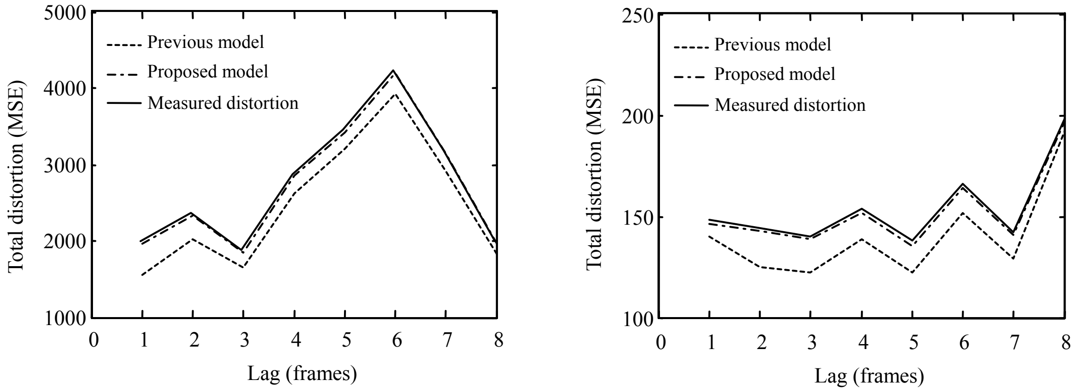
Fig.2 plots the measured versus estimated distortion for two losses separated by different lags for one particular loss event in which the first losses of length 2 occur at frame 80. Due to neglecting the correlation between the first loss and two losses separated by a lag, the previous model obviously underestimates the distortion, while the proposed model can obviously more accurately estimate the distortion.

Fig.3 plots the measured versus estimated total distortion for two burst losses of length 8 separated by different lags, averaged over all loss events. For each loss event, we simulate the loss event starting at different frame in the video sequence, decode and compute the resulting total distortion for each starting frame. The previous model (Liang *et al.*, 2003) estimates the losses event with two general patterns (burst losses of length 7) and a general pattern (two losses separated by a lag). While the proposed model estimates this losses event by concatenating two general patterns defined in this paper. For Foreman, the proposed model estimates the distortion to an accuracy of within ±0.08 dB, while the previous model underestimates the error by up to 0.23 dB. For Claire, the proposed model estimates the distortion to an accuracy of within ±0.1 dB, while the previous model underestimates the error by up to 0.28 dB. To summarize the results for this figure, the proposed model provides much higher accuracy, in particular for small lag and more complex losses event.



(a) (b)

Fig.1 Measured versus estimated total distortion as a function of burst loss length, normalized by total distortion for a single loss. (a) Foreman; (b) Claire



(a) (b)

Fig.2 Total distortion of two losses with a lag. First losses of length 2 at frame 80, and second loss at frame 80+lag. (a) Foreman; (b) Claire

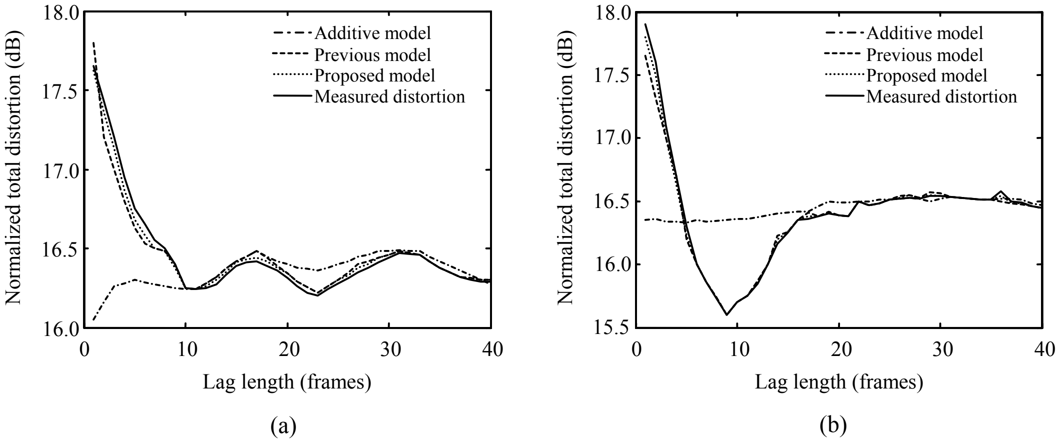


Fig.3 Measured versus estimated total distortion for two burst losses of length 8 separated by a lag, normalized by total distortion for a single loss. (a) Foreman; (b) Claire

CONCLUSION

After analyzing different packet loss patterns over lossy packet networks, we propose a new general loss pattern to approach arbitrary loss events. The simulation result indicated that the new general loss packet pattern can more accurately approach to packet loss events and that the proposed model can more accurately estimate the distortion resulting from different packet losses, in particular when packet loss events consist of some non-consecutive losses spaced apart by small lags. Obviously, these complex packet loss events are more general for video transport over Internet and wireless networks. According to feedback information on packet losses (e.g., indexes of lost packets) from the receiver, the proposed model can accurately model the effect of corresponding losses on the reconstructed video quality. Compared to the previous model, the proposed model is more suitable for video communication over very lossy packet networks. We expect that the error-resilient video communication over lossy packet networks can benefit from the use of the new model.

References

Apostolopoulos, J.G., Tan, W., Wee, S.J., Wornell, G.W., 2002. Modeling Path Diversity for Multiple Description Video Communication. Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP'02. Orlando, FL, **3**:2161-2164.

Chakareski, J., Chou, P.A., Girod, B., 2002. Rate-Distortion Optimized Streaming from the Edge of the Network. Proc. Workshop on Multimedia Signal Processing, IEEE. St. Thomas, US Virgin Islands, p.49-52.

Liang, Y.J., Girod, B., 2002. Low-latency Streaming of Pre-encoded Video Using Channel-adaptive Bitstream Assembly. Proc. Int'l Conf. Multimedia and Exhibition, IEEE. Lausanne, Switzerland, **1**:873-876.

Liang, Y.J., Apostolopoulos, J.G., Girod, B., 2003. Analysis of Packet Loss for Compressed Video: Does Burst-length Matter. Proc. Int'l Conf. Acoustics, Speech, and Signal Processing, IEEE. Hong Kong, China, **5**:684-687.

Stuhlmuller, K., Farber, N., Link, M., Girod, B., 2000. Analysis of video transmission over lossy channels. *IEEE J. Selected Areas in Communications*, **18**(6):1012-1032. [doi:10.1109/49.848253]

Tan, W.T., Zakhor, A., 2001. Video multicast using layered FEC and scalable compression. *IEEE Trans. Circuits and Systems for Video Technology*, **11**(3):373-378. [doi:10.1109/76.911162]

Telecommunication Standardization Sector of ITU, 2003. Video Coding for Low Bit Rate Communication. Draft ITU-T Recommendation H.264.

Wang, Y., Ostermann, J., Zhang, Y.Q., 2002. Video Processing and Communications. Pearson Education, Inc.

Zhang, R., Regunathan, S.L., Rose, K., 2000. Video coding with optimal intel/intra-mode switching for packet loss resilience. *IEEE J. Selected Areas in Communications*, **18**(6):966-976. [doi:10.1109/49.848250]