MODELING OF DISTORTION CAUSED BY MARKOV-MODEL BURST PACKET LOSSES IN VIDEO TRANSMISSION

Zhicheng Li, Jacob Chakareski, Xiaodun Niu, Gaoxi Xiao, Yongjun Zhang, and Wanyi Gu

Communication Research Laboratory
Nanyang Technological University, Singapore

ABSTRACT

This paper addresses the problem of distortion modeling for video transmission over burst-loss channels characterized by a finite state Markov chain. A Distortion Trellis model is proposed, enabling us to estimate at the frame level the expected mean-square error (MSE) distortion caused by Markov-model bursty packet losses. A sliding window algorithm is developed to perform the MSE estimation with low complexity. Simulation results show that the proposed models are accurate for all tested average loss rates and average burst lengths.

Index Terms— distortion modeling, burst-loss channel, Markov-model loss process, error propagation.

1. INTRODUCTION

Internet packet loss often exhibits finite temporal dependency, which leads to bursty packet losses, a characteristic not found on traditional Bernoulli-model loss process, but can be characterized by the finite-state Markov-model loss process [1]. Still, to the best of our knowledge, a complete mathematical model relating the channel-induced distortion in decoded video and the Markov-model bursty packet loss process has not been proposed yet. In particular, very little analytical work has been done on estimating the expected distortion at the encoder given a Markov burst-loss channel model.

So far, many channel-induced distortion models have been proposed [2]. However, these existing works assume that the underlying network packet losses are independent and identically distributed (i.i.d.) and therefore employ a Bernoulli loss model to this end. These models only consider the average loss rate in the absence of another factor, the burst length, and therefore are less efficient for the case of video transmission over burst-loss channels. The work in [3] shows that the burst length does affect the distortion. Similarly, the Distortion Chain model in [4] predicts the end-to-end distortion for arbitrary loss patterns. However, both of these works do not consider explicitly in their analysis the channel correlation that exists between the individual packet losses.

This paper develops a mathematical framework denoted as the Distortion Trellis model for estimating the expected MSE distortion caused by Markov-model bursty packet losses. Without loss of generality and for simplicity, we derive our distortion model for a two-state Markov loss model, or the Gilbert model [5]. The proposed techniques can be easily extended to most finite state Markov loss models.

The paper is organized as follows. Section 2 formulates the problem. In Section 3 presents the framework of the proposed Distortion Trellis model, while Section 4 describes the sliding window algorithm for calculating the MSE distortion with low complexity. In Section 5, we study the performance of the proposed techniques through simulation experiments.

2. PROBLEM FORMULATION

We assume a raw video sequence is separated into groups of pictures (GOPs) and each GOP starts with an I-frame, followed by P-frames. We do not consider B-frames in this paper. All the MBs in a frame are grouped into one slice and each slice is coded into one network packet. The channel losses can be characterized via a Gilbert model, and the channel drops or delivers the packet according to the current channel state. At the decoder, certain temporal error concealment strategy is applied when a P-frame is lost.

For a Gilbert channel [5], let p be the transition probability from error-free state to error state, and q denotes the probability of the opposite transition. The stationary probability for error-free state and error state, denoted by π_0 and π_1, can be computed as $\pi_0 = q/(p+q)$ and $\pi_1 = p/(p+q)$, respectively. Then, the mean packet loss ratio PLR equals π_1, and the average burst length ABL is given by $1/q$.

Let x_n^i and y_n^i be the reconstructed pixel values for frame n and pixel i at the encoder and at the decoder, respectively. Then, the expected distortion of frame n can be defined as

$$d_n = E_c\{d_n^c\} = E_c\{E_{pix}\{(x_n^i - y_n^i)^2\}\}, \qquad (1)$$

where d_n^c denotes the average MSE distortion for frame n for channel realization c, $E_{pix}\{\cdot\}$ denotes the computation of the average MSE over all pixels in frame n, $E_c\{\cdot\}$ denotes the expectation taken over all possible channel realizations. This paper mainly focuses on modeling d_n for successive P-frames in video transmission over a Gilbert channel.

3. PROPOSED DISTORTION TRELLIS MODEL

3.1. Framework of the Distortion Trellis model

Consider the impairments for a transmitted packet (frame) sequence of length n as an n-bit binary random variable $K_n = \{B_j\}_{j=1}^{n}$. The random variable B_j is over the binary alphabet $\{0, 1\}$. $B_j = 1$ indicates that the j-th frame is lost. Then, the total number of all possible values of K_n is 2^n. Define moreover an ordered set $\mathbf{I}_n = \{k_n^r\}$, $r = 1, \ldots, 2^n$, where k_n^r is an n-bit binary number and $k_n^1 = (0 \ldots 0)_{n\text{bits}}$, $k_n^r = 1 + k_n^{r-1}$, $r = 2, \ldots, 2^n$. Furthermore, we assume that the r-th value of K^n is k_n^r, the r-th element in $\mathbf{I}_n$. Note that in our analysis this is an important assumption.

Let $P(k_n^r)$ denote the probability that loss pattern k_n^r occurs. Note that different loss patterns lead to different distortion values. Let d_n^r be the decoder distortion of the n-th frame in a frame sequence of length n under loss pattern k_n^r. Then, d_n^r can be defined as $d_n^r = E_{pix}\{(x_n^i - y_{n,r}^i)^2\}$, where $y_{n,r}^i$ denotes the decoder reconstructed value of pixel i in the n-th frame for an n-length frame sequence under loss pattern k_n^r. From the definition of d_n^r, we can obtain an important probability relation as follows: $\mathrm{P_r}$(at the decoder the distortion of frame n is d_n^r)$= P(k_n^r)$. Fig.1 shows the statistical dependencies between the elements of the set $\{d_n^r, n = 1, 2, \ldots\}$. Since it is in a trellis shape, we refer to the proposed distortion estimation method as the Distortion Trellis model. Then, we can calculate d_n by taking an expectation over all possible decoder distortion values for frame n,

$$d_n = E_c\{d_n^c\} = \sum_{r=1}^{2^n} d_n^r \cdot P(k_n^r), n = 1, 2, \ldots \tag{2}$$

The formula in (2) is the general form of the proposed Distortion Trellis model. From (2), it is clear that the computation of d_n necessitates knowledge of both d_n^r and $P(k_n^r), r = 1, \ldots, 2^n$. In a Gilbert channel, $P(k_n^r)$ can be derived recursively, as follows. From the definition of $\{k_n^r\}$, it is clear that given $P(k_{n-1}^t), t = 1, \ldots, 2^{n-1}$, the loss pattern probabilities can be written as,

$$\begin{cases} P(k_n^{4r-3}) = (1-p) \cdot P(k_{n-1}^{2r-1}) \\ P(k_n^{4r-2}) = p \cdot P(k_{n-1}^{2r-1}) \\ P(k_n^{4r-1}) = q \cdot P(k_{n-1}^{2r}) \\ P(k_n^{4r}) = (1-q) \cdot P(k_{n-1}^{2r}), r = 1, \ldots, 2^{n-2}. \end{cases} \tag{3}$$

The remaining task in this section is how to calculate d_n^r. Given that the loss pattern of the previous $n-1$ frames is k_{n-1}^r, d_n^{2r-1} is the frame-average distortion if the n-th frame is received while d_n^{2r} denotes the same quantity for the case when the n-th frame is lost. Next, we separately consider computing d_n^{2r-1} and d_n^{2r}.

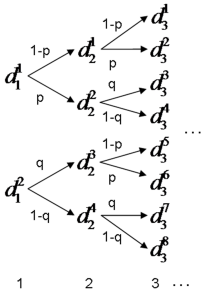

Fig. 1. Statistical dependencies $\{d_n^r, n = 1, 2, \ldots\}$.

3.2. Computation of d_n^{2r} and d_n^{2r-1}

In our assumption, if a frame is lost, all MBs in this frame are recovered using some temporal error concealment strategy, regardless their coding modes. Let $f_l(i)$ denote the index of the l-th pixel in frame $n-1$ that is used to estimate pixel i in frame n. Then the final concealed value of $y_{n,2r}^i$ can be expressed as $\Phi_l(y_{n-1,r}^{f_l(i)})$, where Φ_l represents the pixel operation on $y_{n-1,r}^{f_l(i)}$ for all l used in obtaining the final concealed value of $y_{n,2r}^i$. For example, Φ_l could denote the interpolation operation and the deblocking operation. For previous frame copy concealment, $\Phi_l(y_{n-1,r}^{f_l(i)}) = y_{n-1,r}^i$. It is a reasonable assumption that Φ_l is same for different frames. Then, d_n^{2r}can be derived as follows:

$$\begin{aligned} d_n^{2r} &= E_{pix}\{(x_n^i - \Phi_l(y_{n-1,r}^{f_l(i)}))^2\} \\ &= E_{pix}\{(x_n^i - \Phi_l(x_{n-1}^{f_l(i)}))^2\} \\ &\quad + E_{pix}\{(\Phi_l(x_{n-1}^{f_l(i)}) - \Phi_l(y_{n-1,r}^{f_l(i)}))^2\} \\ &= ECD_n + E_{pix}\{(\Phi_l(x_{n-1}^{f_l(i)} - y_{n-1,r}^{f_l(i)}))^2\}, \end{aligned} \tag{4}$$

where $r = 1, \ldots, 2^{n-1}$, $ECD_n = E_{pix}\{(x_n^i - \Phi_l(x_{n-1}^{f_l(i)}))^2\}$. Note that the second identity in (4) is based on the assumption that the concealment error $x_n^i - \Phi_l(x_{n-1}^{f_l(i)})$ and the propagation error $\Phi_l(x_{n-1}^{f_l(i)}) - \Phi_l(y_{n-1,r}^{f_l(i)})$ are uncorrelated. The third identity is based on the assumption that Φ_l is linear.

Furthermore, when the error in a previous frame propagates into the current frame, it is typically attenuated by the adoption of some coding schemes such as deblocking filtering and sub-pixel motion estimation, whose effect can be regarded as a spatial filter or more precisely as an error attenuator. Following this reasoning, $E_{pix}\{(\Phi_l(x_{n-1}^{f_l(i)} - y_{n-1,r}^{f_l(i)}))^2\}$ in (4) can be approximated as $u \cdot E_{pix}\{(x_{n-1}^i - y_{n-1,r}^i)^2\} = u \cdot d_{n-1}^r$. Then (4) can be rewritten as

$$d_n^{2r} = ECD_n + u \cdot d_{n-1}^r, r = 1, \ldots, 2^{n-1}, \tag{5}$$

where u is the error attenuation factor for a lost frame.

In received inter-coded MBs and intra-coded MBs, the distortions are different. We first consider the case when all

MBs are coded in inter mode. Let $g_l(i)$ denote the index of the l-th pixel in frame $n-1$ that is used to estimate pixel i in frame n. Note that $g_l(i)$ may differ from $f_l(i)$. Then, at the encoder, the predicted value of x_n^i can be expressed as $\Psi_l(x_{n-1}^{g_l(i)})$, where Ψ_l represents the pixel operation on all $x_{n-1}^{g_l(i)}$ used for obtaining the predicted value of x_n^i, such as when performing interpolation or deblocking filtering. We also assume that Ψ_l is linear and has the same form for different frames. Similarly, at the decoder, the predicted value of $y_{n,2r-1}^i$ is $\Psi_l(y_{n-1,r}^{g_l(i)})$. Then, d_n^{2r-1} can be derived as

$$\begin{aligned} d_n^{2r-1} &= E_{pix}\{(\Psi_l(x_{n-1}^{g_l(i)}) - \Psi_l(y_{n-1,r}^{g_l(i)}))^2\} \\ &= E_{pix}\{(\Psi_l(x_{n-1}^{g_l(i)} - y_{n-1,r}^{g_l(i)}))^2\}. \end{aligned} \quad (6)$$

As in the case of d_n^{2r}, the operator Ψ_l can be regarded as a spatial filter that will attenuate the error propagation. Hence, we similarly employ $v_0 \cdot d_{n-1}^r$ to approximate $E_{pix}\{(\Psi_l(x_{n-1}^{g_l(i)} - y_{n-1,r}^{g_l(i)}))^2\}$ and therefore we can rewrite (6) as $d_n^{2r-1} = v_0 \cdot d_{n-1}^r, r = 1, \ldots, 2^{n-1}$, where v_0 is the error attenuation factor for a received frame, in which all MBs are coded in inter mode. The above development assumes that all MBs in a P-frame are coded in an inter mode. However, a P-frame often contains intra-coded MBs, which will effectively restrain the error propagation. The effect of macroblock intra refreshing can also be considered as an attenuator that attenuates the error signal from an impaired previous frame. Therefore, to take this into account we introduce a new constant λ and rewrite (6) as

$$d_n^{2r-1} = v \cdot d_{n-1}^r, r = 1, \ldots, 2^{n-1}, \quad (7)$$

where $v = \lambda \cdot v_0$.

3.3. Recursive computation of d_n

Based on (5) and (7), we can recursively obtain the distortion d_n^r, for $r = 1, \ldots, 2^n$. The loss pattern probability $P(k_n^r)$ can be recursively calculated with (3). Then, using (2), the expected distortion d_n for Gilbert channel packet losses can be estimated as

$$\begin{aligned} d_n &= \sum_{t=1}^{2^n} d_n^t P(k_n^t) = \sum_{r=1}^{2^{n-2}} [P(k_n^{4r}) d_n^{4r} + P(k_n^{4r-1}) d_n^{4r-1} \\ &\quad + P(k_n^{4r-2}) d_n^{4r-2} + P(k_n^{4r-3}) d_n^{4r-3}], \end{aligned} \quad (8)$$

It can be seen that d_n depends on u, v, ECD_n, p, and q. The former three parameters depend on the video sequence. The parameter pair p and q is used to describe the Gilbert channel and is equivalent to another parameter pair, PLR and ABL, which are more commonly used. Then, for video transmission over a Gilbert channel, given PLR, ABL, the initial probability distribution $P(k_1^1)$ and $P(k_1^2)$, and the initial distortion distribution d_1^1 and d_1^2, the expected distortion of each frame in a GOP can be estimated using (8). Often, this model fails to compute d_n within acceptable time.

4. SLIDING WINDOW ALGORITHM

Due to the intra refreshing and the spatial filtering, the propagation of error typically decays in magnitude over the subsequent frames. Based on the fact, we now propose a sliding window (SW) algorithm to calculate d_n for $n > W$ with low complexity, where W is an integer constant. When $n \leq W$, we employ the same approach described in Section 3 to calculate d_n. In most cases, the SW algorithm could provide more than 90% reduction in computational complexity.

Algorithm 1: SW for calculating d_n for $n = 1, \ldots, N$

Input: $PLR, ABL, u, v, W, N, \{ECD_n, n = 1, , N\}$.
Output: the expected distortion d_n for $n = 1, \ldots, N$;
Initialization: $d_1^1 = 0, d_1^2 = ECD_1, P(k_1^1) = 1 - PLR$, $P(k_1^2) = PLR, p = PLR/(ABL(1 - PLR)), q = 1/ABL$;
for $n = 1$ to N **do**
  if $n \leq W$ **then**
    Compute d_n using (8), note that when $n = W$, $P(k_W^j), j = 1, \ldots, 2^W$ are obtained;
  else
    Reset the initial distortion values: $d_1^1 = 0$, $d_1^2 = ECD_{n-W+1}$;
    for $i = 2$ to W **do**
      for $j = 1$ to 2^{i-1} **do**
        $d_i^{2j-1} = vd_{i-1}^j, d_i^{2j} = ECD_{n-W+i} + ud_{i-1}^j$;
      end for
    end for
    After the two for loops, $d_W^j, j = 1, \ldots, 2^W$ are obtained, then the output is $d_n = \sum_{j=1}^{2^W} P(k_W^j) d_W^j$;
  end if
end for

5. SIMULATIONS

The H.264 reference encoder JM12.2 with the Baseline profile is used to encode QCIF "Foreman" at 15fps and "Football" at 30fps, both with QP=28. The first frame is coded as an I-frame, followed by P-frames. At the decoder, frame-copy scheme is used for concealment, so that ECD_n equals to the mean square difference between two successive P-frames. To estimate the model parameters u and v, we use the least square fitting method based on training data.

Fig. 2 plots the average expected distortion for PLR values from 3% to 10% at $ABL = 2$. Due to the high complexity of the original Distortion Trellis model, we encode 20-frame segment starting at different positions in the original sequence. For each tested PLR and ABL pair, we simulate a Gilbert channel and generate 50,000 loss traces for each

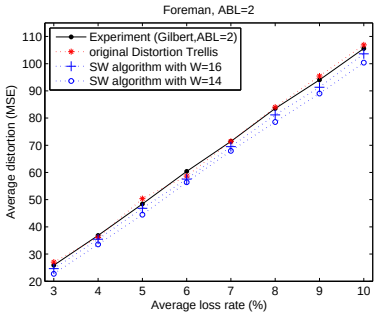

Fig. 2. Average distortion comparison

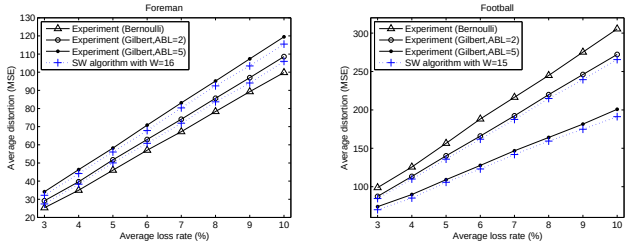

Fig. 3. Average expected distortion versus PLR

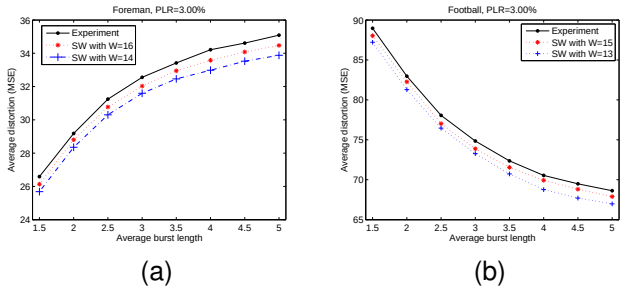

Fig. 4. Average expected distortion versus ABL

segment. The average expected distortion is then obtained by averaging all segments and all loss traces. It can be seen that the original Distortion Trellis model provides better prediction of the expected distortion than the SW algorithm along the whole tested PLR range. Although the SW algorithm is less accurate, it still matches the measured expected distortion quite well. Hereafter, we only use the SW algorithm to estimate the expected distortion.

Next, the average expected distortion over all P-frames versus PLR from 3% to 10%, for both ABL=2 and 5 is plotted in Fig. 3. For "Foreman" the first 390 frames are coded while for "Football" the first 240 frames are coded. For each tested PLR and ABL pair, we simulate a Gilbert channel and generate 90,000 random loss patterns. The average expected distortion for the case of a Bernoulli channel at the same loss rate is also plotted in the same figures for comparison, where 1000 loss traces are generated at each loss rate for this channel model. We see that the SW algorithm accurately estimates the average expected distortion. Interestingly, we observe that for "Foreman" the expected distortion for the Gilbert channel is larger while for "Football" the expected distortion for the Bernoulli channel is larger. Even more interestingly, we observe that increasing the average burst length does NOT always contribute to a larger expected distortion, for a given average loss rate. For example, in the case of "Foreman" a larger average burst length leads to a larger expected distortion, at the same average loss rate. However, the opposite holds in the case of "Football". To the best of our knowledge, the aforementioned experimental result is reported for the first time here.

To study further the impact of the average burst length on the average video quality, Fig. 4 shows the average expected distortion over all P-frames versus ABLs of 1, 1.5, ..., 5, for the same $PLR = 3\%$. For each sequence the first 200 frames are coded and for each tested PLR and ABL pair, 60,000 loss traces are generated. We see that the estimated distortion matches the measured data well along the whole tested ABLs. We clearly observe that the average expected distortion does NOT always increase as the average burst length increases at the same average loss rate. For "Football", increasing the average burst length will reduce the average expected distortion. This confirms the earlier findings from Fig. 3 that at the same average loss rate, a larger average burst length does not always lead to a larger distortion in the case of a Gilbert channel.

6. REFERENCES

[1] M. Yajnik, S. B. Moon, J. Kurose, and D. Towsley, "Measurement and modeling of the temporal dependence in packet loss," in *Proc. IEEE INFOCOM'99*, New York, NY, Mar. 1999, vol. 1, pp. 345–352.

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

[3] Y. J. Liang, J. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: Does burst-length matter," in *Proc. IEEE ICASSP*, Apr 2003, pp. 684–687.

[4] J. Chakareski, J. Apostolopoulos, S. Wee, W.-T. Tan, and B. Girod, "Rate-distortion hint tracks for adaptive video streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 10, pp. 1257–1269, Oct 2005.

[5] E. N. Gilbert, "Capacity of burst-noise channel," *Bell Syst. Tech.*, vol. 39, pp. 1253–1265, Sep 1960.